A brief History of Computer Vision

The history of computer vision began in 1963 with the PhD dissertation of Claude Shannon's student Lawrence Roberts at MIT ("Machine Perception of Three Dimensional Solids," 1963). This was the same Roberts who in 1966 would launch the Arpanet project, later renamed Internet.

Face recognition was pioneered by Woody Bledsoe at his Palo Alto consulting firm Panoramic Research, that mainly worked on military projects ("The Model Method in Facial Recognition", 1964). In 1966 he moved to SRI Intl where he trained Peter Hart, who later worked on Shakey the Robot (and who in 1973 coauthored the textbook "Pattern Classification and Scene Analysis" with Richard Duda).

In 1968 Adolfo Guzman at MIT built programs to detect the constituent objects of a scene ("Computer Recognition of Three-Dimensional Objects in a Visual Scene", 1968). Max Clowes ("On Seeing Things", 1971) at the University of Sussex and David Huffman at UC Santa Cruz ("Impossible Objects as Nonsense Sentences", 1971) independently discovered methods to interpret pictures of polyhedra (solids such as cubes and pyramids), and Alan Mackworth at the University of Sussex developed a program to interpret line drawings as polyhedral scenes ("Interpreting Pictures of Polyhedral Scenes", 1973). Computer vision was mostly about recognizing objects within a picture, and initially the prevailing method was to compare the regions of the picture with templates of typical objects. Martin Fischler and Robert Elschlager at Lockhead's Palo Alto Research Laboratory expanded this method with "stretchable templates" ("The Representation and Matching of Pictorial Structures", 1973). Takeo Kanade graduated from Kyoto University in 1973 with the world's first automated face recognition system ("Picture Processing System by Computer Complex and Recognition of Human Faces", 1973).

By coincidence or not, both Huffman and Kanade were pioneers of mathematical origami, that consists in folding a two-dimensional piece of paper into a three-dimensional object. For example, Takeo Kanade, now at Carnegie Mellon University, published the article "A Theory of Origami World" (1980).

Then David Marr's epochal posthumous book "Vision" (1982) was published. It summarized his research at MIT that the mind understands a scene through a three stage process: a primal two-dimensional sketch that contains the basic components of the scene; a 2.5 dimension sketch of the scene that also contains depth; an the final three-dimensional model. Computer vision became a popular subject. One limitation was that all the classic algorithms only dealt with straight lines.

Jitendra Malik at Stanford ("Interpreting Line Drawings of Curved Objects", 1985) was one of the young scientists who studied how to deal with curved objects. In 1987 Lawrence Sirovich and Michael Kirby, mathematicians at Brown University, used principle component analysis (i.e., linear algebra) to transform images of faces into mathematical vectors called "eigenfaces" ("Low-dimensional Procedure for the Characterization of Human Faces", 1987). This method constituted the basis of the "sliding window" approach of Matthew Turk and Alex Pentland at MIT ("Eigenfaces for Face Detection/Recognition", 1991). In 1987 another cognitive scientist, Irving Biederman at State University of New York in Buffalo, published an influential article to explain how we recognize objects, arguing that objects can be broken down into basic geometric solids called "geons" ("Recognition-by-components, a theory of human image understanding", 1987).

After graduation Malik moved to UC Berkeley, where he founded an important school of computer vision, and one of his students was Pietro Perona, who, in turn, moved to the California Institute of Technology (CalTech) and built another important group in computer vision that over the years refined Perona's "constellation models" for object detection: Thomas Leung ("Finding Faces in Cluttered Scenes Using Labeled Random Graph Matching", 1995), Michael Burl ("Localization via Shape Statistics", 1995), Markus Weber ("Unsupervised Learning of Models for Recognition", 2000), Fei-Fei Li ("A Bayesian Approach to Unsupervised One-Shot learning of Object Categories", 2003), Rob Fergus ("Object Class Recognition by Unsupervised Scale-Invariant Learning", 2003), etc. It was here that in 2003 Fei-Fei Li built the dataset of images Caltech 101 that later evolved into ImageNet.