Image Understanding as a Second Course in AI: Preparing Students for Research

Roxanne Canosa
Rochester Institute of Technology
134 Lomb Memorial Drive
Rochester, New York 14613
(585) 475-5810

rlc@cs.rit.edu

ABSTRACT

This paper describes the development and structure of a second course in artificial intelligence that was developed to meet the needs of upper-division undergraduate and graduate computer science and computer engineering students. These students already have a background in either computer vision or artificial intelligence, and desire to apply that knowledge to the design of algorithms that are able to automate the process of extracting semantic content from either static or dynamic imagery. Theory and methodology from diverse areas were incorporated into the course, including techniques from image processing, statistical pattern recognition, knowledge representation, multivariate analysis, cognitive modeling, and probabilistic inference. Students read selected current literature from the field, took turns presenting the selected literature to the class, and participated in discussions about the literature. Programming projects were required of all students, and in addition, graduate students were required to propose, design, implement, and defend an image understanding project of their own choosing. The course served as preparation for and an incubator of an active research group.

Categories and Subject Descriptors

I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding – *perceptual reasoning.*

General Terms

None.

Keywords

Course design, artificial intelligence education, computer vision, student projects.

1. INTRODUCTION

The Middle States Commission on Higher Education has set forth guidelines for developing research and communication skills across the curriculum and improving information literacy for all students [9]. Information literacy, as defined by the Middle States Commission in [8], refers to the ability of students to access relevant information, apply critical thinking skills to the content of that information, use sophisticated methods to pursue advanced lines of inquiry, and use the newly acquired knowledge for a specific purpose. The Distributed Curriculum Model, as described in the guidelines, suggests that information literacy can be achieved within the context of a specific discipline in upper division and graduate courses. This approach has the advantage of combining information literacy and research skills naturally with student interest. Further, it is essential that students understand research in the broader context of scholarship. In other words, the primary result of a research-oriented course or program should not be merely training in the specific skills relevant to the field of study, but rather an awareness of how the research fits into the larger picture of scholarly work [6].

A first course in artificial intelligence is usually targeted at advanced undergraduate students or graduate students, or perhaps both, depending upon course scheduling constraints. It is not uncommon in a small department for a single section to accommodate both graduates and undergraduates in the same classroom, posing a challenge for the instructor in terms of meeting the need of both student demographics. Undergraduate students usually have less programming experience, have had fewer opportunities to test their skills on large-scale projects, and are not as familiar with published literature in the field. Graduate students, on the other hand, have presumably mastered the broad as well as the fine aspects of the field and require a means of entry into the research community. This entry point usually takes the form of a master's project or thesis, and culminates in the student's defense of the completed project. An advanced course in the student's area of interest can serve as a source for possible thesis topics, however, the instructor must be careful not to alienate undergraduate students in the same classroom whose needs may differ. Image Understanding can be presented as a second course in AI, with the purpose of preparing graduate students for the master's thesis, and at the same time, introducing undergraduate students to research methodologies and publications from the current literature in the field.

2. COURSE OVERVIEW

Image Understanding was offered as an elective seminar course in the Computer Science Department, however it was open to any student who had successfully completed the prerequisite,

Conference'04, Month 1–2, 2004, City, State, Country.

Introduction to Computer Vision and/or Artificial Intelligence. The objectives of the course were defined as:

- Provide the students with the theoretical foundations of the field

- Introduce techniques for automating the extraction of semantic content from imagery

- Review the current literature in the field

- Evaluate the current literature in terms of the advantages and disadvantages of the approach used.

- Implement and present a self-selected programming project demonstrating an application of image understanding (this was required only of graduate students)

Given the objectives, a set of learning outcomes were identified:

- Students will demonstrate a thorough understanding of the algorithms and data structures used in the interpretation of images

- Students will be able to implement selected algorithms, evaluate the effectiveness of the approach, and communicate the result

- Students will be able to interpret, critically evaluate, and discuss the current literature in the field

The duration of the course was ten weeks (one academic quarter) and the topics covered are given as follows:

Week	Topics
1	Introduction, human visual perception and illusions
2	Knowledge representation, shape representation, structural information theory
3	Classification – parametric (Bayes' theorem) and non-parametric techniques, supervised and non-supervised learning
4	Clustering – hierarchical and partitional techniques, *k*-means, isodata
5	Control strategies for image understanding, semantic networks, active contour models
6	Scene labeling, semantic segmentation, genetic interpretation, hidden Markov models
7	3D information from 2D scenes, interpretation trees, shape from x, active vision, purposive vision
8	3D models and matching – geometric, relational, and functional models, octrees, balloons
9	Reasoning about images – Bayesian networks and decision graphs, principle components analysis
10	Student project presentations

Figure 1: Image Understanding Course Syllabus

Resources for the lecture topics were drawn from a variety of sources, including [2,10,13,14]. The class met for two hours twice a week, with 2-3 hours per week devoted to lecture, and 1-2 hours per week devoted to discussion. Discussions centered on a required reading assigned from the previous week. One student each week was selected to present the following week's reading assignment and to lead the discussion. Every student in the class had a chance to lead a discussion. All students were required to submit a 300-word summary of the reading and prepare three questions to pose during the discussion. Some of the assignments also included a programming component to enable the students to gain experience with the implementation of concepts expressed in the reading and lecture. All programming was done in MATLAB®, using the MATLAB® Image Processing Toolbox.

Students were evaluated based on their performance on a 2-hour midterm exam, six reading/programming assignments, class participation, a 2-hour final exam, and a term project (graduate students only). The project included a significant software solution to a current image understanding problem, an 8-10 page report detailing the background, approach, and results of the project, and a 30-minute presentation to the class. Students were allowed to choose their own research topic, however, during week three of the term, each student submitted a brief project proposal and presented the proposal to the class. This was to ensure that the selected project would be neither too difficult nor too easy for the student to complete in a ten-week term. Other students in the class posed questions and made comments about each proposal to provide direction, clarification, and motivation for the project.

3. COURSE CONTENT

Computer systems that automate the task of analyzing and interpreting image data are becoming increasingly valuable in the domains of both fundamental research and industrial settings. As such, computer vision and image understanding techniques bridge the gap between engineering and artificial intelligence, and combine aspects of both disciplines for specific purposes. Current applications include the control of manufacturing processes via parts inspection, remote sensing and interpretation of aerial and satellite imagery, medical diagnoses from x-ray and ultrasound images, content-based image retrieval from web servers, management of digital image archives, the automatic generation of digital photo albums, and robotic aids for the elderly or disabled. Fundamental research in image understanding contributes to and benefits from knowledge in diverse areas such as computational neuroscience, human vision and psychophysics, data mining, robotic planning, and computer graphics.

3.1 Image Understanding as AI

The ultimate goal of image understanding is for the machine to be able to "make sense" of image data, as people do. This goal differs from that of image processing and image analysis, where the image is presented to the system in a useful form and the goal is to extract information from the processed image. Processing includes image capture and digitization, noise removal, detection of luminance differences, detection of edges, and image enhancement. Analysis comes after processing and includes identifying boundaries, finding connected components, labeling regions, segmenting object parts, and grouping those parts together into whole objects. The data and results from processing and analysis are *quantitative* rather than *qualitative* in nature.

Image understanding, on the other hand, seeks to derive qualitative conclusions from quantitative data. Decisions must be made about the data in order to derive meaning and facilitate interpretation, and this usually requires information about the world that is external to, and separate from the image data. Thus, strategies and data structures developed for solving artificial intelligence problems are appropriate at this stage, and knowledge about human visual perception and psychophysics is useful.

Once the processed and analyzed data have been presented to the system, traditional AI techniques such as propositional and predicate calculus, inference rules, graph theory, state space search, nonmonotonic reasoning, fuzzy sets, expert systems, frames, and scripts can be used to reason about the data. These techniques can be combined with well-known theories about how humans perceive their world visually, and used to develop computational algorithms for extracting semantic information from scenes.

3.2 Human Vision and Psychophysics

Students typically assume that since interpretation comes after data acquisition and analysis, bottom-up (data-driven) control is always necessary. With bottom-up control, the flow of data is strictly from lower levels to successively higher levels of abstraction, with no feedback loops. Data is always processed and analyzed before proceeding to the next higher level. Recent studies of human visual perception have shown that this is not how people process visual information [5]. Humans employ top-down (context-driven) control extensively, with feedback loops from higher levels to lower levels. The purpose of top-down processing is to employ abstractions to limit the amount of information that neural hardware must assimilate and interpret. A consequence of top-down processing is that it gives rise to many visual illusions and is responsible for perceptual phenomenon such as illusory contours and amodal completion (Figure 2).

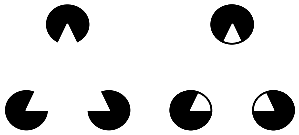

Figure 2: Illusory contours provide evidence that humans impose top-down constraints during perception. (After [5]).

Another lesson from human vision is that people prefer to selectively attend to certain informative parts of the scene, and fail to perceive significant changes in a scene if those changes occur during a visual transient [11]. This phenomenon, known as *change-blindness*, serves as an example of how efficient encoding may adversely affect visual recall. A web-based demonstration of change-blindness [4] was presented in class and convincingly persuaded students that efficient algorithms for image understanding must eschew a strict bottom-up control mechanism.

3.3 Structural Information Theory

Automated systems require quantitative rather than qualitative data. Therefore, although the implications of illusory contours and change blindness are intellectually satisfying, the recognition of the phenomena gives no clue as to the implementation details. Consider the Gestalt principle of figural goodness, or *prägnanz*. This principle states that given an ambiguous figure (one with more than one visual interpretation), we tend to perceive the alternative that gives rise to the simplest solution – i.e., figures that have “gute Gestalt”. For example, illusory contours can be explained with this principle by noting that a triangle floating over three circular black discs has more figural goodness than the alternative of a triangle situated inside three notched black discs. Good shapes are preferred because they can be more efficiently encoded, and are based on global properties of objects – symmetry, order, simplicity, and regularity.

Structural information theory [7], also known as coding theory, is a method for constructing shape descriptors of an ambiguous figure and comparing alternative descriptors to find the one that gives rise to the simplest interpretation. A code is a string that can generate the figure, i.e., it is an abstract and compact representation of the figure. Since a single figure can give rise to multiple interpretations (as in the example of illusory contours), a single figure may also have multiple codes. The minimum length code for a figure (i.e., the one with the lowest “information load”) represents the best interpretation in terms of figural goodness.

The algorithm to compute the minimum code for a figure is as follows:

For each possible interpretation of the figure:

- Construct a primitive code by tracing the contour of the figure (e.g., chain code) and record the contour as a sequence of line segments and the angles between those line segments

- Simplify (reduce) the primitive code by removing redundancies. This is accomplished via a set of semantic operators (rewrite rules):

$S \rightarrow n * (x)$ (Iterator operator)

$S \rightarrow SYM\ (x)$ (Symmetry operator)

$S \rightarrow \langle X \rangle \langle Y \rangle$ (Distribution operator)

- Count the number of parameters (numerical values) in the reduced code. This is the information load for that interpretation.

The interpretation with the lowest information load is the minimum code for the figure. Figures with few abrupt angles and more symmetry will tend to have a lower information load and will be the preferred perception.

Structural information theory is a means to quantify figural goodness and can explain many perceptual phenomena. Also, the technique is independent of the position, size, and orientation of the figure. The disadvantages are that it can only deal with simple lines and curves, it is sensitive to noise in the contour, and, most severely, it is computationally intractable for complex figures. To find the minimum information load, all possible interpretations must be considered. For example, in the case of a partially occluded square, there are an infinite number of possible interpretations because the occluded part cannot be sensed. This disadvantage can be overcome in practice, however, by selecting a small set of likely shapes to describe the occluded part. The purpose of presenting structural coding theory to the students was to introduce a computational technique for describing qualitative information, and to provide motivation for exploring similar ideas for potential research projects.

3.4 Blocks World Revisited

Blocks world [12] was an early attempt to interpret three-dimensional scene data. A scene in blocks world consists only of objects with trihedral corners, i.e., all surfaces are planar and all corners are formed by the intersection of three surfaces. The goal is to interpret a scene by classifying all edges according to the environmental situation that produced them. For example, an edge may be the result of a shadow, an occlusion, or contact with another surface.

Two types of edges are most important for 3D scene interpretation – orientation edges, which are due to two surfaces touching each other, and depth edges, which indicate a spatial discontinuity between two surfaces, for example when one surface occludes another. There are two kinds of orientation edges – convex and concave, and two kinds of depth edges – closer and farther. Thus when labeling an image, there are four possible labels for each edge. For a figure with *n* edges, there are 4^n possible labelings, corresponding to 4^n qualitatively different depth interpretations. This is an intractably large number of logically possible interpretations for realistic scenes, yet most people will perceive only one or two (Figure 3). Independent studies in [1,3] found that most *logically* possible interpretations are not *physically* possible. In fact, there exist only 16 physically possible combinations of edge labels for any given trihedral corner. A further constraint is that each edge must have a consistent labeling along the surface, eliminating all but a few possibilities for the final interpretation.

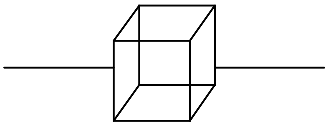

Figure 3: Cube with more than one 3D interpretation

An interpretation tree is used to find all physically possible labelings for a given 3D trihedral figure. To create an interpretation tree, begin with an arbitrary vertex as the root of the tree. Next, generate children that represent all physically possible labelings of that vertex. For each of those labelings, choose an adjacent vertex and assign labels to that vertex which are consistent with the already assigned label(s). Many of the 16 possibilities will be eliminated at this point due to inconsistencies along the edge. Continue choosing adjacent vertices and assign edge labels until all vertices have consistent labels. When the procedure is finished, the leaves of the tree will represent the consistent labelings, which correspond to the different physically possible interpretations of the figure. It is possible that no consistent labeling exists for the figure. This situation corresponds to an “impossible figure”, as shown in Figure 4, which is a plausible 3-dimensional figure that has no physical analogue.

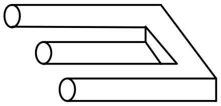

Figure 4: Impossible figure

4. STUDENT PROJECTS

Structural information theory and interpretation trees are just two techniques for incorporating semantics into visual representations. Many other techniques exist, as shown in the course syllabus given in Figure 1. The purpose of the student project was to allow the student to select a technique, implement it in software, analyze the advantages and disadvantages of the approach, and attempt to overcome limitations inherent in the selected technique. Thus, the project must incorporate a creative component and show originality of approach, while building upon existing work. These are basic component of any research project, and provide a means for students to experience and prepare for comprehensive research projects at a higher intellectual level. The following is a list of the student projects for the Spring 2005 Image Understanding class:

- “Impossible figure detection”
- “Solving the EZ-Gimpy CAPTCHA”
- “Scene classification for natural images”
- “Face detection with shape contexts”
- “Eye detection”
- “Eye detection in multi-scale color images”
- “Breaking a visual CAPTCHA”
- “Iris location in an eye image”

The two students in the class who were registered as undergraduates were not required to submit a project. All students successfully completed the course, and four of the ten students in the class expressed a strong interest in continuing research in this area after completion of the course. Two of these students are currently working on independent study research projects with the author on the topic of automated game rule learning from video-streamed perceptual observations. All four students desire to pursue a master’s thesis or project in this area. Of the six who did not express an interest in continuing, one is an undergraduate, three have a previous commitment to another thesis project in this field, and two are currently undecided.

5. STUDENT FEEDBACK

Student comments at the completion of the course are given below (all comments were made anonymously):

“The class was great, definitely liked the conversational atmosphere.”

“Some of the readings (well, only one or two) were just a bit too long/dense for my tastes.”

"More programming assignments would be helpful, should continue giving papers to read. Some review about algorithms like corner detection, Hough transform general, and important algorithms review and understanding could be helpful."

"The labs were good and helpful. I liked the readings and then talking about them."

"An interesting survey of the material. Paper reading was the best part of the course."

"During this course we went through various research papers from well-known authors and organizations. Also the homework assignments were very helpful. The tests could be simpler."

"A very informative class. The small class size was great for discussions and let everyone have a voice. Reading papers every week was very helpful to see what was going on in the field. However, more programming assignments would have been helpful."

"Very interesting course! Reviewing current papers is a good idea to gain knowledge of what has been done in the field."

6. CONCLUSION

Preparing students for research involves guiding them toward an independent investigation of a currently unsolved problem, encouraging them to think creatively and critically about proposed solutions, helping them to develop the ability to implement a solution, and requiring them to communicate the results to others. Image Understanding, as a second course in artificial intelligence, is an example of a course that was well suited to helping students meet this challenge. The course proved to be successful as a means for bridging the educational gap that exists between the first few years of post-secondary coursework and upper-division/graduate classes, where the emphasis is less on the assimilation of new information, and more on information literacy and expanding the body of knowledge.

7. ACKNOWLEDGMENTS

I would like to thank the students who participated in making this course a success – Dave Wollenhaupt, Christopher Orogvany, Tim Lebo, Alex Jarocha-Ernst, Justin Hnatow, Jonathan Lareau, Sneha Sharma, Sahil Dhawan, Tayeb Karim, and Yury Luzanov. You have proved to me that I have as much to learn from my students, as you have to learn from your teachers.

8. REFERENCES

[1] Clowes, M.B. On seeing things. *Artificial Intelligence*, 2, 1971, 79-116.

[2] Gose, E., Johnsonbaugh, R., and Gose, S. *Pattern Recognition and Image Analysis*. Prentice-Hall, Upper Saddle River, NJ, 1996.

[3] Huffman, D.A. Impossible objects as nonsense sentences. In M. Meltzer & D. Michie (Eds.) *Machine Intelligence (vol. 6)*. Edinburgh University Press, Edinburgh, Scotland, 1971.

[4] http://www.usd.edu/psyc301/ChangeBlindness.htm

[5] Kellman, P.J., and Shipley, T.F. A theory of visual interpolation in object perception. *Cognitive Psychology*, 23(2), 1991, 141-221.

[6] LaPidus, J.B. *Doctoral Education: Preparing for the Future*. (September 1997). http://www.cgsnet.org/pdf/doctoraledpreparing.pdf

[7] Leeuwenberg, E.L.J. A perceptual coding language for visual and auditory patterns. *American Journal of Psychology*, 84(3), 1971, 307-349.

[8] Middle States Commission on Higher Education. Characteristics of Excellence in Higher Education: Eligibility Requirements and Standards for Accreditation, 2002.

[9] Middle States Commission on Higher Education. Developing Research & Communication Skills: Guidelines for Information Literacy in the Curriculum – Executive Summary, 2002.

http://www.msache.org/msache/content/pdf_files/devskill.pdf

[10] Palmer, S.E. *Vision Science: Photons to Phenomenology*. MIT Press, Cambridge, MA, 1999.

[11] Rensink, R.A. Seeing, sensing, and scrutinizing. *Vision Research*, 40, 2000, 1469-1487.

[12] Roberts, L.G. Machine perception of three-dimensional solids. In J.T. Tippett, D.A. Berkowitz, L.C. Clapp et al. (Eds.) *Optical and Electro-optical Information Processing*. MIT Press, Cambridge, MA, 1965.

[13] Shapiro, L.G., and Stockman, G.C. *Computer Vision*. Prentice-Hall, Upper Saddle River, NJ, 2001.

[14] Sonka, M., Hlavac, V., and Boyle, R. *Image Processing, Analysis, and Machine Vision, 2nd Edition*, Brooks/Cole Publishing, Pacific Grove, CA, 1999.