Real Image of Computer Vision Application and its Impact: Future and Challenges

Article · December 2020

1 author:

Nakul Shahdadpuri
ABV-Indian Institute of Information Technology and Management Gwalior

Real Image of Computer Vision Application and its Impact: Future and Challenges

Nakul Shahdadpuri*

Indian Institute of Information Technology, Gwalior, India

Abstract: Since decades, computer vision, or the ability of artificially smart devices to 'see' like humans, has been a topic of growing interest and thorough study. In the field of computer vision, research aims to create machines that can automate tasks that involve visual awareness as a way of simulating the human visual system. The method of deciphering images, however, is much more complicated than understanding other types of binary information because of the substantially greater amount of multi-dimensional data that requires interpretation. This makes it more complicated to create AI systems that can recognize visual data. The future of computer vision seems to be full of hope and unimaginable performance, with equally amazing feats of AI with computer vision technology becoming increasingly common in various industries. Computer vision tasks include methods for digital images to be obtained, processed, interpreted and understood, and high-dimensional data extracted from the real world in order to generate numerical or symbolic information, e.g. in the form of decisions. In this sense, comprehension implies the transformation of visual representations (the input of the retina) into world descriptions that can interface with other processes of thinking and evoke effective action." This interpretation of images can be seen as the separation of symbolic data from image data using models developed with the aid of geometry, physics, statistics, and learning theory. This article was an overview of the real image of computer vision mission exploration past, present, future, impact and its challenges.

Keywords: Real Image, Computer Vision, Application, Impact, Future, Challenges.

Article History

*Received: 16/11/2020; Accepted: 27/12/2020

Corresponding author: Nakul Shahdadpuri

I.INTRODUCTION

Nowadays the increased use of mobile cameras today means a steady stream of images and videos, and the technology of Computer Vision has become readily available, making it even more attractive to companies. In less than a decade, accuracy rates for object recognition and classification have soared from 50% to 99% and today's systems are even more accurate than humans [1]. Computer vision is, by definition, a field that involves methods for collecting, processing, analyzing , and interpreting images and, in general, high-dimensional real-world data for the production of numerical or symbolic information , e.g. in decision-making forms. In the 1950s, when early neural networks started to detect the edges of objects and to organize them by their types, Computer Vision took its first steps. The first commercial Computer Vision systems were used in the 1970s, using optical character recognition (OCR) to read written text for the blind. Huge collections of images became available online for review as the internet evolved in the 1990s, driving the development of facial recognition programmes.

II. CONCEPT OF COMPUTER VISION AND ITS EVOLUTION

Computer vision is an interdisciplinary scientific field that deals with how visual images or videos can help computers achieve high-level understanding. In this context, comprehension implies the transformation of visual images (the retina input) into world explanations that make sense of thought processes and can evoke effective action. This interpretation of images can be seen as the disengagement of symbolic knowledge from image data using models developed with the assistance of geometry, physics, statistics, and the theory of learning[2-9]. To construct computer vision systems, the technical discipline of computer vision aims to apply its theories and models. Site reconstruction, event identification, video tracking, object recognition, 3D pose estimation, learning, indexing, motion estimation, visual servo, 3D scene modeling, and image recovery are sub-domains of computer vision[7]. Fast image acquisition enables 3D measurement and feature tracking to be realized when paired with a high-speed projector.[10] Egocentric vision systems are composed of a wearable camera that takes photographs from a first-person perspective automatically. In addition to CPUs and graphics processing units (GPUs) in this position, vision processing units are emerging as a new processor class as of 2016[11].

- *How does it work?*

In both Neuroscience and Machine Learning, one of the key unanswered questions is: How exactly do our brains function and how do we estimate it with our own algorithms? The truth is that there are very few working and systematic brain computation theories; so despite the truth that neural networks are supposed to 'mimic the way the brain functions,' no one is very sure if that's true. Let's leave our fuzzy cat friends on the side for a moment and let's have more technique. Below is a clear example of the buffer of the grayscale image that stores our Abraham Lincoln image. A single 8-bit number ranging from 0 (black) to 255 (white) reflects each pixel's brightness:

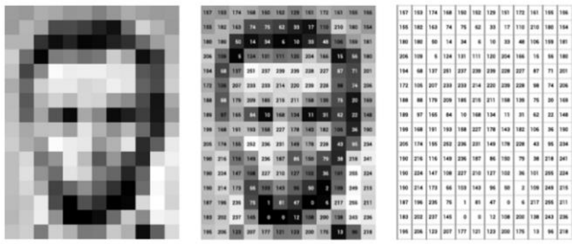

Figure 1 Data diagram for pixels. Our image of Lincoln on the left; the pixels labelled with numbers from 0-255 at the middle, reflecting their brightness; and those numbers on the right by themselves. Pixel values, in fact, are almost uniformly stored in a one-dimensional array at the hardware level. The data from the above picture, for example, is stored in a way similar to this long list of unsigned chars.

```
{157, 153, 174, 168, 150, 152, 129, 151, 172, 161, 155, 156,
 155, 182, 163,  74,  75,  62,  33,  17, 110, 210, 180, 154,
 180, 180,  50,  14,  34,   6,  10,  33,  48, 106, 159, 181,
 206, 109,   5, 124, 131, 111, 120, 204, 166,  15,  56, 180,
 194,  68, 137, 251, 237, 239, 239, 228, 227,  87,  71, 201,
 172, 105, 207, 233, 233, 214, 220, 239, 228,  98,  74, 206,
 188,  88, 179, 209, 185, 215, 211, 158, 139,  75,  20, 169,
 189,  97, 165,  84,  10, 168, 134,  11,  31,  62,  22, 148,
 199, 168, 191, 193, 158, 227, 178, 143, 182, 106,  36, 190,
 205, 174, 155, 252, 236, 231, 149, 178, 228,  43,  95, 234,
 190, 216, 116, 149, 236, 187,  86, 150,  79,  38, 218, 241,
 190, 224, 147, 108, 227, 210, 127, 102,  36, 101, 255, 224,
 190, 214, 173,  66, 103, 143,  96,  50,   2, 109, 249, 215,
 187, 196, 235,  75,   1,  81,  47,   0,   6, 217, 255, 211,
 183, 202, 237, 145,   0,   0,  12, 108, 200, 138, 243, 236,
 195, 206, 123, 207, 177, 121, 123, 200, 175,  13,  96, 218};
```

Figure 2 data image

This way of storing image data can be contradictory to your standards, because when it is viewed, the data definitely appears to be two-dimensional. This is the case, however, since computer memory is simply an ever-increasing linear array of address spaces.

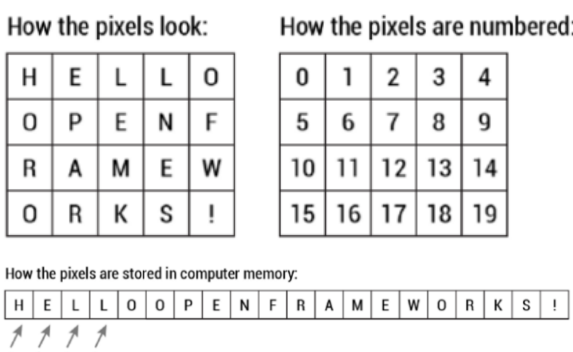

Figure 3 How to pixels look like in number or figure form

Let's go back to the first picture again and imagine a coloured one being added. Things are beginning to get more nuanced now. Computers normally read colour on the same 0-255 scale as a sequence of 3 values, red, green, and blue (RGB). Now, in addition to its location, each pixel actually has 3 values for the machine to store. That would lead to 12 x 16 x 3 values, or 576 numbers, if we were to colourize President Lincoln.

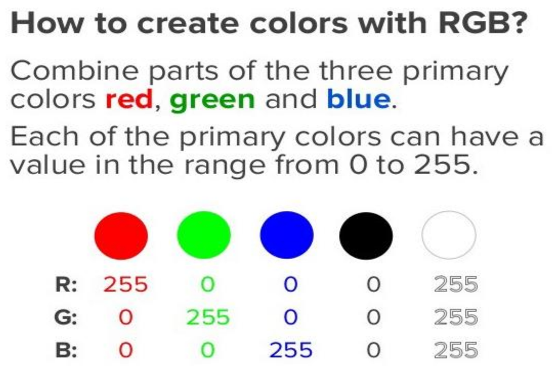

Figure 3 how colours look like in in number or figure form

That's a lot of memory for one image, and a lot of pixels for iterating over an algorithm.

III. TYPES OF METHOD, APPLICATIONS AND ITS IMPACT

A computer vision system organization is highly application-dependent. A computer vision system's precise implementation often depends on whether its functionality is pre-specified or whether any aspect of it can be learned or altered during service. Most functions are unique to the programme. There are, however, standard functions found in many systems for computer vision.

- **Image acquisition** – A digital image is created by one or more image sensors that include range sensors, tomography instruments, radar, ultrasonic cameras, etc., in addition to different types of light-sensitive cameras. The resulting image data is an ordinary 2D image, a 3D amount, or an image series, depending on the type of sensor. Typically, pixel values correspond to light intensity in one or more spectral bands (grey images or colour images), but may also be correlated with different physical measurements, such as sonic or electromagnetic wave depth, absorption or reflectance, or nuclear magnetic resonance.
- **Pre-processing** – In order to retrieve a particular piece of information, before a computer vision system can be applied to image data, it is typically important to process the data in order to ensure that it follows some assumptions implied by the system.

- **Feature extraction –** From the image data, image features at different levels of complexity are extracted. Texture, form or motion can be related to more complex features.
- **Detection/segmentation** – A decision about which image points or regions of the image are important for further processing is taken at some point in the processing. Examples are:
  - Selection of a special set of points of interest.
  - Segmentation of one or more regions of the image which contain a particular object of interest.
  - Image segmentation into nested scene architecture comprising foreground, groups of objects, single objects or parts of prominent objects (also referred to as hierarchy of spatial-taxon scene), while visual salience is often used as spatial and temporal attention.
  - Segmentation or co-segmentation into a sequence of per-frame foreground masks of one or more videos, while preserving its temporal semantic continuity.
- **High-level processing** – The input is usually a small collection of data at this level, such as a set of points or an image region that is presumed to contain a particular entity. For example, the remaining processing addresses:
  - Verifying that the knowledge follows model-based and application-specific assumptions.
  - Application-specific parameter estimation, such as object pose or object size.
  - Image recognition-classification of a detected object into various groups.
  - Image registration: two different views of the same object are compared and merged.
- **Decision making:** Making the final decision needed for the request, for instance:
  - Pass / fail on applications with automated inspection.
  - In recognition applications, match / no-match.
  - Flag for more human examination in applications for medical, military, protection and recognition.

Image-understanding systems

Three abstraction levels are used in image-understanding systems (IUS) as follows: low level includes image primitives such as borders, texture components, or regions; intermediate level includes boundaries, surfaces, and volumes; and high level includes objects, scenes, or events.

Impact of technology

The effect of technology is felt through a wide variety of fields that depend on computers to interpret images. This includes the fields of military, manufacturing, healthcare, automobile, information and retail. The applications of its technology seem almost infinite as Computer Vision continues to mature.

- *Military*

Computer Vision is a crucial enabling technology for modern armies that help security systems identify enemy troops or saboteurs and improves guided missile systems' targeting capabilities. In order to provide battlefield intelligence used for tactical decision-making, military principles such as situational awareness depend heavily on image sensors. The areas of autonomous vehicles that need to traverse difficult terrain and recognize adversaries are another key Computer Vision application.

- *Industry*

For automatic inspections, the detection of faulty goods on the production line and remote inspections of pipelines and machinery, the closely allied field of machine vision has long been used in manufacturing. The technology is often used, by flagging unusual events or discrepancies, to automate and optimize organizational and control processes. Predictive maintenance, product assembly, package inspection, barcode reading for successful monitoring, text interpretation and control of robotic employees are examples of computer vision in the industry.

- *Healthcare*

Because 90% of all medical data is image-based, numerous applications for Computer Vision have emerged in the healthcare field. With a far higher degree of accuracy than medical practitioners can achieve, the system can detect anomalies in imagery derived from MRI and CAT scans.

- *Education*

Initiatives in this area have been to enhance the perception of the learner through the use of computer vision. Integrating AR assists students with various learning skills. In addition, computer vision technologies can play an important role in enhancing the efficacy of conventional classroom instruments such as books and research materials and seek to enhance expertise in specific fields.

- *Agriculture and manufacturing*

It could also support agriculture and manufacturing. There is already autonomous machinery in agriculture that uses computer vision to identify grain quality and find the best route through the crops. The use of computer vision to detect weeds in order to apply herbicides directly to particular areas is another possible agricultural use, thereby decreasing the amount of herbicides used.

- *Training and manufacturing*

Product consistency if any manufacturing process is of major concern. The quality control division plays a major role in any facility. Although these tasks have historically been performed by humans, it is now possible for computer vision to make decisions regarding quality control. Cameras and lighting capture images that are then compared to a predefined image or quality norm algorithmically, thus removing human error. AR is also present in practises that are too dangerous for humans alone, such as mining, firefighting, disposal of mines and handling of radioactive materials.

- *Automotive*

Driverless cars, which rely heavily on Computer Vision and Deep Learning, are one field that has captured the imagination of the public. Although not yet completely replacing the human driver, over the past few years, autonomous vehicle technology has progressed significantly. AI analyses data gathered from millions of motorists, learning from driver behaviour, estimating road curvature, detecting hazards, and interpreting traffic signs and signals, to automate lane finding. For example, by driving seven million miles on public roads, Waymo has trained its Computer Vision algorithms.

- *Data Processing*

Computer Vision tools and Deep Learning models must be trained to help people with recognition tasks and organizing information, requiring enormous amounts of labelled data. This is normally performed by humans, a method that is time-consuming, costly and sometimes unreliable. As Deep Learning algorithms evolve, the manual tagging process is largely replaced by an approach known as expertise crowdsourcing-the automated real-time collection and tagging of data generated by professionals as they go about their daily work.

- *Customer Experience – the Killer App*

Retail trends such as Amazon Go have recently grabbed the headlines, but Computer Vision applications and technology have been effectively incorporated into the CRM domain over the past few years, from sales and marketing to customer service and retention. Computer Vision can be a retail force multiplier, offering useful insights into consumer behaviour and helping to upsell and cross-sell. Based on visual data from smart telematics systems, a game-changer for insurance and utility firms, it can add vital details to a client's profile.

- *Self-Service & Remote Visual Assistance*

There has been a paradigm shift in the conventional customer assistance model as customers increasingly adopt smart home technologies to fit their lifestyle, entertainment, safety and security preferences. New levels of interoperability and sophistication are being generated by the IoT, and consumers will need more assistance to mount, run and manage their home computer ecosystems. Therefore, businesses will have to offer new levels of self-service, one of the most exciting emerging Machine Vision technologies, to manage rapidly growing call volumes.

- *Societal Impact of Computer Vision Technologies*

Cote and Albu (2017) looked at the social impact of computer vision technology in computer vision from the perspective of future minds: senior students in engineering. Training in engineering has historically concentrated on technological skills and expertise. The need to educate engineers in socio-technical skills and reflective thinking is now being recognized, especially on the bright and dark sides of the technology they are creating.

IV. FUTURE AND CHALLENGES

By feeding thousands of photographs of the same object presented in various ways into computers, scientists are attempting to solve this obstacle. In each image, the machine can then evaluate the pixels. They evaluate the pixel data and compare it to the pixel data from previous images when images of the same object are presented to them in the future. Even if the object is viewed in a different way, by comparing the pixels that are the same, the machine will determine if it is the same object. Computer vision studies first started in the 1950s, says Forbes. There are many applications for this technology now, and by 2022, the computer vision market is expected to reach $48.6 billion. While the most talked-about use of computer vision is often autonomous vehicles, this is far from the only use for this technology.

Deep learning, the progress of which is driven by the creation of large-scale datasets, the creation of powerful models, and the availability of vast computing resources, is the driving force behind recent developments in image recognition. Carefully developed deep neural networks have far exceeded previous approaches focused on hand-crafted image features for a number of image recognition tasks. However, despite the great success of deep learning in image recognition to date, several obstacles remain to be overcome before it can be used for wider use.

- *Improving generalization of models*

One of these difficulties is how to train models that generalize well to situations in the real world that haven't been used in training. In real-world applications, however, the test images can come from distributions of data that vary from those used in training. For example, in viewing angles, object sizes, scene configurations, and camera properties, the unseen data can differ. A recent study shows that such a gap in data delivery across a wide variety of deep network architectures can lead to significant drops in accuracy. In vital applications, such as autonomous vehicle navigation, the vulnerability of current models to natural variations in the distribution of data can be a significant drawback.

- *Exploiting data on a small and ultra-large scale*

How to properly leverage small-scale training data is another current problem. Although deep learning has shown great success in different tasks with a large amount of labeled data, if few labelled examples are available, current techniques typically break down. At the other extreme is how ultra-large-scale data can be efficiently scaled to the efficiency of recognition algorithms. The price of recognition errors is very high for important applications like autonomous driving. So massive datasets, comprising hundreds of millions of richly annotated images, are created in hopes of significantly improving the accuracy of the trained models. However, a recent study indicates that such ultra-large-scale data can not necessarily be used as effectively by current algorithms. The performance of a variety of deep networks only increases logarithmically with regard to the amount of training data on the JFT dataset containing 300 million annotated images (Figure 4). The diminishing benefits of greater large-scale training data are a major problem to solve.

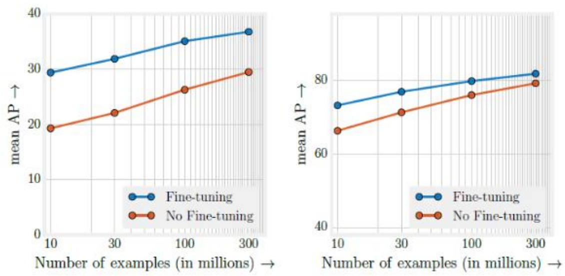

Figure 4 On the COCO minival test set, the left graph uses the mAP @[0.5, 0.95] metric, and on the PASCAL VOC 2007 test set, the right graph uses the mAP @0.5 metric.

- *Extensive understanding of scenes*

Comprehensive scene comprehension is an important subject for investigation, in addition to issues related to training data and generalization. Humans also infer object-to-object relationships, part-to-whole object hierarchies, object attributes, and 3D scene structure in addition to identifying and locating objects in a scene. Not only does this activity entail awareness of the scene, it also includes a cognitive understanding of the physical world. To achieve this goal, there is a long way to go (Figure 5).

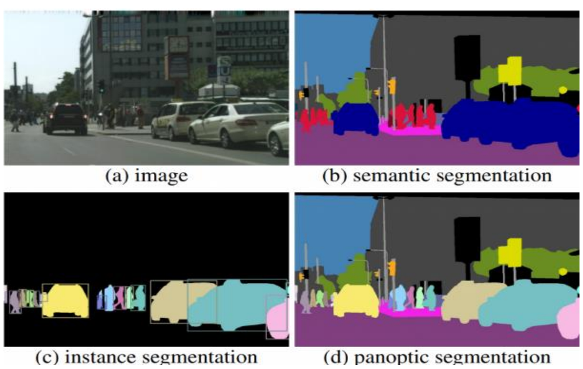

Figure 5 At many stages, a picture needs to be understood.

- *Automating engineering for networks*

The need to automate network engineering is a final challenge that we wish to note. The field has seen its focus change from the development of improved functionality to the design of new network architectures in recent years. Architecture engineering, however, is a repetitive process that deals with various hyper-parameters and choices of architecture. Tuning these elements by skilled engineers takes a huge amount of time and effort. The search space for current approaches is very limited, since they are searching for a locally optimal combination of existing network modules (e.g., convolutions of depth and connexions of identity) and are unable to discover new modules (Figure 6). It is uncertain if these existing methods are suitable for more advanced tasks.

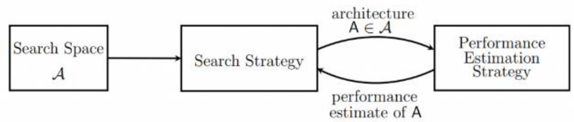

Figure 6 Neural structure search algorithm abstract diagram

We also believe in the immense potential of deep learning for image recognition, despite these challenges. Opportunities abound to solve these challenges and to drive the field forward quickly. The following explains some of these directions.

- *Embedding common sense*

The incorporation of common sense into deep learning is one important path. Deep learning is currently mainly used as a strictly data-driven process, where the network fits a non-linear function with annotated samples given in the training set, and then applies the learned function at test time to image pixels. No information is used outside of the training package.

- *Geometrical Logic*

The combined execution of image recognition and geometric reasoning is another promising path. Only 2D appearance information is considered in the leading models for image recognition. In comparison, humans interpret 3D scene layouts along with the semantic categories under which they reside. Not only binocular vision, but also geometric thinking on 2D input, such as when people look at images, can derive a 3D structure. Joint identification of images and reasoning of geometry provides shared benefits. In cases of unseen viewpoints, distortions, and appearance, the 3D architecture calculated from geometric reasoning can help direct identification. Unreasonable semantic layouts can also be removed and categories identified by their 3D form or functions can be recognized.

- *Relationship simulation*

Relational modelling holds great promise as well. It is important to model the relationships and interactions between the object entities that are present in order to understand a scene comprehensively (Figure 7 and 8). Consider two pictures, each containing an individual and a horse. If one displays the person riding the horse and the other displays the horse trampling the person, there is a totally different interpretation to what is seen in these pictures. In addition, when current deep learning methods falter due to limited data, the underlying scene structure extracted from relational modelling may help to compensate. Although efforts for this problem are already underway, the analysis is still preliminary and there is plenty of space for exploration.

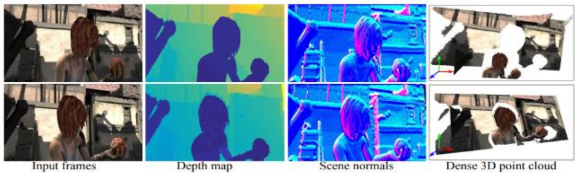

Figure 7 Rebuilding point clouds from two separate video frames of complex dynamic scenes.

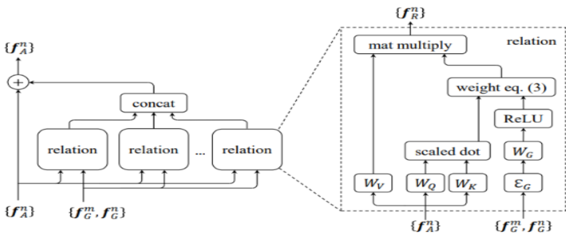

Figure 8 Relational networks in detection of targets. The external characteristics of objects are expressed by f A, and f G represents the geometric characteristics of objects.

- *Educating to learn*

Meta learning, which aims at learning the learning process, is an additional path to note. Recently, this subject has attracted considerable attention and neural architecture search can be considered as one meta learning application. However, the research on meta learning is still at an early stage, as the structures, representations, and algorithms are currently primitive for modelling the learning process. It is limited to simple combinations of existing network modules only, taking neural architecture search as an example. In order to construct novel network modules, the meta learner cannot capture the subtle intuition and sharp insight required. The full potential of automated architecture design can be unleashed with advances in meta learning, leading to network configurations that exceed those obtained through manual engineering (Figure 9).

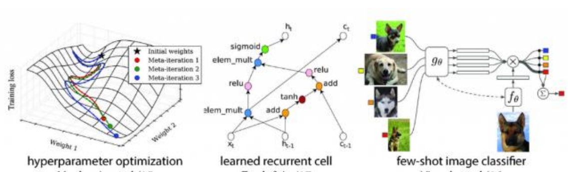

Figure 9 Latest developments in meta learning.

There are meta parametric optimization from left to right, neural structure search, and limited classification of sample images, respectively. Working on image recognition, a time full of opportunities to move the field forward and affect futuristic applications, is an exciting time. We look forward to the change that is coming soon and expect these emerging developments to change our lives in profound and incredible ways.

V. CONCLUSION

Until recently, only limited-capacity computer vision worked. The field has been able to take great strides in recent years, thanks to developments in artificial intelligence and advancements in deep learning and neural networks, and has been able to overtake humans in certain tasks related to detecting and marking objects. The amount of data we produce today that is then used to train and make computer vision better is one of the driving forces behind the growth of computer vision. The computational power needed to analyze the data is now available; along with a huge amount of visual data (more than 3 billion pictures are exchanged online every day). As new hardware and algorithms have evolved in the field of computer vision, so have the precision rates for object recognition. Today's systems have achieved 99 percent accuracy in less than a decade, from 50 percent, making them more effective than humans in responding rapidly to visual inputs. Computers are pre-programmed to solve a specific task in many computer vision applications, but learning-based approaches are also becoming increasingly popular. Hence computer vision is used in the analysis of play and strategy, and as part of real-time monitoring to determine the success and ratings of a player. Further use of brand sponsorship in sports broadcasts in the world of sports to track exposure.

REFERENCES

1. Andrew Mort, AI in Customer Service, Posted on (Mar 11, 2020). How Computer Vision Applications are changing the World. https://techsee.me/blog/computer-vision
2. Dana H. Ballard; Christopher M. Brown (1982). Computer Vision. Prentice Hall. ISBN 978-0-13-165316-0.
3. Huang, T. (1996). Vandoni, Carlo, E (ed.). Computer Vision : Evolution And Promise (PDF). 19th CERN School of Computing. Geneva: CERN. pp. 21–25. doi:10.5170/CERN-1996-008.21. ISBN 978-9290830955.
4. Milan Sonka; Vaclav Hlavac; Roger Boyle (2008). Image Processing, Analysis, and Machine Vision. Thomson. ISBN 978-0-495-08252-1.
5. Reinhard Klette (2014). Concise Computer Vision. Springer. ISBN 978-1-4471-6320-6.

6. Linda G. Shapiro; George C. Stockman (2001). Computer Vision. Prentice Hall. ISBN 978-0-13-030796-5.
7. Tim Morris (2004). Computer Vision and Image Processing. Palgrave Macmillan. ISBN 978-0-333-99451-1.
8. Bernd Jähne; Horst Haußecker (2000). Computer Vision and Applications, A Guide for Students and Practitioners. Academic Press. ISBN 978-0-13-085198-7.
9. David A. Forsyth; Jean Ponce (2003). Computer Vision, A Modern Approach. Prentice Hall. ISBN 978-0-13-085198-7.
10. Kagami, Shingo (2010). 'High-speed vision systems and projectors for real-time perception of the world'. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops. IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops. 2010. pp. 100–107.
11. Seth Colaner (January 3, 2016). 'A Third Type Of Processor For VR/AR: Movidius' Myriad 2 VPU'. www.tomshardware.com.
12. Richard Szeliski (30 September 2010). Computer Vision: Algorithms and Applications. Springer Science & Business Media. pp. 10–16. ISBN 978-1-84882-935-0.
13. Papert, Seymour (1966-07-01). 'The Summer Vision Project'. MIT AI Memos (1959 - 2004). hdl:1721.1/6125.
14. Margaret Ann Boden (2006). Mind as Machine: A History of Cognitive Science. Clarendon Press. p. 781. ISBN 978-0-19-954316-8.
15. Takeo Kanade (6 December 2012). Three-Dimensional Machine Vision. Springer Science & Business Media. ISBN 978-1-4613-1981-8.
16. Nicu Sebe; Ira Cohen; Ashutosh Garg; Thomas S. Huang (3 June 2005). Machine Learning in Computer Vision. Springer Science & Business Media. ISBN 978-1-4020-3274-5.
17. William Freeman; Pietro Perona; Bernhard Scholkopf (2008). 'Guest Editorial: Machine Learning for Computer Vision'. International Journal of Computer Vision. 77 (1): 1. doi:10.1007/s11263-008-0127-7. ISSN 1573-1405.
18. E. Roy Davies (2005). Machine Vision: Theory, Algorithms, Practicalities. Morgan Kaufmann. ISBN 978-0-12-206093-9.