Using Domain Knowledge for Low Level Vision

Theo Pavlidis
Distinguished Professor Emeritus
Stony Brook University

t.pavlidis@ieee.org
http://www.theopavlidis.com/
An Industrial Vision Problem*

- Capture a single image of a rectangular shipping box and provide an estimate of its three dimensions (Height, Width, Depth)
- Device includes two laser beams whose spots on the box are captured and used to estimate absolute size.
- Relative size of H, W, and D must be found from the analysis of a single image.

* Symbol Technologies, Holtsville, NY, circa 2001
Basic Idea: Because the three edges meeting at a vertex are mutually perpendicular we can compute their relative size from one view.
Typical Image of Interest

Our goal is to use image analysis to go from the above image to a line drawing such as that shown in the previous slide.
A paradox

• Human viewers have no trouble identifying the box and its edges.

• Application of Edge Detection or Segmentation produces a “mess:”
  – Contrast inside the box may be higher than contrast between the box and the background.

• What does this observation imply about Machine Vision?
Do we really understand human vision?
Reading Demo - 1

It is hard to explain the human ability of reading dot-matrix print and fine laser print by purely bottom up processes.
New York State lacks proper facilities for the mentally ill.

The New York Jets won Superbowl III.

• Human readers may ignore entirely the shape of individual letters if they can infer the meaning through context.
Reading Demo - 3

The behavior of Machines
Tentative binding on the letter shapes (bottom up) is finalized once a word is recognized (top down). Word shape and meaning over-ride early cues.
What Neuroscientist Say - 1

• “In real-life situations, bottom-up and top-down processes are interwoven in intricate ways," and "progress in psychobiology is ... hampered ... by our inability to find the proper levels of complexity for describing mental phenomena”

What Neuroscientist Say - 2

• “Perceptions emerge as a result of reverberations of signals between different levels of the sensory hierarchy, indeed across different senses”. The authors then go on to criticize the view that “sensory processing involves a one-way cascade of information (processing)”

Using Domain Knowledge for Low Level Vision
Using Domain Knowledge for Low Level Vision
Using Domain Knowledge for Low Level Vision

- Image
- Feature Extraction
- Final Decision
- Prior Knowledge
Back the Box Case

- **Challenge:** Contrast within a box is often higher than contrast between box and background.

- **Facilitating factor:** We know that the box occupies most of the image.
  - The device is aimed at the box and there is auditory feedback (beep) when the measurement is completed.
An Inspiration from Nature

• In a classical paper J. Letvin *et al* showed that the frog’s visual system responds to only two kinds of stimuli:
  – fast moving, high contrast small shapes (food) or
  – decrease in the ambient illumination (danger).

  *[Proceedings of IRE, 1959]*
An Inspiration from Nature translated to the box dimension problem

• The system should look only for hexagonal shapes occupying most of the image.
• This means that the only edges of interest should be lines of length comparable to the dimensions of the field of view.
• Such lines should form a convex set.
• The convex set should be a hexagon.
Another Challenge

• The system must work **ALL THE TIME** in the hands of “blue collar” workers.
  – (Not only on a group of selected images with the system operated by PhD candidates.)

• *Therefore:* There is no way to obtain an adequate “training” set of images.
Methodology

• In order to deal with the contrast issues we designed the low level vision part on the basis of top level (domain) knowledge.

• In order to deal with the lack of a training set we kept heuristics to a minimum and relied on mathematically rigorous algorithms.
Acknowledgments

• The project was carried out at Symbol Technologies in collaboration with Ke-Fei Lu, Eugene Joseph, Jackson D. He, and Ed Hatton during 2000-2002.

• Symbol Technologies no longer exists. In January 2007 it was acquired by Motorola.
Publications


• On the Web: http://www.theopavlidis.com/technology/BoxDimensions/overview.htm
We use (Long) Line Detection as the first step (rather than segmentation or edge detection)
Line Finder

• In a given area find the pixel $P$ with the maximum gradient.
• We select a line through $P$, perpendicular to the gradient that divides the area into two parts.
• For each part we calculate its mean and we keep the line only if the two means are significantly different.
• All parameters are determined adaptively.
Proximity Clusters

• The line segments found are merged to find long lines (we look at co-linearity for that).
• The lines found are then clustered into proximity clusters.
• A proximity cluster is defined as a set of line segments $L$ with the property that for each $s$ in $L$, there is a $t$ in $L$, such that $t$ and $s$ have at least a pair of endpoints near each other.
Examples of Proximity Clusters
Convex Hull

• Next we find the convex hull of each cluster as well as that of groups of clusters. (We use a standard algorithm for the process.)
Editing the Convex Hull (Main Heuristic)

• Line segments of the convex hull are assigned a confidence level that is high if they are nearly collinear to a line segment of the cluster.

• Line segments with low confidence (red in figures) are removed together with all line segments that contributed to them.
Editing Example
Editing Example
Editing Example
Editing Example
Editing Continued

• We also check how closely the convex hull resembles a hexagon (the projection of a rectangular object) and remove edges that reduce the quality.
Sequence of Editing Operations
More on Editing

• From the hexagon we can infer the “Y” around a vertex and thus the relative dimensions of the rectangular box.

• After the line segments have been found the rest of the operations (clustering, convex hull finding and editing, dimension estimation) are very fast because we deal with very few objects (20-30 line segments) rather than 480x640 pixels!
Technological Conclusions

• Field tests proved that the system was reliable.
• Symbol Technologies was hoping to sell such gadgets to shipping companies (such as UPS).
• Drivers could measure immediately the size of pick ups and radio the information to the basis. There a program would compute allocating packages to containers.
• However customers were not interested without a demonstration of the whole system (including the “bin packing” part) that was never prototyped.
Business Conclusions

- Other applications: The device could be used in a hub to measure dimensions of boxes while on a conveyor belt (customers are charged both by weight and size).
- Not clear how cost effective that would be. Also few units would be needed and Symbol Technologies lost interest.
- Around that time the company was also rocked by accounting scandals.
Scientific Conclusions

• Research in Image Analysis (or Machine Vision) has been going on for over 40 years.
• We still do not have good and general segmentation or object outlining algorithms. Probably they do not exist.
• It is best to derive special low level processing algorithms for each application based on top level knowledge.
General Challenges to Machine Vision

• We need to replicate complex transformations that the (human/animal) brain has evolved to do over hundreds of millions of years.

• We have to deal with the fact the processing is not unidirectional and also affected by other factors besides input (context both inside and outside the image). Visual illusions (far more common than auditory illusions) attest to that fact.
Why is Machine Vision so Hard?

- Organisms with complex visual systems have existed for over 300 million years.

- Speech has existed for less than 200 thousand years.
A Malady: “Proof” by Example

• An algorithm is applied to a set of images and its parameters are chosen to give a satisfactory results on subset of these images (*Learning Subset*). Then the algorithm is tested on the remaining images of the set (*Testing Subset*) and if the results are satisfactory the algorithm is considered a success.

• However, that particular set is unlikely to be a representative example of all images.

• The space of all possible image is huge!!!
What is the Number of All Possible Images?

• $10^{56}$ is a very conservative lower bound to the number of all possible meaningful/valid images. The number of all meaningful/valid images is at least as high as $10^{400}$.

• See:
  – T. Pavlidis "The Number of All Possible Meaningful or Discernible Pictures" *Pattern Recognition Letters*, vol. 30 (2009), pp. 1413-1415.
An Illustration - 1

• A few years ago I worked on a method for Image Retrieval (CBIR). The method did quite well on a set of about 5,000 images.

• I expanded that set by a factor of about 100 by generating new images from the originals by simulating over- and under-exposure, shadows, and other visual artifacts.

• The method did very poorly on the set of 500,000 images.
The picture in the middle is a brightened version of the picture on the left but two different sets of feature measures classify it as being closer to the picture on the right. The values were:

First feature set: \( \text{dist}(\text{Left, Mid}) = 25, \text{dist}(\text{Left, Right}) = 25, \text{dist}(\text{Mid, Right}) = 0. \)

Second feature set: \( \text{dist}(\text{Left, Mid}) = 42, \text{dist}(\text{Left, Right}) = 50, \text{dist}(\text{Mid, Right}) = 23. \)

http://www.theopavlidis.com/technology/CBIR/PaperB/vers3.htm
Conclusions

• We need mathematical models of the scenes we try to interpret.

• Such models would allow:
  – Design of effective low level vision algorithms.
  – Analytical validation of the results.

• While precise models may be hard to come by, approximate models provide many of their benefits.