Computer vision

From Wikipedia, the free encyclopedia

Computer vision is a field that includes methods for acquiring, processing, analyzing, and understanding images and, in general, high-dimensional data from the real world in order to produce numerical or symbolic information, *e.g.*, in the forms of decisions.[1][2][3][4] A theme in the development of this field has been to duplicate the abilities of human vision by electronically perceiving and understanding an image.[5] This image understanding can be seen as the disentangling of symbolic information from image data using models constructed with the aid of geometry, physics, statistics, and learning theory.[6] Computer vision has also been described as the enterprise of automating and integrating a wide range of processes and representations for vision perception.[7][8]

As a scientific discipline, computer vision is concerned with the theory behind artificial systems that extract information from images. The image data can take many forms, such as video sequences, views from multiple cameras, or multi-dimensional data from a medical scanner. As a technological discipline, computer vision seeks to apply its theories and models to the construction of computer vision systems.

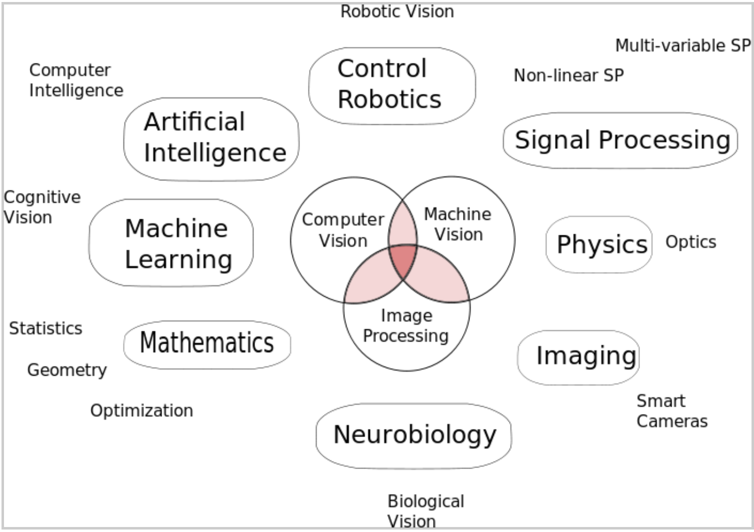
Sub-domains of computer vision include scene reconstruction, event detection, video tracking, object recognition, learning, indexing, motion estimation, and image restoration.

Contents

Related fields

Areas of artificial intelligence deal with autonomous planning or deliberation for robotical systems to navigate through an environment. A detailed understanding of these environments is required to navigate through them. Information about the environment could be provided by a computer vision system, acting as a vision sensor and providing high-level information about the environment and the robot.

Artificial intelligence and computer vision share other topics such as pattern recognition and learning techniques. Consequently, computer vision is sometimes seen as a part of the artificial intelligence field or the computer science field in general.



Relation between computer vision and various other fields

Solid-state physics is another field that is closely related to computer vision. Most computer vision systems rely on image sensors, which detect electromagnetic radiation which is typically in the form of either visible or infra-red light. The sensors are designed using quantum physics. The process by which light interacts with surfaces is explained using physics. Physics explains the behavior of optics which are a core part of most imaging systems. Sophisticated image sensors even require quantum mechanics to provide a complete understanding of the image formation process. Also, various measurement problems in physics can be addressed using computer vision, for example motion in fluids.

A third field which plays an important role is neurobiology, specifically the study of the biological vision system. Over the last century, there has been an extensive study of eyes, neurons, and the brain structures devoted to processing of visual stimuli in both humans and various animals. This has led to a coarse, yet complicated, description of how "real" vision systems operate in order to solve certain vision related tasks. These results have led to a subfield within computer vision where artificial systems are designed to mimic the processing and behavior of biological systems, at different levels of complexity. Also, some of the learning-based methods developed within computer vision (*e.g.* neural net and deep learning based image and feature analysis and classification) have their background in biology.

Some strands of computer vision research are closely related to the study of biological vision – indeed, just as many strands of AI research are closely tied with research into human consciousness, and the use of stored knowledge to interpret, integrate and utilize visual information. The field of biological vision studies and models the physiological processes behind visual perception in humans and other animals. Computer vision, on the other hand, studies and describes the processes implemented in software and hardware behind artificial vision systems. Interdisciplinary exchange between biological and computer vision has proven fruitful for both fields.

Yet another field related to computer vision is signal processing. Many methods for processing of one-variable signals, typically temporal signals, can be extended in a natural way to processing of two-variable signals or multi-variable signals in computer vision. However, because of the specific nature of images there are many methods developed within computer vision which have no counterpart in processing of one-variable signals. Together with the multi-dimensionality of the signal, this defines a subfield in signal processing as a part of computer vision.

Beside the above-mentioned views on computer vision, many of the related research topics can also be studied from a purely mathematical point of view. For example, many methods in computer vision are based on statistics, optimization or geometry. Finally, a significant part of the field is devoted to the implementation aspect of computer vision; how existing methods can be realized in various combinations of software and hardware, or how these methods can be modified in order to gain processing speed without losing too much performance.

The fields most closely related to computer vision are image processing, image analysis and machine vision. There is a significant overlap in the range of techniques and applications that these cover. This implies that the basic techniques that are used and developed in these fields are more or less identical, something which can be interpreted as there is only one field with different names. On the other hand, it appears to be necessary for research groups, scientific journals, conferences and companies to present or market themselves as belonging specifically to one of these fields and, hence, various characterizations which distinguish each of the fields from the others have been presented.

Computer vision is, in some ways, the inverse of computer graphics. While computer graphics produces image data from 3D models, computer vision often produces 3D models from image data. There is also a trend towards a combination of the two disciplines, *e.g.*, as explored in augmented reality.

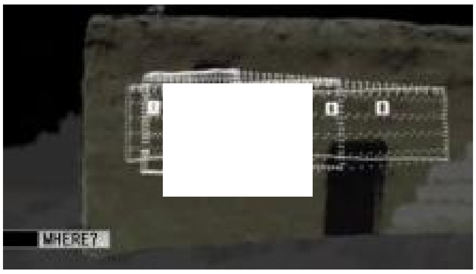
The following characterizations appear relevant but should not be taken as universally accepted:

- Image processing and image analysis tend to focus on 2D images, how to transform one image to another, *e.g.*, by pixel-wise operations such as contrast enhancement, local operations such as edge extraction or noise removal, or geometrical transformations such as rotating the image. This characterization implies that image processing/analysis neither require assumptions nor produce interpretations about the image content.
- Computer vision includes 3D analysis from 2D images. This analyzes the 3D scene projected onto one or several images, *e.g.*, how to reconstruct structure or other information about the 3D scene from one or several images. Computer vision often relies on more or less complex assumptions about the scene depicted in an image.
- Machine vision is the process of applying a range of technologies & methods to provide imaging-based automatic inspection, process control and robot guidance[9] in industrial applications.[10] Machine vision tends to focus on applications, mainly in manufacturing, *e.g.*, vision based autonomous robots and systems for vision based inspection or measurement. This implies that image sensor technologies and control theory often are integrated with the processing of image data to control a robot and that real-time processing is emphasised by means of efficient implementations in hardware and software. It also implies that the external conditions such as lighting can be and are often more controlled in machine vision than they are in general computer vision, which can enable the use of different algorithms.
- There is also a field called imaging which primarily focus on the process of producing images, but sometimes also deals with processing and analysis of images. For example, medical imaging includes substantial work on the analysis of image data in medical applications.
- Finally, pattern recognition is a field which uses various methods to extract information from signals in general, mainly based on statistical approaches and artificial neural networks. A significant part of this field is devoted to applying these methods to image data.

Applications for computer vision

Applications range from tasks such as industrial machine vision systems which, say, inspect bottles speeding by on a production line, to research into artificial intelligence and computers or robots that can comprehend the world around them. The computer vision and machine vision fields have significant overlap. Computer vision covers the core technology of automated image analysis which is used in many fields. Machine vision usually refers to a process of combining automated image analysis with other methods and technologies to provide automated inspection and robot guidance in industrial applications. In many computer vision applications, the computers are pre-programmed to solve a particular task, but methods based on learning are now becoming increasingly common. Examples of applications of computer vision include systems for:

- Controlling processes, *e.g.*, an industrial robot;
- Navigation, *e.g.*, by an autonomous vehicle or mobile robot;
- Detecting events, *e.g.*, for visual surveillance or people counting;
- Organizing information, *e.g.*, for indexing databases of images and image sequences;
- Modeling objects or environments, *e.g.*, medical image analysis or topographical modeling;
- Interaction, *e.g.*, as the input to a device for computer-human interaction, and
- Automatic inspection, *e.g.*, in manufacturing applications.



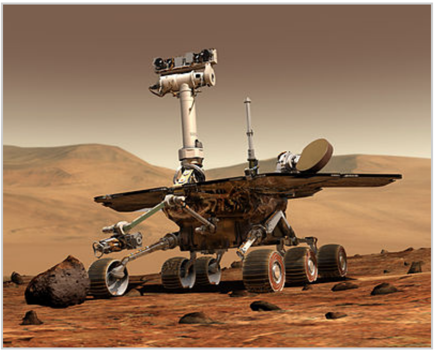
DARPA's Visual Media Reasoning concept video

One of the most prominent application fields is medical computer vision or medical image processing. This area is characterized by the extraction of information from image data for the purpose of making a medical diagnosis of a patient. Generally, image data is in the form of microscopy images, X-ray images, angiography images, ultrasonic images, and tomography images. An example of information which can be extracted from such image data is detection of tumours, arteriosclerosis or other malign changes. It can also be measurements of organ dimensions, blood flow, etc. This application area also supports medical research by providing new information, *e.g.*, about the structure of the brain, or about the quality of medical treatments. Applications of computer vision in the medical area also includes enhancement of images that are interpreted by humans, for example ultrasonic images or X-ray images, to reduce the influence of noise.

A second application area in computer vision is in industry, sometimes called machine vision, where information is extracted for the purpose of supporting a manufacturing process. One example is quality control where details or final products are being automatically inspected in order to find defects. Another example is measurement of position and orientation of details to be picked up by a robot arm. Machine vision is also heavily used in agricultural process to remove undesirable food stuff from bulk material, a process called optical sorting.

Military applications are probably one of the largest areas for computer vision. The obvious examples are detection of enemy soldiers or vehicles and missile guidance. More advanced systems for missile guidance send the missile to an area rather than a specific target, and target selection is made when the missile reaches the area based on locally acquired image data. Modern military concepts, such as "battlefield awareness", imply that various sensors, including image sensors, provide a rich set of information about a combat scene which can be used to support strategic decisions. In this case, automatic processing of the data is used to reduce complexity and to fuse information from multiple sensors to increase reliability.

One of the newer application areas is autonomous vehicles, which include submersibles, land-based vehicles (small robots with wheels, cars or trucks), aerial vehicles, and unmanned aerial vehicles (UAV). The level of autonomy ranges from fully autonomous (unmanned) vehicles to vehicles where computer vision based systems support a driver or a pilot in various situations. Fully autonomous vehicles typically use computer vision for navigation, i.e. for knowing where it is, or for producing a map of its environment (SLAM) and for detecting obstacles. It can also be used for detecting certain task specific events, *e.g.*, a UAV looking for forest fires. Examples of supporting systems are obstacle warning systems in cars, and systems for autonomous landing of aircraft. Several car manufacturers have demonstrated systems for autonomous driving of cars, but this technology has still not reached a level where it can be put on the market. There are ample examples of military autonomous vehicles ranging from advanced missiles, to UAVs for recon missions or missile guidance. Space exploration is already being made with autonomous vehicles using computer vision, *e.g.*, NASA's Mars Exploration Rover and ESA's ExoMars Rover.



Artist's Concept of Rover on Mars, an example of an unmanned land-based vehicle. Notice the stereo cameras mounted on top of the Rover.

Other application areas include:

- Support of visual effects creation for cinema and broadcast, *e.g.*, camera tracking (matchmoving).
- Surveillance.

Typical tasks of computer vision

Each of the application areas described above employ a range of computer vision tasks; more or less well-defined measurement problems or processing problems, which can be solved using a variety of methods. Some examples of typical computer vision tasks are presented below.

Recognition

The classical problem in computer vision, image processing, and machine vision is that of determining whether or not the image data contains some specific object, feature, or activity. This task can normally be solved robustly and without effort by a human, but is still not satisfactorily solved in computer vision for the general case – arbitrary objects in arbitrary situations. The existing methods for dealing with this problem can at best solve it only for specific objects, such as simple geometric objects (*e.g.*, polyhedra), human faces, printed or hand-written characters, or vehicles, and in specific situations, typically described in terms of well-defined illumination, background, and pose of the object relative to the camera.

Different varieties of the recognition problem are described in the literature:

- **Object recognition** – one or several pre-specified or learned objects or object classes can be recognized, usually together with their 2D positions in the image or 3D poses in the scene. Google Goggles provides a stand-alone program illustration of this function.
- **Identification** – an individual instance of an object is recognized. Examples include identification of a specific person's face or fingerprint, identification of handwritten digits, or identification of a

specific vehicle.

- **Detection** – the image data are scanned for a specific condition. Examples include detection of possible abnormal cells or tissues in medical images or detection of a vehicle in an automatic road toll system. Detection based on relatively simple and fast computations is sometimes used for finding smaller regions of interesting image data which can be further analyzed by more computationally demanding techniques to produce a correct interpretation.

Several specialized tasks based on recognition exist, such as:

- **Content-based image retrieval** – finding all images in a larger set of images which have a specific content. The content can be specified in different ways, for example in terms of similarity relative a target image (give me all images similar to image X), or in terms of high-level search criteria given as text input (give me all images which contains many houses, are taken during winter, and have no cars in them).

* **Pose estimation** – estimating the position or orientation of a specific object relative to the camera. An example application for this technique would be assisting a robot arm in retrieving objects from a conveyor belt in an assembly line situation or picking parts from a bin.

Computer vision for people counter purposes in public places, malls, shopping centres

- **Optical character recognition** (OCR) – identifying characters in images of printed or handwritten text, usually with a view to encoding the text in a format more amenable to editing or indexing (*e.g.* ASCII).
- **2D Code reading** Reading of 2D codes such as data matrix and QR codes.
- **Facial recognition**
- **Shape Recognition Technology** (SRT) in people counter systems differentiating human beings (head and shoulder patterns) from objects

Motion analysis

Several tasks relate to motion estimation where an image sequence is processed to produce an estimate of the velocity either at each points in the image or in the 3D scene, or even of the camera that produces the images . Examples of such tasks are:

- **Egomotion** – determining the 3D rigid motion (rotation and translation) of the camera from an image sequence produced by the camera.
- **Tracking** – following the movements of a (usually) smaller set of interest points or objects (*e.g.*, vehicles or humans) in the image sequence.
- **Optical flow** – to determine, for each point in the image, how that point is moving relative to the image plane, i.e., its apparent motion. This motion is a result both of how the corresponding 3D point is moving in the scene and how the camera is moving relative to the scene.

Scene reconstruction

Given one or (typically) more images of a scene, or a video, scene reconstruction aims at computing a 3D model of the scene. In the simplest case the model can be a set of 3D points. More sophisticated methods produce a complete 3D surface model. The advent of 3D imaging not requiring motion or scanning, and related processing algorithms is enabling rapid advances in this field. Grid-based 3D sensing can be used to acquire 3D images from multiple angles. Algorithms are now available to stitch multiple 3D images together into point clouds and 3D models.

Image restoration

The aim of image restoration is the removal of noise (sensor noise, motion blur, etc.) from images. The simplest possible approach for noise removal is various types of filters such as low-pass filters or median filters. More sophisticated methods assume a model of how the local image structures look like, a model which distinguishes them from the noise. By first analysing the image data in terms of the local image structures, such as lines or edges, and then controlling the filtering based on local information from the analysis step, a better level of noise removal is usually obtained compared to the simpler approaches.

An example in this field is the inpainting.

Computer vision system methods

The organization of a computer vision system is highly application dependent. Some systems are stand-alone applications which solve a specific measurement or detection problem, while others constitute a sub-system of a larger design which, for example, also contains sub-systems for control of mechanical actuators, planning, information databases, man-machine interfaces, etc. The specific implementation of a computer vision system also depends on if its functionality is pre-specified or if some part of it can be learned or modified during operation. Many functions are unique to the application. There are, however, typical functions which are found in many computer vision systems.

- **Image acquisition** – A digital image is produced by one or several image sensors, which, besides various types of light-sensitive cameras, include range sensors, tomography devices, radar, ultra-sonic cameras, etc. Depending on the type of sensor, the resulting image data is an ordinary 2D image, a 3D volume, or an image sequence. The pixel values typically correspond to light intensity in one or several spectral bands (gray images or colour images), but can also be related to various physical measures, such as depth, absorption or reflectance of sonic or electromagnetic waves, or nuclear magnetic resonance.[11]
- **Pre-processing** – Before a computer vision method can be applied to image data in order to extract some specific piece of information, it is usually necessary to process the data in order to assure that it satisfies certain assumptions implied by the method. Examples are
  - Re-sampling in order to assure that the image coordinate system is correct.
  - Noise reduction in order to assure that sensor noise does not introduce false information.
  - Contrast enhancement to assure that relevant information can be detected.
  - Scale space representation to enhance image structures at locally appropriate scales.
- **Feature extraction** – Image features at various levels of complexity are extracted from the image

data.[11] Typical examples of such features are
    - Lines, edges and ridges.
    - Localized interest points such as corners, blobs or points.

  More complex features may be related to texture, shape or motion.

- **Detection/segmentation** – At some point in the processing a decision is made about which image points or regions of the image are relevant for further processing.[11] Examples are
    - Selection of a specific set of interest points
    - Segmentation of one or multiple image regions which contain a specific object of interest.
- **High-level processing** – At this step the input is typically a small set of data, for example a set of points or an image region which is assumed to contain a specific object.[11] The remaining processing deals with, for example:
    - Verification that the data satisfy model-based and application specific assumptions.
    - Estimation of application specific parameters, such as object pose or object size.
    - Image recognition – classifying a detected object into different categories.
    - Image registration – comparing and combining two different views of the same object.
- **Decision making** Making the final decision required for the application,[11] for example:
    - Pass/fail on automatic inspection applications
    - Match / no-match in recognition applications
    - Flag for further human review in medical, military, security and recognition applications

Computer vision hardware

There are many kind of computer vision systems, nevertheless all of them contains these basic elements: power source, at least one image acquisition device (i.e. camera, ccd, etc.), processor as well as control and communication cables or some kind of wireless interconnection mechanism. In addition a practical vision system contains software for application and develop as well as a display in order to monitor what the system does. Vision system for inner spaces, as most industrial ones, contains in addition an illumination system and in most cases a controlled environment, specially on external lighting. Furthermore, a completed system includes many accessories like camera supports, cables and connectors.

See also

- AI effect
- Applications of artificial intelligence
- Machine vision glossary
- Artificial neural networks
- Deep Learning

Lists

- List of computer vision topics
- List of emerging technologies
- Outline of artificial intelligence

References

1. ^ Reinhard Klette (2014). *Concise Computer Vision*. Springer. ISBN 978-1-4471-6320-6.
2. ^ Linda G. Shapiro and George C. Stockman (2001). *Computer Vision*. Prentice Hall. ISBN 0-13-030796-3.
3. ^ Tim Morris (2004). *Computer Vision and Image Processing*. Palgrave Macmillan. ISBN 0-333-99451-5.
4. ^ Bernd Jähne and Horst Haußecker (2000). *Computer Vision and Applications, A Guide for Students and Practitioners*. Academic Press. ISBN 0-13-085198-1.
5. ^ Milan Sonka, Vaclav Hlavac and Roger Boyle (2008). *Image Processing, Analysis, and Machine Vision*. Thomson. ISBN 0-495-08252-X.
6. ^ David A. Forsyth and Jean Ponce (2003). *Computer Vision, A Modern Approach*. Prentice Hall. ISBN 0-13-085198-1.
7. ^ Dana H. Ballard and Christopher M. Brown (1982). *Computer Vision* (http://homepages.inf.ed.ac.uk/rbf/BOOKS/BANDB/bandb.htm). Prentice Hall. ISBN 0-13-165316-4.
8. ^ Barghout, Lauren, and Jacob Sheynin. "Real-world scene perception and perceptual organization: Lessons from Computer Vision." Journal of Vision 13.9 (2013): 709-709.
9. ^ Turek, Fred (June 2011). "Machine Vision Fundamentals, How to Make Robots See". *NASA Tech Briefs magazine* **35** (6). pages 60–62
10. ^ Steger, Carsten, Markus Ulrich, and Christian Wiedemann (2008). *Machine Vision Algorithms and Applications* (http://books.google.com/books?id=bvSgjky9lBYC&lpg=PP1&pg=PA1#v=onepage&q&f=false). Weinheim: Wiley-VCH. p. 1. ISBN 978-3-527-40734-7. Retrieved 2010-11-05.
11. ^ ***a b c d e*** E. Roy Davies (2005). *Machine Vision: Theory, Algorithms, Practicalities*. Morgan Kaufmann. ISBN 0-12-206093-8.

Further reading

- David Marr (1982). *Vision*. W. H. Freeman and Company. ISBN 0-7167-1284-9.
- Azriel Rosenfeld and Avinash Kak (1982). *Digital Picture Processing*. Academic Press. ISBN 0-12-597301-2.
- Barghout, Lauren, and Lawrence W. Lee (2003). *Perceptual information processing system*. U.S. Patent Application 10/618,543. ISBN 0-262-08159-8.
- Berthold Klaus Paul Horn (1986). *Robot Vision*. MIT Press. ISBN 0-262-08159-8.
- Olivier Faugeras (1993). *Three-Dimensional Computer Vision, A Geometric Viewpoint*. MIT Press. ISBN 0-262-06158-9.
- Tony Lindeberg (1994). *Scale-Space Theory in Computer Vision* (http://www.nada.kth.se/~tony/book.html). Springer. ISBN 0-7923-9418-6.
- James L. Crowley and Henrik I. Christensen (Eds.) (1995). *Vision as Process*. Springer-Verlag. ISBN 3-540-58143-X.

- Gösta H. Granlund and Hans Knutsson (1995). *Signal Processing for Computer Vision*. Kluwer Academic Publisher. ISBN 0-7923-9530-1.
- Reinhard Klette, Karsten Schluens and Andreas Koschan (1998). *Computer Vision – Three-Dimensional Data from Images* (http://www.cs.auckland.ac.nz/~rklette/Books/SpringerCV98/Springer98.html). Springer, Singapore. ISBN 981-3083-71-9.
- Emanuele Trucco and Alessandro Verri (1998). *Introductory Techniques for 3-D Computer Vision*. Prentice Hall. ISBN 0-13-261108-2.
- Bernd Jähne (2002). *Digital Image Processing*. Springer. ISBN 3-540-67754-2.
- Richard Hartley and Andrew Zisserman (2003). *Multiple View Geometry in Computer Vision*. Cambridge University Press. ISBN 0-521-54051-8.
- Gérard Medioni and Sing Bing Kang (2004). *Emerging Topics in Computer Vision*. Prentice Hall. ISBN 0-13-101366-1.
- R. Fisher, K Dawson-Howe, A. Fitzgibbon, C. Robertson, E. Trucco (2005). *Dictionary of Computer Vision and Image Processing*. John Wiley. ISBN 0-470-01526-8.
- Nikos Paragios and Yunmei Chen and Olivier Faugeras (2005). *Handbook of Mathematical Models in Computer Vision* (http://www.mas.ecp.fr/vision/Personnel/nikos/paragios-chen-faugeras/). Springer. ISBN 0-387-26371-3.
- Wilhelm Burger and Mark J. Burge (2007). *Digital Image Processing: An Algorithmic Approach Using Java* (http://www.imagingbook.com/). Springer. ISBN 1-84628-379-5.
- Pedram Azad, Tilo Gockel, Rüdiger Dillmann (2008). *Computer Vision – Principles and Practice* (http://ivt.sourceforge.net/book.html). Elektor International Media BV. ISBN 0-905705-71-8.
- Richard Szeliski (2010). *Computer Vision: Algorithms and Applications* (http://szeliski.org/Book/). Springer-Verlag. ISBN 978-1848829343.
- J. R. Parker (2011). *Algorithms for Image Processing and Computer Vision (2nd ed.)* (http://www.amazon.ca/Algorithms-Image-Processing-Computer-Vision/dp/0470643854). Wiley. ISBN 978-0470643853.

External links

- USC Iris computer vision conference list (http://iris.usc.edu/Information/Iris-Conferences.html)
- Computer vision papers on the web (http://www.cvpapers.com/index.html) A complete list of papers of the most relevant computer vision conferences.
- Computer Vision Online (http://www.computervisiononline.com/) News, source code, datasets and job offers related to computer vision.
- Keith Price's Annotated Computer Vision Bibliography (http://iris.usc.edu/Vision-Notes/bibliography/contents.html)
- CVonline (http://homepages.inf.ed.ac.uk/rbf/CVonline/) Bob Fisher's Compendium of Computer Vision.

- The British Machine Vision Association (http://www.bmva.org/) Supporting computer vision research within the UK via the BMVC and MIUA conferences, Annals of the BMVA open-source journal, BMVA Summer School and one-day meetings

Categories: Artificial intelligence | Image processing | Computer vision | Packaging machinery

---