

Computer Vision

Computer vision tasks include methods for acquiring, processing, analyzing, and understanding digital images, and extraction of high-dimensional data from the real world in order to produce numerical or symbolic information, e.g. in the form of decisions.[1][2][3][4] "Understanding" in this context signifies the transformation of visual images (the input to the retina) into descriptions of the world that make sense to thought processes and can elicit appropriate action. This image understanding can be seen as the disentangling of symbolic information from image data using models constructed with the aid of geometry, physics, statistics, and learning theory.

The scientific discipline of computer vision is concerned with the theory behind artificial systems that extract information from images. Image data can take many forms, such as video sequences, views from multiple cameras, multi-dimensional data from a 3D scanner, 3D point clouds from LiDaR sensors, or medical scanning devices. The technological discipline of computer vision seeks to apply its theories and models to the construction of computer vision systems.

Subdisciplines of computer vision include scene reconstruction, object detection, event detection, activity recognition, video tracking, object recognition, 3D pose estimation, learning, indexing, motion estimation, visual servoing, 3D scene modeling, and image restoration.

Computer vision is an interdisciplinary field that deals with how computers can be made

to gain high-level understanding from digital images or videos. From the perspective of engineering, it seeks to automate tasks that the human visual system can do.[5][6][7] "Computer vision is concerned with the automatic extraction, analysis, and understanding of useful information from a single image or a sequence of images. It involves the development of a theoretical and algorithmic basis to achieve automatic visual understanding." [8] As a scientific discipline, computer vision is concerned with the theory behind artificial systems that extract information from images. The image data can take many forms, such as video sequences, views from multiple cameras, or multi-dimensional data from a medical scanner.[9] As a technological discipline, computer vision seeks to apply its theories and models for the construction of computer vision systems. Machine vision refers to a systems engineering discipline, especially in the context of factory automation. In more recent times, the terms computer vision and machine vision have converged to a greater degree.[10]:13

In the late 1960s, computer vision began at universities that were pioneering artificial intelligence. It was meant to mimic the human visual system as a stepping stone to endowing robots with intelligent behavior.[11] In 1966, it was believed that this could be achieved through an undergraduate summer project,[12] by attaching a camera to a computer and having it "describe what it saw".[13][14]

What distinguished computer vision from the prevalent field of digital image processing at that time was a desire to extract three-dimensional structure from images with the goal of achieving full scene understanding. Studies in the 1970s formed the early foundations for many of the computer vision algorithms that exist today, including extraction of edges from images, labeling of lines, non-polyhedral and polyhedral

modeling, representation of objects as interconnections of smaller structures, optical flow, and motion estimation.[11]

The next decade saw studies based on more rigorous mathematical analysis and quantitative aspects of computer vision. These include the concept of scale-space, the inference of shape from various cues such as shading, texture and focus, and contour models known as snakes. Researchers also realized that many of these mathematical concepts could be treated within the same optimization framework as regularization and Markov random fields.[15]

By the 1990s, some of the previous research topics became more active than others. Research in projective 3-D reconstructions led to better understanding of camera calibration. With the advent of optimization methods for camera calibration, it was realized that a lot of the ideas were already explored in bundle adjustment theory from the field of photogrammetry. This led to methods for sparse 3-D reconstructions of scenes from multiple images. Progress was made on the dense stereo correspondence problem and further multi-view stereo techniques. At the same time, variations of graph cut were used to solve image segmentation. This decade also marked the first time statistical learning techniques were used in practice to recognize faces in images (see Eigenface). Toward the end of the 1990s, a significant change came about with the increased interaction between the fields of computer graphics and computer vision. This included image-based rendering, image morphing, view interpolation, panoramic image stitching and early light-field rendering.[11]

Recent work has seen the resurgence of feature-based methods used in conjunction with machine learning techniques and complex optimization frameworks.[16][17]

The advancement of Deep Learning techniques has brought further life to the field of computer vision. The accuracy of deep learning algorithms on several benchmark computer vision data sets for tasks ranging from classification,[18] segmentation and optical flow has surpassed prior methods. [citation needed][19]

Solid-state physics is another field that is closely related to computer vision. Most computer vision systems rely on image sensors, which detect electromagnetic radiation, which is typically in the form of either visible, infrared or ultraviolet light. The sensors are designed using quantum physics. The process by which light interacts with surfaces is explained using physics. Physics explains the behavior of optics which are a core part of most imaging systems. Sophisticated image sensors even require quantum mechanics to provide a complete understanding of the image formation process.[11] Also, various measurement problems in physics can be addressed using computer vision, for example, motion in fluids.

Neurobiology has greatly influenced the development of computer vision algorithms. Over the last century, there has been an extensive study of eyes, neurons, and brain structures devoted to the processing of visual stimuli in both humans and various animals. This has led to a coarse yet convoluted description of how natural vision systems operate in order to solve certain vision-related tasks. These results have led to a sub-field within computer vision where artificial systems are designed to mimic the processing and behavior of biological systems at different levels of complexity. Also, some of the learning-based methods developed within computer vision (e.g. neural net and deep learning based image and feature analysis and classification) have their background in neurobiology. The Neocognitron, a neural network developed in the

1970s by Kunihiro Fukushima, is an early example of computer vision taking direct inspiration from neurobiology, specifically the primary visual cortex.

Some strands of computer vision research are closely related to the study of biological vision—indeed, just as many strands of AI research are closely tied with research into human intelligence and the use of stored knowledge to interpret, integrate, and utilize visual information. The field of biological vision studies and models the physiological processes behind visual perception in humans and other animals. Computer vision, on the other hand, develops and describes the algorithms implemented in software and hardware behind artificial vision systems. An interdisciplinary exchange between biological and computer vision has proven fruitful for both fields.[21]

Yet another field related to computer vision is signal processing. Many methods for processing one-variable signals, typically temporal signals, can be extended in a natural way to the processing of two-variable signals or multi-variable signals in computer vision. However, because of the specific nature of images, there are many methods developed within computer vision that have no counterpart in the processing of one-variable signals. Together with the multi-dimensionality of the signal, this defines a subfield in signal processing as a part of computer vision.

Robot navigation sometimes deals with autonomous path planning or deliberation for robotic systems to navigate through an environment.[22] A detailed understanding of these environments is required to navigate through them. Information about the environment could be provided by a computer vision system, acting as a vision sensor

and providing high-level information about the environment and the robot

Besides the above-mentioned views on computer vision, many of the related research topics can also be studied from a purely mathematical point of view. For example, many methods in computer vision are based on statistics, optimization or geometry. Finally, a significant part of the field is devoted to the implementation aspect of computer vision; how existing methods can be realized in various combinations of software and hardware, or how these methods can be modified in order to gain processing speed without losing too much performance. Computer vision is also used in fashion eCommerce, inventory management, patent search, furniture, and the beauty industry.[23]

The fields most closely related to computer vision are image processing, image analysis and machine vision. There is a significant overlap in the range of techniques and applications that these cover. This implies that the basic techniques that are used and developed in these fields are similar, something which can be interpreted as there is only one field with different names. On the other hand, it appears to be necessary for research groups, scientific journals, conferences, and companies to present or market themselves as belonging specifically to one of these fields and, hence, various characterizations which distinguish each of the fields from the others have been presented. In image processing, the input is an image and the output is an image as well, whereas in computer vision, an image or a video is taken as an input and the output could be an enhanced image, an understanding of the content of an image or even behavior of a computer system based on such understanding.

Computer graphics produces image data from 3D models, and computer vision often produces 3D models from image data.[24] There is also a trend towards a combination of the two disciplines, e.g., as explored in augmented reality.

The following characterizations appear relevant but should not be taken as universally accepted:

Photogrammetry also overlaps with computer vision, e.g., stereophotogrammetry vs. computer stereo vision.

Applications range from tasks such as industrial machine vision systems which, say, inspect bottles speeding by on a production line, to research into artificial intelligence and computers or robots that can comprehend the world around them. The computer vision and machine vision fields have significant overlap. Computer vision covers the core technology of automated image analysis which is used in many fields. Machine vision usually refers to a process of combining automated image analysis with other methods and technologies to provide automated inspection and robot guidance in industrial applications. In many computer-vision applications, computers are pre-programmed to solve a particular task, but methods based on learning are now becoming increasingly common. Examples of applications of computer vision include systems for:

One of the most prominent application fields is medical computer vision, or medical image processing, characterized by the extraction of information from image data to diagnose a patient. An example of this is the detection of tumours, arteriosclerosis or other malign changes, and a variety of dental pathologies; measurements of organ dimensions, blood flow, etc. are another example. It also supports medical research by providing new information: e.g., about the structure of the brain or the quality of medical treatments. Applications of computer vision in the medical area also include enhancement of images interpreted by humans—ultrasonic images or X-ray images, for example—to reduce the influence of noise.

A second application area in computer vision is in industry, sometimes called machine vision, where information is extracted for the purpose of supporting a production process. One example is quality control where details or final products are being automatically inspected in order to find defects. One of the most prevalent fields for such inspection is the Wafer industry in which every single Wafer is being measured and inspected for inaccuracies or defects to prevent a computer chip from coming to market in an unusable manner. Another example is a measurement of the position and orientation of details to be picked up by a robot arm. Machine vision is also heavily used in the agricultural processes to remove undesirable foodstuff from bulk material, a process called optical sorting.[32]

Military applications are probably one of the largest areas of computer vision[citation needed]. The obvious examples are the detection of enemy soldiers or vehicles and missile guidance. More advanced systems for missile guidance send the missile to an area rather than a specific target, and target selection is made when the missile

reaches the area based on locally acquired image data. Modern military concepts, such as "battlefield awareness", imply that various sensors, including image sensors, provide a rich set of information about a combat scene that can be used to support strategic decisions. In this case, automatic processing of the data is used to reduce complexity and to fuse information from multiple sensors to increase reliability.

One of the newer application areas is autonomous vehicles, which include submersibles, land-based vehicles (small robots with wheels, cars, or trucks), aerial vehicles, and unmanned aerial vehicles (UAV). The level of autonomy ranges from fully autonomous (unmanned) vehicles to vehicles where computer-vision-based systems support a driver or a pilot in various situations. Fully autonomous vehicles typically use computer vision for navigation, e.g., for knowing where they are or mapping their environment (SLAM), for detecting obstacles. It can also be used for detecting certain task-specific events, e.g., a UAV looking for forest fires. Examples of supporting systems are obstacle warning systems in cars, cameras and LiDAR sensors in vehicles, and systems for autonomous landing of aircraft. Several car manufacturers have demonstrated systems for autonomous driving of cars. There are ample examples of military autonomous vehicles ranging from advanced missiles to UAVs for reconnaissance missions or missile guidance. Space exploration is already being made with autonomous vehicles using computer vision, e.g., NASA's Curiosity and CNSA's Yutu-2 rover.

Materials such as rubber and silicon are being used to create sensors that allow for applications such as detecting microundulations and calibrating robotic hands. Rubber can be used in order to create a mold that can be placed over a finger, inside of this

mold would be multiple strain gauges. The finger mold and sensors could then be placed on top of a small sheet of rubber containing an array of rubber pins. A user can then wear the finger mold and trace a surface. A computer can then read the data from the strain gauges and measure if one or more of the pins are being pushed upward. If a pin is being pushed upward then the computer can recognize this as an imperfection in the surface. This sort of technology is useful in order to receive accurate data on imperfections on a very large surface.[33] Another variation of this finger mold sensor are sensors that contain a camera suspended in silicon. The silicon forms a dome around the outside of the camera and embedded in the silicon are point markers that are equally spaced. These cameras can then be placed on devices such as robotic hands in order to allow the computer to receive highly accurate tactile data.[34]

Other application areas include:

Each of the application areas described above employ a range of computer vision tasks; more or less well-defined measurement problems or processing problems, which can be solved using a variety of methods. Some examples of typical computer vision tasks are presented below.

Computer vision tasks include methods for acquiring, processing, analyzing and understanding digital images, and extraction of high-dimensional data from the real world in order to produce numerical or symbolic information, e.g., in the forms of decisions.[1][2][3][4] Understanding in this context means the transformation of visual images (the input of the retina) into descriptions of the world that can interface with

other thought processes and elicit appropriate action. This image understanding can be seen as the disentangling of symbolic information from image data using models constructed with the aid of geometry, physics, statistics, and learning theory.[39]

The classical problem in computer vision, image processing, and machine vision is that of determining whether or not the image data contains some specific object, feature, or activity. Different varieties of recognition problem are described in the literature.[40]

Currently, the best algorithms for such tasks are based on convolutional neural networks. An illustration of their capabilities is given by the ImageNet Large Scale Visual Recognition Challenge; this is a benchmark in object classification and detection, with millions of images and 1000 object classes used in the competition.[41] Performance of convolutional neural networks on the ImageNet tests is now close to that of humans.[41] The best algorithms still struggle with objects that are small or thin, such as a small ant on the stem of a flower or a person holding a quill in their hand. They also have trouble with images that have been distorted with filters (an increasingly common phenomenon with modern digital cameras). By contrast, those kinds of images rarely trouble humans. Humans, however, tend to have trouble with other issues. For example, they are not good at classifying objects into fine-grained classes, such as the particular breed of dog or species of bird, whereas convolutional neural networks handle this with ease.[citation needed]

Several specialized tasks based on recognition exist, such as:

Several tasks relate to motion estimation, where an image sequence is processed to produce an estimate of the velocity either at each points in the image or in the 3D scene or even of the camera that produces the images. Examples of such tasks are:

Given one or (typically) more images of a scene, or a video, scene reconstruction aims at computing a 3D model of the scene. In the simplest case, the model can be a set of 3D points. More sophisticated methods produce a complete 3D surface model. The advent of 3D imaging not requiring motion or scanning, and related processing algorithms is enabling rapid advances in this field. Grid-based 3D sensing can be used to acquire 3D images from multiple angles. Algorithms are now available to stitch multiple 3D images together into point clouds and 3D models.[24]

Image restoration comes into the picture when the original image is degraded or damaged due to some external factors like lens wrong positioning, transmission interference, low lighting or motion blurs, etc., which is referred to as noise. When the images are degraded or damaged, the information to be extracted from them also gets damaged. Therefore, we need to recover or restore the image as it was intended to be. The aim of image restoration is the removal of noise (sensor noise, motion blur, etc.) from images. The simplest possible approach for noise removal is various types of filters, such as low-pass filters or median filters. More sophisticated methods assume a model of how the local image structures look to distinguish them from noise. By first analyzing the image data in terms of the local image structures, such as lines or edges, and then controlling the filtering based on local information from the analysis step, a better level of noise removal is usually obtained compared to the simpler approaches.

An example in this field is inpainting.

The organization of a computer vision system is highly application-dependent. Some systems are stand-alone applications that solve a specific measurement or detection problem, while others constitute a sub-system of a larger design which, for example, also contains sub-systems for control of mechanical actuators, planning, information databases, man-machine interfaces, etc. The specific implementation of a computer vision system also depends on whether its functionality is pre-specified or if some part of it can be learned or modified during operation. Many functions are unique to the application. There are, however, typical functions that are found in many computer vision systems.

Image-understanding systems (IUS) include three levels of abstraction as follows: low level includes image primitives such as edges, texture elements, or regions; intermediate level includes boundaries, surfaces and volumes; and high level includes objects, scenes, or events. Many of these requirements are entirely topics for further research.

The representational requirements in the designing of IUS for these levels are: representation of prototypical concepts, concept organization, spatial knowledge, temporal knowledge, scaling, and description by comparison and differentiation.

While inference refers to the process of deriving new, not explicitly represented facts from currently known facts, control refers to the process that selects which of the many inference, search, and matching techniques should be applied at a particular stage of processing. Inference and control requirements for IUS are: search and hypothesis activation, matching and hypothesis testing, generation and use of expectations, change and focus of attention, certainty and strength of belief, inference and goal satisfaction.[48]

There are many kinds of computer vision systems; however, all of them contain these basic elements: a power source, at least one image acquisition device (camera, ccd, etc.), a processor, and control and communication cables or some kind of wireless interconnection mechanism. In addition, a practical vision system contains software, as well as a display in order to monitor the system. Vision systems for inner spaces, as most industrial ones, contain an illumination system and may be placed in a controlled environment. Furthermore, a completed system includes many accessories, such as camera supports, cables, and connectors.

Most computer vision systems use visible-light cameras passively viewing a scene at frame rates of at most 60 frames per second (usually far slower).

A few computer vision systems use image-acquisition hardware with active illumination or something other than visible light or both, such as structured-light 3D scanners, thermographic cameras, hyperspectral imagers, radar imaging, lidar scanners,

magnetic resonance images, side-scan sonar, synthetic aperture sonar, etc. Such hardware captures "images" that are then processed often using the same computer vision algorithms used to process visible-light images.

While traditional broadcast and consumer video systems operate at a rate of 30 frames per second, advances in digital signal processing and consumer graphics hardware has made high-speed image acquisition, processing, and display possible for real-time systems on the order of hundreds to thousands of frames per second. For applications in robotics, fast, real-time video systems are critically important and often can simplify the processing needed for certain algorithms. When combined with a high-speed projector, fast image acquisition allows 3D measurement and feature tracking to be realized.[49]

Egocentric vision systems are composed of a wearable camera that automatically take pictures from a first-person perspective.

As of 2016, vision processing units are emerging as a new class of processors to complement CPUs and graphics processing units (GPUs) in this role.[50]