Teaching Image Analysis at DIKU

Peter Johansen

Datalogisk Institut, University of Copenhagen (DIKU)
Universitetsparken 1, 2100 Copenhagen, Denmark
peterjo@diku.dk

Abstract. The early development of computer vision at Department of Computer Science at University of Copenhagen (DIKU) is briefly described. The different disciplines in computer vision are introduced, and the principles for teaching two courses, an image analysis course, and a robot lab class are outlined.

Keywords: Computer vision, image analysis, pattern recognition, robot vision, University of Copenhagen

1 Introduction

Teaching image analysis at DIKU began in 1979 when the author returned from a one-year sabbatical at the University of Hawaii at Manoa. The university had astronomy as an area of excellence. Hawaii has the best "seeing" on earth, meaning that the image of a point source like a star in a telescope is subject to the least atmospheric disturbance anywhere on earth. The reason of course is that the shortest distance from the Islands of Hawaii to the nearest continent is about 3,000 miles, resulting in minimal air pollution. At that time, a telescope was built with a very sensitive digital sensor, perhaps the first digital sensor on an astronomical telescope. A project was underway as a joint project between USA, Switzerland, and the USSR. The purpose was to classify the surface of the moon from multispectral images and if possible, to detect mineral deposits.

2 Signal Analysis

At that time, the author had no experience in digital analysis of images, but found the concept so intriguing that he decided to embark upon analysis of images by programming a computer. As a first step he signed up for a course in optics - as it turned out not a very useful beginning, then starting to familiarize himself with digital signal analysis. As it later dawned on him, signal analysis was mainly concerned with linear processing of 1D time series, and in particular, applications of the Fourier transform. The fast Fourier transform is a beautiful algorithm, and it was tempting to start a voyage of discovery into signal analysis of 2D signals. Filtering an image of a

star in order to undo the damage of atmospheric distortion seemed a good starting point. This way imaging defects by an optical sensor could be minimized. It was a laudable endeavor, but not central to image analysis. Of course, information in an image should be as credible as possible, but the real problem was what to do with an image once it had been massaged to provide optimal information. Furthermore, linear methods in signal processing originated in analysis of electrical circuits, and concepts and terminology was characterized by electrical engineering. No doubt, signal processing was important but, as the author was to discover, statistical analysis was also central. An important initial application of image analysis was processing of microscope images in medical research. Immense gains could be made by automating manual visual classification of slides. In this application, so-called mathematical morphology was used very early. This approach is complementary to signal processing with origin in electrical engineering. Mathematical morphology is non-linear, and this extension of classical concepts in signal processing was very much needed in image analysis. Images could be cleaned in a much better way, and non-linear filters opened up for new applications.

3 Pattern Recognition

Astronomers used spectral images. An area was imaged in different bands of the electromagnetical spectrum. This way each location (picture element or pixel for short) was associated numbers expressing the reflected light at the different bands. Mathematically, a digital image was an array of multiples of numbers, not just of single numbers. The basic systematic way to classify an area was to group number multiples into groups of similar multiples. This is the basic problem in multispectral pattern recognition. At the Technical University of Copenhagen (DTU) in Denmark, satellite images were being analyzed this way. Images of Mestersvig in Greenland were investigated for possible mineral deposits. For a mathematician turned computer scientist this application opened up a bewildering set of problems. A pixel did not contain precise information but was the filtered response of a sensor by its point spread function. A pixel pertained to a point location on the ground, and was influenced by neighboring points. In an application, geologists would know the minerals at selected spots. Thus, in the image one could designate areas with known deposits as *training areas*. Thus, all multiples at the training areas were known to reflect a certain mineral. This meant that educated guesses could be made for minerals at unvisited locations by computing the training area with most similar multiples. "Most similar" had to be defined, and here mathematical statistics was the necessary tool. One describes the distribution of multiples as a multidimensional probability density distribution, and one estimates its parameters from information obtained from relevant training areas. Even, if no training areas were known, it was often possible to group similar multiples into classes, and, this way segment the image into different classes. The discipline is called unsupervised pattern recognition and supervised, when training areas are known.

4 Image Scale

If images were taken from a different height, what was the relation between the pictures? This problem was important, when information from different sensors should be combined. In 1985, Ruzena Bajczy visited Copenhagen University, and in an invited lecture, she mentioned the work of Andrew Witkin on Scale space. This meant a mathematically fruitful way of defining the effect of varying spatial aperture of a sensor. When an image is analyzed, a description must combine an overall description with small significant details: different scales are necessary for interpretation. As an example, shadow boundaries are sometimes more and sometimes less sharp. In order to locate a less sharp boundary one has to analyze a larger environment of the image, than when the boundary is well defined. The Perona-Malik filter uses a small aperture at image points with large gradient and a larger aperture when the image gradient is smaller. In this manner, different points in an image are processed at different scales.

5 Human Visual Perception

A human makes the interpretation of an image, be it a medical doctor screening an X-ray image or a biologist studying a histological microscope slide. Thus, even more complexity is introduced into automatic analysis, since one has to understand and preferably model human visual perception. Not only that, but the interpretation is conditioned on previous experience. A study of human visual perception and memory was called for. Fortunately, the neurology of visual perception was much studied. The Hubel and Wiesel discovery of brain areas with specialized and specific visual functions initiated an avalanche of new discoveries of how the brain decodes images. The book *Vision* by David Marr from 1982 had international impact, and hid DIKU as a bomb. He treated biological vision from a computational point of view. As an example, the function of a ganglion cell in the retina can be described as a linear filter, and ganglion cells come with different apertures. This permits the retina to analyze an image at different scales.

6 Information Theory

A fundamental question presents itself: Interpretation of image data depends on the application. In general, one wants to model the phenomenon that one wants to understand. Then one determines the parameters of the model. A computer scientist would ask, which program (model) would generate the data that one observes. Without constraints, this question has many answers, including the trivial one: The model is a listing of the data. The question has to be rephrased: Find the program given reasonable restrictions. A possible restriction is to find a simple model. This way the model would describe one's understanding of the data. For example, Kepler's equations for motion of the planets express our understanding of planet motion.

Likewise, one would like to obtain a description of a 3D scene from a 2D image, and one wants a simple description. The vocabulary of the description is elements of the language for building the model. One would like to obtain a description of the pathology of a medical image in concepts relevant for the illness investigated. The precise formulation of how to find the simplest model for a set of data was first made by Kolmogorov, when he stated that the complexity of a set of data is the number of bits needed to represent it. Thus, one is forced into information theory in order to understand how data can be represented using the fewest number of bits. This approach provides a mathematical definition of parsimony, also called Occam's razor: One describes a phenomenon by the simplest explanation. Obviously, the shortest program to generate a set of data is the optimal way to compress a data set. We need to understand data compression as well. Jorma Rissanen made statistical models explicit, stating in his *Minimum Description Length Principle (MDL)* that the best model is the one that possesses the least joint complexity of the model and the data that it generates. Here the complexity of a model is the information needed to describe its parameters.

7 Active Vision

In 1983, Ruzena Bajczy gave a lecture at the Scandinavian conference on image analysis. She did research on robotics that is programming motion of a machine based on visual input. Until then, at DIKU we had concentrated on signal analysis including the fast Fourier transform as the basic theory of image understanding. Ruzena Bajczy introduced the concept of active vision. How does a device – a robot – decide where to direct its sensor? Analysis of its sensors, in particular its images, should tell it where to direct its camera next in order to obtain maximum information about the object to describe or to avoid. This introduces the perception-action cycle into the pêle-mêle of knowledge needed for image analysis! One wants that information from an image that is sufficient to compute the next position of the sensor – and for a movable robot, the robot's actuators. This threw us into description of the 3D world from image inputs. We started by two fundamental investigations: Generating depth from two images of a stationary object taken from different angles, and computing the motion of a moving object from a stationary camera. For this purpose, we built equipment that made it possible to program the motion of an object and to locate a camera at different position. We settled on a turntable that we could move along a rail. On the table, we could position either the object, or the camera. The equipment permitted us to quantify the precision of our computation of depth and motion, since we could set up an experiment with known parameters for position and motion. Four PhD theses resulted from this early development [1–4].

Later we changed to small movable robots, that is platforms on which were mounted a camera and a computer that controlled actuators for the platform's wheels. The prototype project would be to program the robots as a swarm according to the rules used by Reynolds' "boids": Agents should move at the same speed, close to each other, but not too close. Furthermore, we mounted communication equipment on the platforms, thus permitting the robots to communicate.

A robot interacting with its environment is a dynamical system. The theory of dynamical systems predicts that for certain values of system parameters, infinitesimal changes of parameter values could make the system change behavior from well behaved to chaotic. A challenge, which we never came to realize in the lab, would be to implement the transition from period doubling to unpredictability.

8 Teaching Computer Vision

How is one to teach this enormous complexity inside a computer science curriculum? One approach is to teach the basic disciplines first, followed by their application to image analysis. We discovered by experience that, since the involved disciplines are so many, it is difficult to decide if one should select a few or teach a little of each. Furthermore, we found that focus is easily lost over too long a time span, that is, when too many (≥ 2) courses are needed. At DIKU a successful lecture series has resulted, which is built on a set of weekly exercises. Each exercise introduces necessary theory in connection with an application. An exercise is a program that should be modified or extended. This way theory is seen as necessary for results, theory and application were seen as depending on each other, the students' attention were caught and their creativity unleashed. The same concept was used for image analysis and for active robot vision.

This setup functions well in a computer science education, since the students would be proficient in programming in advance of classes, permitting focus on the subject matter, and permitting more advanced applications.

Two classes have crystallized from our efforts: An image analysis class, and a robot lab class. Both classes were built on the same concept: In advance a set of weekly exercises were programmed as small applications. The students were to apply the programs and to do application motivated modifications. The lectures presented the theory in advance of the exercise. This way theoretical concepts and applications were always together. Advanced theory would always be demonstrated with practical applications. It is surprising how advanced theory a simple application needs, if it has to be well understood, and how advanced applications can be build from little well understood theory. The experience was that if a firm ground could be made for a simple and convincing application, the students themselves would extend the theory and the application.

Exercises were carefully designed, and constituted the backbone of the classes. The exercises were non-trivial. For instance, the first exercise in image analysis was to segment a grey level image. This is possible with threshold logic for a simple image; the application was to compute the total value of coins on an image. For an image analysis expert trivial, but for a novice without experience in image analysis, a surprisingly large set of concepts are introduced by a single exercise. Then segmentation of a color image followed, then image segmentation, and motion analysis.

The robot lab class would start by implementing avoidance behaviour by fuzzy control. Then students would program self localization using particle filtering. The

final exercise would be on robot cooperation, but so far we have not yet succeeded to program the robots to behave like one of Reynolds' swarms.

9 Literature

Below are mentioned a selection of textbooks and papers that have influenced the development of computer vision at DIKU in the period 1980–2008. Edda Sveinsdottir's research in computer tomography can be found in [5, 6]. Image analysis and pattern recognition was typically taught using a background textbook [7] supplemented by lecture notes. The book by Bishop [8] on pattern recognition was very welcome, when is appeared in 2006. Its flavour is mathematical statistics, and computer science students would need some mathematical background to fully profit from the book. Andrew Witkin's paper on scale space [9] introduced us to the idea of analysing a signal at different scales. Koenderink [10] extended scale from 1D signals to 2D signals, that is to images. Its application to image analysis is well explained by Lindeberg in his paper [11], in which he gives instructive examples. Information theory was introduced by Claude Shannon [12], and his initial text is still very readable. The minimum description length (MDL) principle is well described by Hansen and Yu [13]. Kolmogorov complexity and its applications are explained in the textbook by Li and Vitanýi [14]. Active vision introduced by Bajczy is well presented in a paper by K. Pahlavan and T. Uhlin and J.-O. Eklundh [15]. The elegant emerging flocking behaviour of locally acting agents is presented in a paper by Reynolds [16]. Devaney [17] is an introductory textbook on dynamical systems from a mathematical point of view. The elegant and powerful technique of particle filtering can be found in the paper by Isard and Blake [18].

References

1. Arnspang, J.: Local Differential Kinematics in Surface Vision. PhD thesis, DIKU, University of Copenhagen (1987)
2. Olsen, I.: The Design and Analysis of a Feature Based Stereo Algorithm. PhD thesis, DIKU, University of Copenhagen (1988)
3. Henriksen, K.: Projective Geometry and Straight Lines in Computational Vision. PhD thesis, DIKU, University of Copenhagen (1990)
4. Nielsen, M.: From Paradigm to Algorithms in Computer vision. PhD thesis, DIKU, University of Copenhagen (1995)

5. Høedt-Rasmussen, K., Sveinsdottir, E., Lassen, N.A.: Regional Cerebral flow in Man Determined ny Intra-arterial Injection of Radioactive Inert Gas. Circulation Research XVIII, no. 3, 237–247 (1966)
6. Sveinsdottir, E., Larsen, B., Rommer, P., Lassen, N.A.: A Multidetector Scintillation Camera with 254 Channels. The Journal of Nuclear Medicine 18, no. 2, 168–174 (1977)
7. Sonka, M., Hlavac, V., Boyle, R.: Image Processing, Analysis, and Machine Vision. Brooks/Cole Thompson (1999)
8. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer (2006)
9. Witkin, A.: Scale Space Filtering. In: Proceedings of the International Joint Conference on Artificial Intelligence. Karlsruhe, pp. 1019–1021 (1993)
10. Koenderink, J.J.: The Structure of Images. Biological Cybernetics 50, no. 5, 363–370 (1984)
11. Lindeberg, T.: Feature Detection with Automatic Scale Selection. In: International Journal of Computer Vision. 30, no. 2, 79–116 (1998)
12. Shannon, C.E., Weaver, W.: The Mathematical Theory of Communication. Illini Books edition (1963)
13. Hansen, M.H., Yu, B.: Model Selection and the Principle of Minimum Description Length. Journal of the American Statistical Association 96, no. 454, 746–774 (2001)
14. Li, M., Vitanýi, P.: An Introduction to Kolmogorov Complexity and its Applications. North Holland (1993)
15. Pahlavan, K., Uhlin, T., Eklundh, J.-O.: Active Vision as a Methodology. In: Aloimonos, Y. (ed.) Purposive Vision, Chapter 1. Lawrence Erlbaum (1994)
16. Reynolds, C.: Flocks, Herds and Schools: A Distributed Behavioral Model. In: SIGGRAPH '87: Proceedings of the 14th Annual Computer Graphics and Interactive Techniques (Association for Computing Machinery), pp. 25–34 (1987)
17. Devaney, R.L.: An Introduction to Chaotic Dynamical Systems. Addison Wesley, (1989)
18. Isard, M., Blake, A.: Condensation – Conditional Density Propagation for Visual Tracking. International Journal of Computer Vision 29, no. 2, 5–28 (1998)