An Introduction to Computer Vision

Ying Wu
Electrical Engineering & Computer Science
Northwestern University
Evanston, IL 60208
yingwu@ece.northwestern.edu

Contents

1 What Motivates Us? 2

2 What Is This Area About? 2
2.1 Application Areas . . . 2
2.1.1 Human-Computer Interaction . . . 3
2.1.2 Intelligent Environments . . . 3
2.1.3 Multimedia . . . 4
2.1.4 Intelligent Robots . . . 4
2.2 Fundamental Research Issues . . . 4
2.2.1 Image Processing and Computer Vision . . . 4
2.2.2 Machine Learning and Pattern Recognition . . . 4

3 What Is Computer Vision? 5
3.1 Image Formation . . . 6
3.2 Low-level Image Processing . . . 6
3.3 Low-level Vision . . . 6
3.4 Middle-level Vision . . . 6
3.5 High-level Vision . . . 7

4 What Is This Course Going To Cover? 7

1 What Motivates Us?

An interesting question we always ask is what the next generation of computers is going to be like. To answer this question, let's recall our first touch of computer. At least, my experience was that I waved my hands and said "how are you" to a machinery. Obviously, no answer at all.

It was a dream that computers would be able to see and think, which has been driving us to explore various research issues to make this dream come true. Although computers become faster and faster, they are still quite dull, since they can neither see nor even perform simple reasonings. Obviously, we are not satisfied to just use our computers as a calculator, a word processer, a CD player, or a game station; instead, we expect computers to do more intelligent things like our human beings. For example,

- Can computers identify me by looking at my face or even my gait?
- Can computers know where I am looking at and what I am doing?
- Can computers tell what is a car and what is not a car?
- Can computers learn something by themselves?
- Can computers summarize a video for me?
- ...

2 What Is This Area About?

Obviously, with an interdisciplinary nature, this area involves fundamental research in image processing, computer vision/graphics, machine learning, pattern recognition, biomechanics and even psychology. Figure 1 shows a big picture of this area. On top of it are several major application areas such as human-computer interaction, robotics, virtual environments, and multimedia. The common foundation for such applications include computer vision, image processing and speech processing. Instead of taking some *ad hoc* approaches to audio and visual processing when the area was in its infantile stage, we are currently pursuing some intelligent ways by machine learning and pattern recognition, trying to achieve a kind of artificial intelligence.

2.1 Application Areas

We can imagine what a visually-capable and intelligent computer can do! We expect a revolution in next generation of computer: we do not use mice and keyboards anymore. Computers could understand our actions and our languages, they could think and feedback to us some kind of smart results in response to our commands, and they could even perform some missions on behalf of our human beings. Least but not last, we expect a rapid progress in the near future in such areas as intelligent human-computer interaction, robotics, virtual environments, intelligent environments, and multimedia.

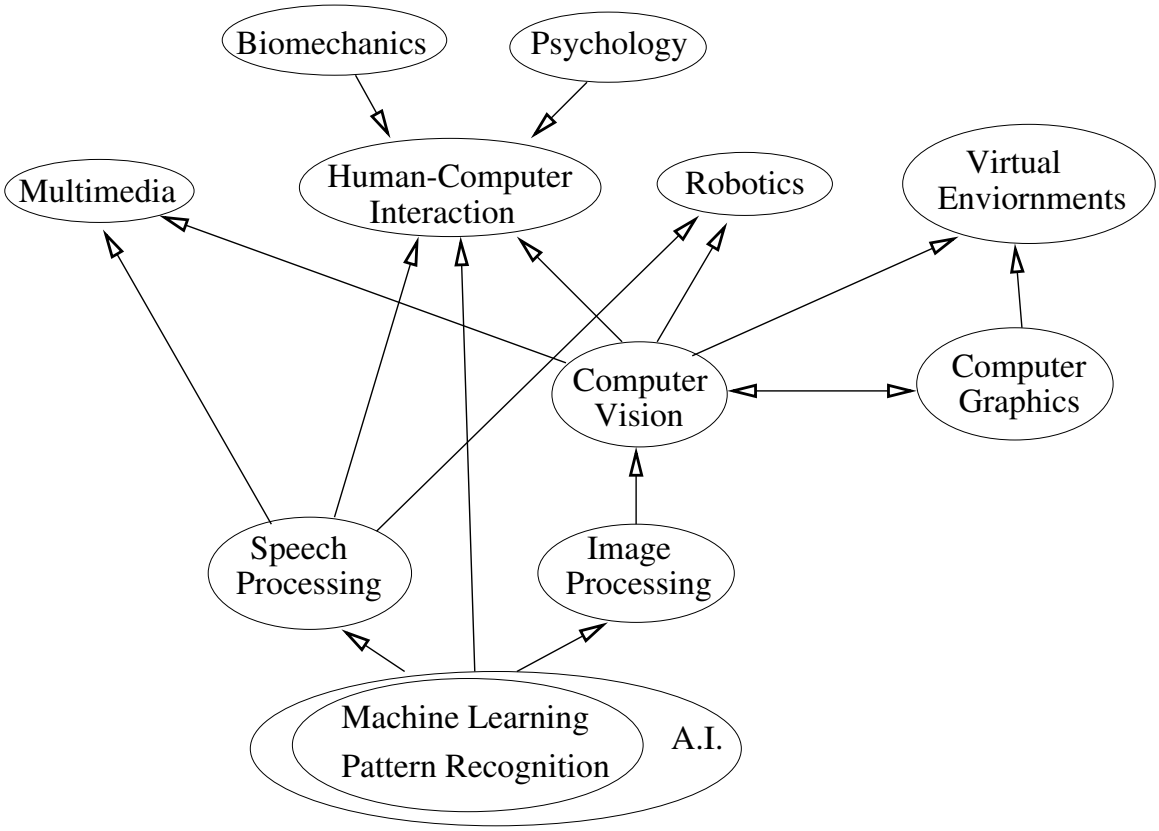

Figure 1: The big picture of the entire area

2.1.1 Human-Computer Interaction

The research of human-computer interaction is no longer the design of devices and psychological experiments of windows layouts, but evolutes to a new stage: intelligent interaction. One aspects is that computers should be able to accept audio and visual sensory inputs, and then make some kind of analysis and interpretation, and then provide intuitive feedbacks by synthesizing speech, video or actions. Fundamentally, besides speech recognition, computers should be able to recognize, interpret and understand human actions and behaviors from visual inputs.

2.1.2 Intelligent Environments

Intelligent environments, or smart environments, refer to some physical spaces that could automatically or intelligently react according to human activities. For example, when a person enters, the system could tell a people comes in and even identify who s/he is, and then turn on the lights. When the people sits on a sofa and points to a TV, the TV will be turned on. When s/he says "I want some news", the TV will be switched to a channel that is broadcasting news at that moment.

2.1.3 Multimedia

Multimedia is a vague term. Different people have different emphasis. We are particularly interested in the analysis of the content of multimedia. An interesting question we ask is what is inside this picture or what this video means, which involves a quite challenging task of image/video understanding. Many appealing applications have been proposed, but yet to be accomplished. When given just a photo of Sophie Marceau, without knowing her name, computers could search the Internet and get tones of her photos and movies. When you get tied of watching a long movie, computers could automatically summarize the movie in maybe five minutes.

2.1.4 Intelligent Robots

Robots have been giving quite good mechanical ability, but they are still machinery because they are neither able to see nor able to think. Honda has built a humanoid robot, ASIMO, which can walk like a human being. However, he is blind, dump and dull. We expect to see that ASIMO moves by itself.

2.2 Fundamental Research Issues

The fundamental research in image processing, computer vision, machine learning and pattern recognition is important part of the foundation of these application topics.

2.2.1 Image Processing and Computer Vision

Image processing is a quite board research area, not just filtering, compression, and enhancement. Besides, we are even interested in the question, "what is in images?", i.e., content analysis of visual inputs, which is part of the main task of computer vision. The study of computer vision could make possible such tasks as 3D reconstruction of scenes, motion capturing, and object recognition, which are crucial for even higher-level intelligence such as image and video understanding, and motion understanding.

2.2.2 Machine Learning and Pattern Recognition

Vision perception itself is an intelligent process, not just an imaging process. Through vision, human beings are able to perceive the lighting, color, texture, shape and motion of the outside world. The intelligence lies in the inference of such high-level concepts based on imaging. It is quite easy for human beings, but it is still very unclear how computers can achieve that level of intelligence. Recognition is one of the most fundamental problems for machine, i.e., recognizing a pre-stored pattern in new situations by comparing inputs with a set of templates or models. However, the problem is how to construct these templates or models. For example, what will be the appropriate templates to recognize faces even under different view directions or different lightings? The most challenging aspect for visual recognition lies in the fact that there are too many aspects that affects imaging, and it is impossible to model every aspects such as lighting and motion. So, people ask, "can computers 'learn' the

model from examples?" such that models could be learned implicitly, instead of constructed explicitly.

3 What Is Computer Vision?

According to my understanding, computer vision, basically, is to infer different factors such as camera model, lighting, color, texture, shape and motion that affect images and videos, from visual inputs. A rough structure of machine vision could be illustrated by Figure 2. In

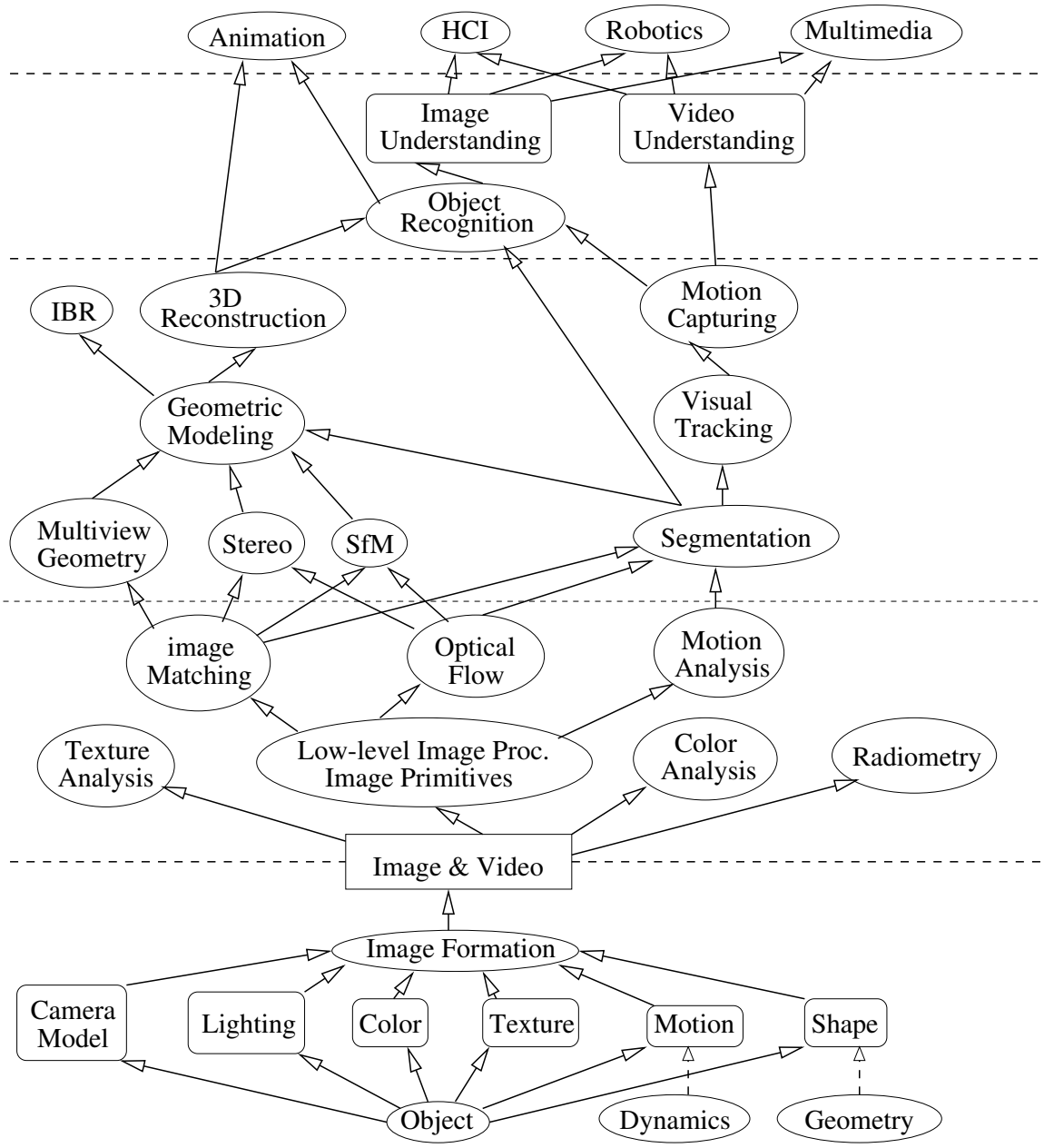

Figure 2: What is computer vision?

a word, computer vision is an inverse processing of the forward process of image formation and graphics. In this sense, as many people agree, vision is a much more challenging problem than computer graphics, because it is full of uncertainties.

3.1 Image Formation

Image formation studies the forward process of producing images and videos. It is an important research topic for both vision and graphics. To produce a real image, the nature of the visual sensors, i.e., cameras, should be studied. In terms of geometrical aspects of camera, people have been looking into pinhole cameras, cameras with lenses and even omnidirectional cameras. In terms of physical aspects, factors such as focal lengthes and dynamic ranges of CCD and CMOS cameras have been investigated.

Besides the imaging device, it is also important to study the factors from objects and scenes themselves, such as lighting, color, texture, motion and shape, which largely affect the appearance of images and video.

3.2 Low-level Image Processing

Low-level image processing is not vision, but the pre-processing steps for vision. The basic task is to extract fundamental image primitives for further processing, including edge detection, corner detection, filtering, and morphology, etc.

3.3 Low-level Vision

Based on low-level image processing, low-level vision tasks could be preformed, such as image matching, optical flow computation and motion analysis. Image matching basically is to find correspondences between two or more images. These images could be the same scene taken from different view points, or a moving scene taken by a fixed camera, or both. Constructing image correspondences is a fundamentally important problem in vision for both geometry recovery and motion recovery. Without exaggeration, image matching is part of the base for vision.

Optical flow is a kind of image observation of motion, but it is not the true motion. Since it only measure the optical changes in images, an aperture problem is unavoidable. But based on optical flows, camera motion or object motion could be estimated.

3.4 Middle-level Vision

There are two major aspects in middle-level vision: (1) inferring the geometry and (2) inferring the motion. These two aspects are not independent but highly related. A simple question is "can we estimate geometry based on just one image?". The answer is obvious. We need at least two images. They could be taken from two cameras or come from the motion of the scene.

Some fundamental parts of geometric vision include multiview geometry, stereo and structure from motion (SfM), which fulfill the step of *from 2D to 3D* by inferring 3D scene information from 2D images. Based on that, geometric modelling is to construct 3D models for

objects and scenes, such that 3D reconstruction and image-based rendering could be made possible.

Another task of middle-level vision is to answer the question "how the object moves". Firstly, we should know which areas in the images belong to the object, which is the task of image segmentation. Image segmentation has been a challenging fundamental problem in computer vision for decades. Segmentation could be based on spatial similarities and continuities. However, uncertainty can not be overcome for static image. When considering motion continuities, we hope the uncertainty of segmentation could be alleviated. On top of that is visual tracking and visual motion capturing, which estimate 2D and 3D motions, including deformable motions and articulated motions.

3.5 High-level Vision

High-level vision is to infer the semantics, for example, object recognition and scene understanding. A challenging question in many decades is that how to achieve invariant recognition, i.e., recognize 3D object from different view directions. There have been two approaches for recognition: model-based recognition and learning-based recognition. It is noticed that there was a spiral development of these two approaches in history.

Even higher level vision is image understanding and video understanding. We are interested in answering questions like "Is there a car in the image?" or "Is this video a drama or an action?", or "Is the person in the video jumping?" Based on the answers of these questions, we should be able to fulfill different tasks in intelligent human-computer interaction, intelligent robots, smart environment and content-based multimedia.

4 What Is This Course Going To Cover?

This course is going to cover most fundamental aspects in computer vision and machine learning. Details are available in course syllabus.