**“INTRODUCTION TO COMPUTER VISION”**

**BY**

**ANEEL MAHMOOD**

**COMP 840 FALL 2019**

**Abstract:**

In this paper I give a somewhat personal and perhaps biased overview of the field of Computer Vision. First, I define computer vision and give a very brief history of it. Then, we outline some of the reasons why computer vision is a very difficult research field. Finally, we discuss past, present, and future applications of computer vision. Especially, we give some examples of future applications which we think are very promising.

1. **What is Computer Vision?**

If I asked you to name the objects in the picture below, you would probably come up with a list of words such as “tablecloth, basket, grass, boy, girl, man, woman, orange juice bottle, tomatoes, lettuce, disposable plates…” without thinking twice. Now, if I told you to describe the picture below, you would probably say, “It’s the picture of a family picnic” again without giving it a second thought.



Those are two very easy tasks that any person with below-average intelligence and above the age of six or seven could accomplish. However, in the background, a very complicated process takes place. The human vision is a very intricate piece of organic technology that involves our eyes and visual cortex, but also takes into account our mental models of objects, our abstract understanding of concepts and our personal experiences through billions and trillions of interactions we’ve made with the world in our lives.

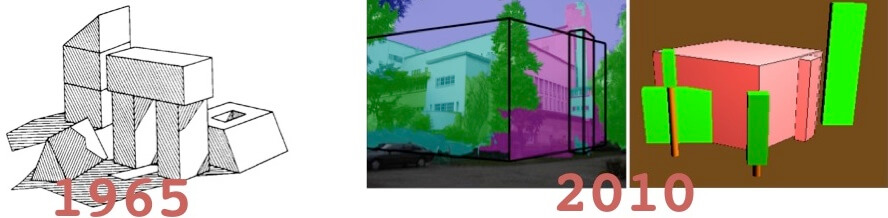
Digital equipment can capture images at resolutions and with detail that far surpasses the human vision system. Computers can also detect and measure the difference between colors with very high accuracy. But making sense of the content of those images is a problem that computers have been struggling with for decades. To a computer, the above picture is an array of pixels, or numerical values that represent colors.

Computer vision is the field of computer science that focuses on replicating parts of the complexity of the human vision system and enabling computers to identify and process objects in images and videos in the same way that humans do. Until recently, computer vision only worked in limited capacity.

Thanks to advances in artificial intelligence and innovations in deep learning and neural networks, the field has been able to take great leaps in recent years and has been able to surpass humans in some tasks related to detecting and labeling objects.

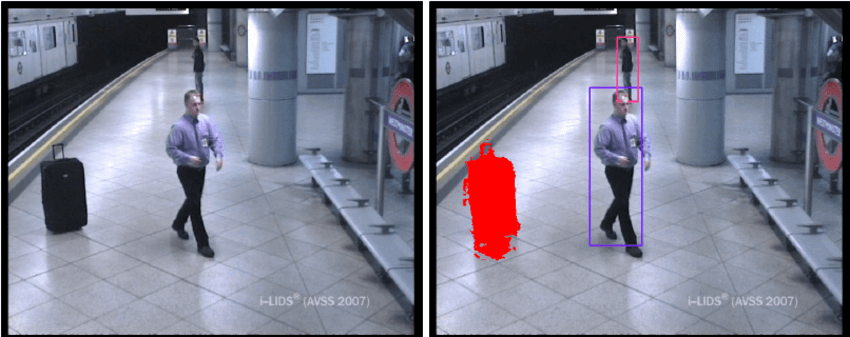
1. **History of Computer Vision**

When computer vision started to take shape as a field in the 1960s, its aim was to try and mimic human vision systems and ask computers to tell us what they see, automating the process of image analysis. This kind of technology is the precursor to artificially intelligent image recognition. Before, any kind of image analysis had to be done manually, from x-rays to MRIs to hi-res space photography.



Just like animals, computers “see” the world differently from us humans: basically, they count the number of pixels, try to discern borders between objects by measuring shades of color, and estimate spatial relations between objects.

As computer vision evolved, algorithms started to be programmed to solve individual challenges, and they become better at doing the job the more they repeat the task.



*Example: CCTV cameras in London Underground stations programmed to spot static / abandoned objects*

Fast forward to 2010 (and beyond), we have seen an acceleration in improved deep learning techniques and technology. With deep learning, we’re now able to program supercomputers to train themselves, self-improve over time and provide portions of these capabilities to businesses as online applications, like cloud-based apps.

In order for these machines to learn, they need to be fed data.

In a world where the biggest players like Facebook and Instagram limit how much of their content other actors are able to tap into, there’s been a rise in open-source projects such as ImageNet. ImageNet’s mission is to create a large-scale image database that researchers can tap into in order to train and manufacture their algorithms.

The challenge is that in order for computers to index and catalogue these huge sets of data, they initially need to have some human input in terms tagging and classifying their ‘training images’. Deep learning algorithms then use this information to create benchmarks to compare future images with but need to be fed large quantities of training images, as many as 10s of millions.

1. **Why is Computer Vision Difficult?**

The most difficult things to understand are the ones that are fundamental. We live in a visual world. We see things and instantly understand what they are and what we can do with them. We can not only identify them but also understand their particular attributes and the classification they belong in. Without thinking too deeply about it we can look at a book and understand the process of the idea, the thinking, the writing, the writer, the agent, the editor, the publishing house, the printer, the marketer, the salesman, the bookshop, the economy. And this is a far from exhaustive list of association that revolve around a book.

Even more to the point we can each look at a book and understand what it is and what it does even if we have not all got the same nuanced understanding of what goes into its making and what happens around it. Our depth of knowledge (or lack of it) will only play a role in specific contexts (like a book convention or a forum on the economy) but for most everyday purposes a number of us that is sufficiently large to be called the majority, will be able to ascribe most of the key attributes that make the book into an entity.

A book may be a great read if it is well written but, in the cases, where this is not the case it also makes a great doorstop.

So vision, really, is a knowledge thing as opposed to an eye thing and this is where things get very interesting. Knowledge is based on what is accepted of the real world and that includes both factual and imaginary things. We can all agree, for instance, on who Harry Potter is what he did and why he did it while we all also agree that he is an imaginary being. This means that in order to understand what we see we do not just use deductive reasoning whereby what we reach as a conclusion is 100% true, we also use inductive reasoning where we extrapolate from premises that are probably true and reach conclusions that are likely: “A book may be a great read if it is well written but in the cases where this is not the case it also makes a great doorstop.”

In that last sentence we imagine not only instances of success and failure but also a world where wit and sarcasm play a part in describing quality. Computers can be equipped with hardware that captures data (as in light in this case) in a far broader spectrum than our organic eyes, and they can also be equipped with algorithms that do such a marvelous job of interpreting this data that just by studying patterns of light they can learn to see around corners. Computers can also be equipped to perform deductive reasoning.

In a paper presented at the 10th International Conference on Artificial General Intelligence, AGI 2017, held in Melbourne, Australia, Army Laboratory researcher Douglas Summers Stay presented his paper on “deductive reasoning on a semantically embedded knowledge graph” with an abstract that read: “Representing knowledge as high-dimensional vectors in a continuous semantic vector space can help overcome the brittleness and incompleteness of traditional knowledge bases. We present a method for performing deductive reasoning directly in such a vector space, combining analogy, association, and deduction in a straightforward way at each step in a chain of reasoning, drawing on knowledge from diverse sources and ontologies.” — reasoning using a combination of “analogy, association, and deduction” and “drawing on knowledge from diverse sources and ontologies.” is exactly what semantic search is designed to do and notice how Stay refers to “a continuous semantic vector space” which mirrors the way the brain achieves infinite storage capacity in a finite space.

A close up of a logo

Description automatically generated

Image from US7519200B2 awarded to Google Inc shows the variables that must be recognized in order to assign a high degree of accuracy to the reading of a face by a computer.

So, now it sounds like we have the problem virtually solved. Yes and no. Yes, in that search has become very good at understanding objects in images (and sometimes, video) and building an index which when coupled to its knowledge base for context can create a pretty good sense of search being intelligent. Two Google patents in particular point out just how this is done in a way that make it appear seamless.

Both of these highlight the issue that is the major stumbling block: namely context and, by association, inductive reasoning. Because knowledge constantly evolves and morphs into analogies that contain meaning but make no sense (like my example above of using a badly written book as a doorstop) computer logic stumbles on inductive reasoning when it is unsupervised and this, also affects, computer vision, at least, where specific contexts are involved. So, it may be OK to have a self-driving vehicle where the on-board telemetry provides a system of vision that is superior to anything a human could bring to bear (think of a car that could ‘see’ through fog, know what’s coming around corners and can even tap into traffic sensors as well as traffic reports to optimize the journey) but a robot nanny is more problematic (because children are inherently unpredictable and in their unpredictability lie risks to their safety).

1. **Application of Computer Vision**

Past and present applications of computer vision include: Autonomous navigation, robotic assembly, and industrial inspections. At best, the results have been mixed. (I am excluding industrial inspection applications which involve only 2D image processing and pattern. recognition.) The main difficulty is that computer vision algorithms are almost all brittle; an algorithm may work in some cases but not in others. My opinion is that in order for a computer vision application to be potentially successful, it has to satisfy two criteria: 1)Possibility of human interaction. 2) Forgiving (i.e., some mistakes are tolerable). It also needs to be emphasized that in many applications vision should be combined with other modalities (such as audio) to achieve the goals. Measured against these two criteria, some of the exciting computer vision applications which can be potentially very successful include: Image/video databases-Image content-based indexing and retrieval. Vision-based human computer interface - e.g., using gesture (combined with speech) in interacting with virtual environments. Virtual agent/actor - generating scenes of a synthetic person based on parameters extracted from video sequences of a real person. It is heartening to see that a number of researchers in computer vision have already started to delve into these and related applications.

1. **Conclusion**

Computer Vision is more than 30 years old. Although as a research field it has been offering many challenging and exciting problems, in terms of successful engineering applications it has been rather disappointing. However, more recently, several very exciting applications have appeared where computer vision I believe can make major contributions.

1. **References**
2. *A Gentle Introduction to Computer Vision*

[*https://machinelearningmastery.com/what-is-computer-vision/*](https://machinelearningmastery.com/what-is-computer-vision/)

1. *Huang, T. (1996-11-19). Vandoni, Carlo, E (ed.).*[*Computer Vision : Evolution And Promise*](http://cds.cern.ch/record/400313/files/p21.pdf)*(PDF).*[*19th CERN School of Computing*](http://cds.cern.ch/record/300250)*. Geneva: CERN. pp. 21–25.*[*doi*](https://en.wikipedia.org/wiki/Digital_object_identifier)*:*[*10.5170/CERN-1996-008.21*](https://doi.org/10.5170%2FCERN-1996-008.21)*.*[*ISBN*](https://en.wikipedia.org/wiki/International_Standard_Book_Number)[*978-9290830955*](https://en.wikipedia.org/wiki/Special:BookSources/978-9290830955)*.*
2. [*Linda G. Shapiro*](https://en.wikipedia.org/wiki/Linda_Shapiro); George C. Stockman (2001). Computer Vision. Prentice Hall. [*ISBN*](https://en.wikipedia.org/wiki/International_Standard_Book_Number) [*978-0-13-030796-5*](https://en.wikipedia.org/wiki/Special:BookSources/978-0-13-030796-5).
3. Nicu Sebe; Ira Cohen; Ashutosh Garg; Thomas S. Huang (3 June 2005). [*Machine Learning in Computer Vision*](https://books.google.com/books?id=lemw2Rhr_PEC). Springer Science & Business Media. [*ISBN*](https://en.wikipedia.org/wiki/International_Standard_Book_Number) [*978-1-4020-3274-5*](https://en.wikipedia.org/wiki/Special:BookSources/978-1-4020-3274-5).
4. E. Roy Davies (2005). Machine Vision: Theory, Algorithms, Practicalities. Morgan Kaufmann. [*ISBN*](https://en.wikipedia.org/wiki/International_Standard_Book_Number) [*978-0-12-206093-9*](https://en.wikipedia.org/wiki/Special:BookSources/978-0-12-206093-9).