Robert B. Fisher
Toby P. Breckon
Kenneth Dawson-Howe
Andrew Fitzgibbon
Craig Robertson
Emanuele Trucco
Christopher K. I. Williams

Dictionary of COMPUTER VISION and IMAGE PROCESSING

Second Edition

WILEY

Contents

Dictionary of Computer Vision and Image Processing

Second Edition

R. B. Fisher
University of Edinburgh, UK

T. P. Breckon
Durham University, UK

K. Dawson-Howe
Trinity College Dublin, Ireland

A. Fitzgibbon
Microsoft Research, UK

C. Robertson
Epipole Ltd., UK

E. Trucco
University of Dundee, UK

C. K. I. Williams
University of Edinburgh, UK

WILEY

This edition first published 2014

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

Library of Congress Cataloging-in-Publication Data

Dictionary of computer vision and image processing / R. B. Fisher, T. P. Breckon,
K. Dawson-Howe, A. Fitzgibbon, C. Robertson, E. Trucco, C. K. I. Williams. – 2nd edition.
pages cm
Includes bibliographical references.
ISBN 978-1-119-94186-6 (pbk.)
1. Computer vision–Dictionaries. 2. Image processing–Dictionaries. I. Fisher, R. B.
TA1634.I45 2014
006.3′703–dc23

2013022869

A catalogue record for this book is available from the British Library.

ISBN: 9781119941866

From Bob to Rosemary, Mies, Hannah, Phoebe and Lars

From Toby to Alison, my parents and Amy

From Ken to Jane, William and Susie

From AWF to Liz, to my parents, and again to D

From Craig to Karen, Aidan and Caitlin

From Manuel to Emily, Francesca, and Alistair

Preface

This dictionary arose out of a continuing interest in the resources needed by students and researchers in the fields of image processing, computer vision and machine vision (however you choose to define these overlapping fields). As instructors and mentors, we often found confusion about what various terms and concepts mean for the beginner. To support these learners, we have tried to define the key concepts that a competent generalist should know about these fields.

This second edition adds approximately 1000 new terms to the more than 2500 terms in the original dictionary. We have chosen new terms that have entered reasonably common usage (e.g., those which have appeared in the index of influential books) and terms that were not included originally. We are pleased to welcome Toby Breckon and Chris Williams into the authorial team and to thank Andrew Fitzgibbon and Manuel Trucco for all their help with the first edition.

One innovation in the second edition is the addition of reference links for a majority of the old and new terms. Unlike more traditional dictionaries, which provide references to establish the origin or meaning of the word, our goal here was instead to provide further information about the term.

Another innovation is to include a few videos for the electronic version of the dictionary.

This is a dictionary, not an encyclopedia, so the definitions are necessarily brief and are not intended to replace a proper textbook explanation of the term. We have tried to capture the essentials of the terms, with short examples or mathematical precision where feasible or necessary for clarity.

Further information about many of the terms can be found in the references. Many of the references are to general textbooks, each providing a broad view of a portion of the field. Some of the concepts are quite recent; although commonly used in research publications, they may not yet have appeared in mainstream textbooks. Subsequently, this book is also a useful source for recent terminology and concepts. Some concepts are still missing from the dictionary, but we have scanned textbooks and the research literature to find the central and commonly used terms.

The dictionary was intended for beginning and intermediate students and researchers, but as we developed the dictionary it was clear that we also had some confusions and vague understandings of the concepts. It surprised us that some terms had multiple usages. To improve quality and coverage, each definition was reviewed during development by at least two people besides its author. We hope that this has caught any errors and vagueness, as well as providing alternative meanings. Each of the co-authors is quite experienced in the topics covered here, but it was still educational to learn more about our field in the process of compiling the dictionary. We hope that you find using the dictionary equally valuable.

To help the reader, terms appearing elsewhere in the dictionary are underlined in the definitions. We have tried to be reasonably thorough about this, but some terms, such as 2D, 3D, light, camera, image, pixel, and color were so commonly used that we decided not to cross-reference all of them.

We have tried to be consistent with the mathematical notation: italics for scalars (s), arrowed italics for points and vectors $(\vec{v})$, and bold for matrices (**M**).

The authors would like to thank Xiang (Lily) Li, Georgios Papadimitriou, and Aris Valtazanos for their help with finding citations for the content from the first edition. We also greatly appreciate all the support from the John Wiley & Sons editorial and production team!

Numbers

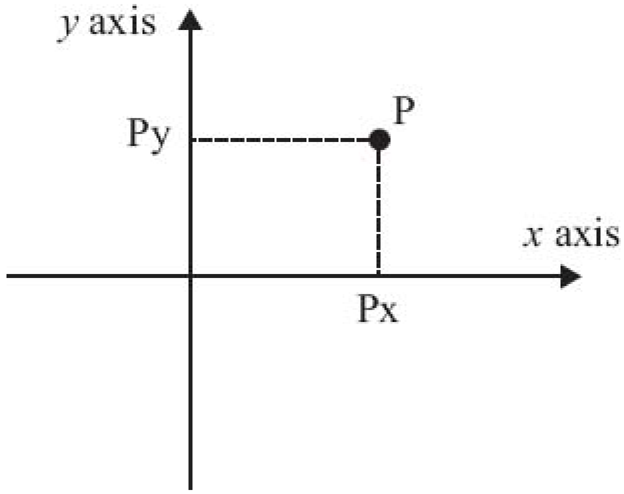
1D: One dimensional, usually in reference to some structure. Examples include: a signal *x*(*t*) that is a function of time *t*; the dimensionality of a single property value; and one degree of freedom in shape variation or motion. [Hec87:2.1]

1D projection: The projection of data from a higher dimension to a single dimensional representation (line).

1-norm: A specific case of the *p*-norm, the sum of the absolute values of the entries of a given vector $\vec{x},\ \|\vec{x}\|_1 = \sum_{i=0}^{n-1} |\vec{x}_i|$, of length *n*. Also known as the taxicab (Manhattan) norm or the L1 norm. [Sho07]

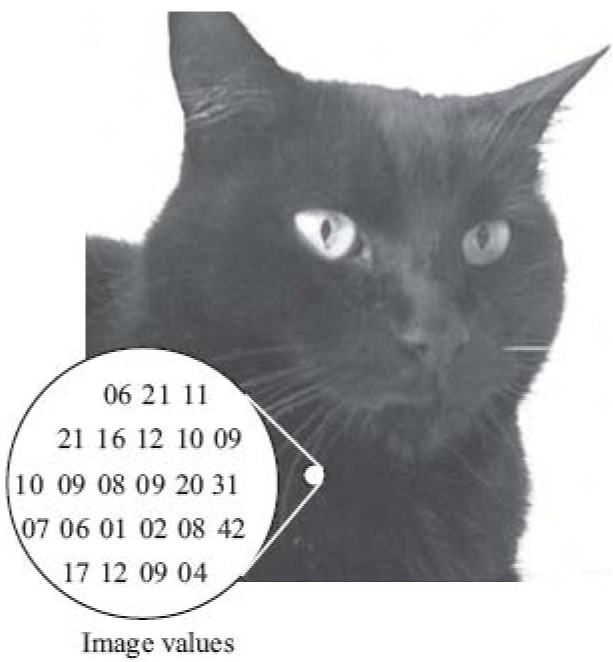
2D: Two dimensional. A space describable using any pair of orthogonal basis vectors consisting of two elements. [WP: Two-dimensional_space]

2D coordinate system: A system uniquely associating two real numbers to any point of a plane. First, two intersecting lines (axes) are chosen on the plane, usually perpendicular to each other. The point of intersection is the origin of the system. Second, metric units are established on each axis (often the same for both axes) to associate numbers to points. The coordinates P_x and P_y of a point, *P*, are obtained by projecting *P* onto each axis in a direction parallel to the other axis and reading the numbers at the intersections: [JKS95:1.4]



2D Fourier transform: A special case of the general Fourier transform often used to find structures in images. [FP03:7.3.1]

2D image: A matrix of data representing samples taken at discrete intervals. The data may be from a variety of sources and sampled in a variety of ways. In computer vision applications, the image values are often encoded color or monochrome intensity samples taken by digital cameras but may also be range data. Some typical intensity values are: [SQ04:4.1.1]



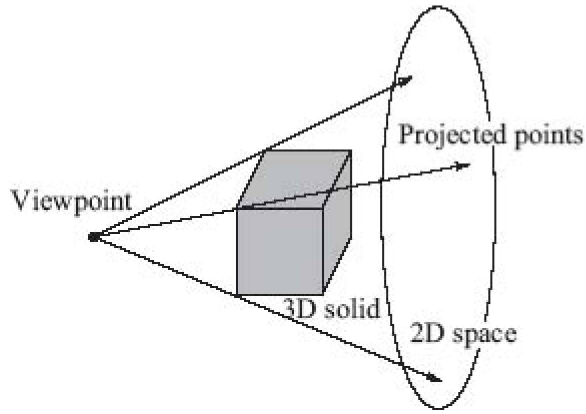
2D input device: A device for sampling light intensity from the real world into a 2D matrix of measurements. The most popular two-dimensional imaging device is the charge-coupled device (CCD) camera. Other common devices are flatbed scanners and X-ray scanners. [SQ04:4.2.1]

2D point: A point in a 2D space, i.e., characterized by two coordinates; most often, a point on a plane, e.g., an image point in pixel coordinates. Notice, however, that two coordinates do not necessarily imply a plane: a point on a 3D surface can be expressed either in 3D coordinates or by two coordinates given a surface parameterization (see surface patch). [JKS95:1.4]

2D point feature: Localized structures in a 2D image, such as interest points, corners and line meeting points (e.g., X, Y and T shaped). One detector for these features is the SUSAN corner finder. [TV98:4.1]

2D pose estimation: A special case of 3D pose estimation. A fundamental open problem in computer vision where the correspondence between two sets of 2D points is found. The problem is defined as follows: Given two sets of points $\{\vec{x}_j\}$ and $\{\vec{y}_k\}$, find the Euclidean transformation $\{\mathbf{R}, \vec{t}\}$ (the pose) and the match matrix $\{\mathbf{M}_{jk}\}$ (the correspondences) that best relates them. A large number of techniques has been used to address this problem, e.g., tree-pruning methods, the Hough transform and geometric hashing. [HJL+89]

2D projection: A transformation mapping higher dimensional space onto two-dimensional space. The simplest method is to simply discard higher dimensional coordinates, although generally a viewing position is used and the projection is performed.



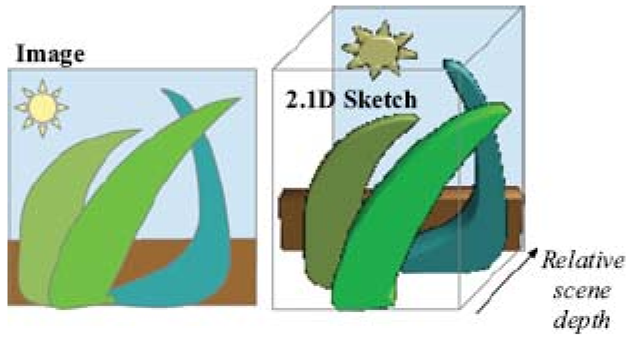
For example, the main steps for a computer graphics projection are as follows: apply normalizing transform to 3D point world coordinates; clip against canonical view volume; project onto projection plane; transform into viewport in 2D device coordinates for display. Commonly used projection functions are parallel projection and perspective projection. [JKS95:1.4]

2D shape descriptor (local): A compact summary representation of object shape over a localized region of an image. See shape descriptor. [Blu67]

2D shape representation (global): A compact summary representation of image shape features over the entire image. See shape representation. [FP03:28.3]

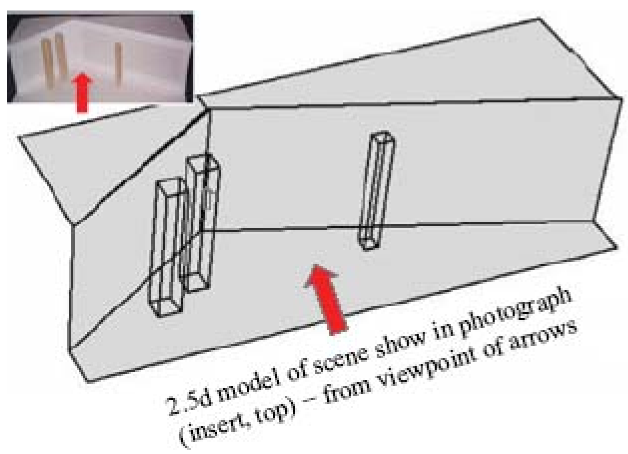
2D view: Planar aspect view or planar projected view (such as an image under perspective projection) such that positions within its spatial representation can be indexed in two dimensions. [SB11:2.3.1]

2.1D sketch: A lesser variant of the established 2.5D sketch, which captures the relative depth ordering of (possibly self-occluding) scene regions in terms of their front-to-back relationship within the scene. By contrast, the 2.5D sketch captures the relative scene depth of regions, rather than merely depth ordering: [NM90]



2.5D image: A range image obtained by scanning from a single viewpoint. It allows the data to be represented in a single image array, where each pixel value encodes the distance to the observed scene. The reason this is not called a 3D image is to make explicit the fact that the back sides of the scene objects are not represented. [SQ04:4.1.1]

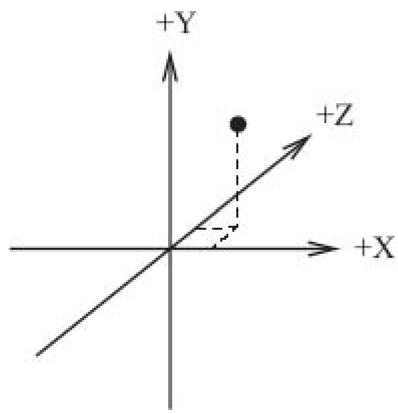
2.5D model: A geometric model representation corresponding to the 2.5D image representation used in the model to (image) data matching problem of model-based recognition: [Mar82] An example model is:



2.5D sketch: Central structure of Marr's Theory of vision. An intermediate description of a scene indicating the visible surfaces and their arrangement with respect to the viewer. It is built from several different elements: the contour, texture and shading information coming from the primal sketch, stereo information and motion. The description is theorized to be a kind of buffer where partial resolution of the objects takes place. The name 2.5D sketch stems from the fact that, although local changes in depth and discontinuities are well resolved, the absolute distance to all scene points may remain unknown. [FP03:11.3.2]

3D: Three dimensional. A space describable using any triple of mutually orthogonal basis vectors consisting of three elements. [WP: Three-dimensional_space]

3D coordinate system: Same as 2D coordinate system but in three dimensions: [JKS95:1.4]



3D data: Data described in all three spatial dimensions. See also range data, CAT and NMR. [WP: 3D_data_acquisition_and_object_reconstruction] An example of a 3D data set is:

3D data acquisition: Sampling data in all three spatial dimensions. There is a variety of ways to perform this sampling, e.g., using structured light triangulation. [FP03:21.1]

3D image: See range image.

3D imaging: Any of a class of techniques that obtain three-dimensional information using imaging equipment. Active vision techniques generally include a source of structured light (or other electromagnetic or sonar radiation) and a sensor, such as a camera or a microphone. Triangulation and time-of-flight computations allow the distance from the sensor system to be computed. Common technologies include laser scanning, texture projection systems and moiré fringe methods. Passive sensing in 3D depends only on external (and hence unstructured) illumination sources. Examples of such systems are stereo reconstruction and shape from focus techniques. See also 3D surface imaging and 3D volumetric imaging. [FMN+91]

3D interpretation: A 3D model, e.g., a solid object that explains an image or a set of image data. For instance, a certain configuration of image lines can be explained as the perspective projection of a polyhedron; in simpler words, the image lines are the images of some of the polyhedron's lines. See also image interpretation. [BB82:9.1]

3D model: A description of a 3D object that primarily describes its shape. Models of this sort are regularly used as exemplars in model-based recognition and 3D computer graphics. [TV98:10.6]

3D model-based tracking: An extension of model-based tracking using a 3D model of the tracked object. [GX11:5.1.4]

3D moments: A special case of moment where the data comes from a set of 3D points. [GC93]

3D motion estimation: An extension of motion estimation whereby the motion is estimated as a displacement vector $\vec{d}$ in $\mathbb{R}^3$. [LRF93]

3D motion segmentation: An extension to motion segmentation whereby motion is segmented within an $\mathbb{R}^3$ dataset. [TV07]

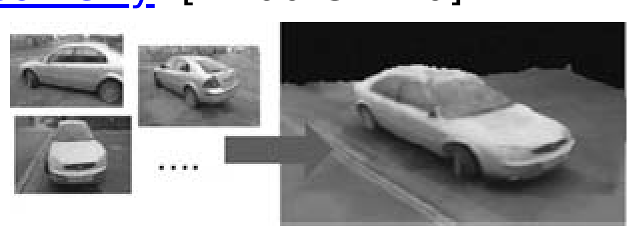
3D object: A subset of $\mathbb{R}^3$. In computer vision, often taken to mean a volume in $\mathbb{R}^3$ that is bounded by a surface. Any solid object around you is an example: table, chairs, books, cups; even yourself. [BB82:9.1]

3D point: An infinitesimal volume of 3D space. [JKS95:1.4]

3D point feature: A point feature on a 3D object or in a 3D environment. For instance, a corner in 3D space. [RBB09]

3D pose estimation: The process of determining the transformation (translation and rotation) of an object in one coordinate frame with respect to another coordinate frame. Generally, only rigid objects are considered; models of those objects exist *a priori* and we wish to determine the position of the object in an image on the basis of matched features. This is a fundamental open problem in computer vision where the correspondence between two sets of 3D points is found. The problem is defined as follows: Given two sets of points $\{\vec{x}_j\}$ and $\{\vec{y}_k\}$, find the parameters of a Euclidean transformation $\{\mathbf{R}, \vec{t}\}$ (the pose) and the match matrix $\{\mathbf{M}_{jk}\}$ (the correspondences) that best relates them. Assuming the points correspond, they should match exactly under this transformation. [TV98:11.2]

3D reconstruction: The recovery of 3D scene information and organization into a 3D shape via e.g., multi-view geometry: [HZ00:Ch. 10]

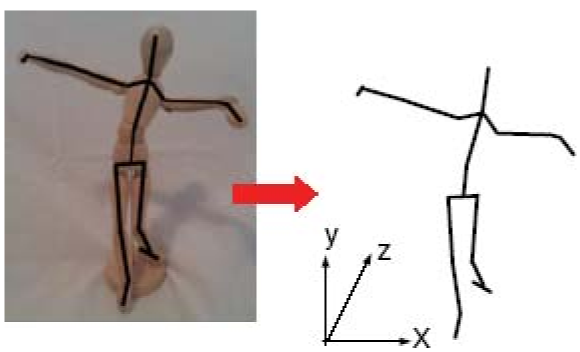


3D shape descriptor: An extension to regular shape descriptor approaches to consider object shape in $\mathbb{R}^3$. [Pri12:Ch. 17]

3D shape representation: A compact summary representation of shape extending shape representation to consider object shape in R^3. [Pri12:Ch. 17]

3D SIFT: A 3D extension of the SIFT operator defined for use over voxel data. [FBM10]

3D skeleton: A 3D extension of an image skeleton defining a tree-like structure of the medial axes of a 3D object (akin to the form of a human stick figure in the case of considering a person as a 3D object). See also medial axis skeletonization: [Sze10:12.6] See example below:



3D stratigraphy: A modeling and visualization tool used to display different underground layers. Often used for visualizations of archaeological sites or for detecting rock and soil structures in geological surveying. [PKVG00]

3D structure recovery: See 3D reconstruction.

3D SURF: A 3D extension to the SURF descriptor that considers the characterization of local image regions in $\mathbb{R}^3$ via either a volumetric voxel-based or a surface-based representation. [KPW+10]

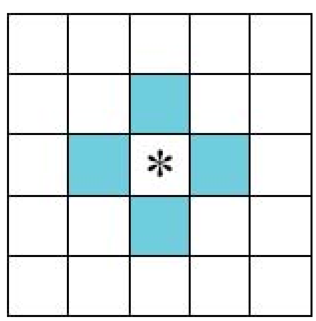
3D surface imaging: Obtaining surface information embedded in a 3D space. See also 3D imaging and 3D volumetric imaging.

3D texture: The appearance of texture on a 3D surface when imaged, e.g., the fact that the density of texels varies with distance because of perspective effects. 3D surface properties (e.g., shape, distances and orientation) can be estimated from such effects. See also shape from texture and texture orientation. [DN99]

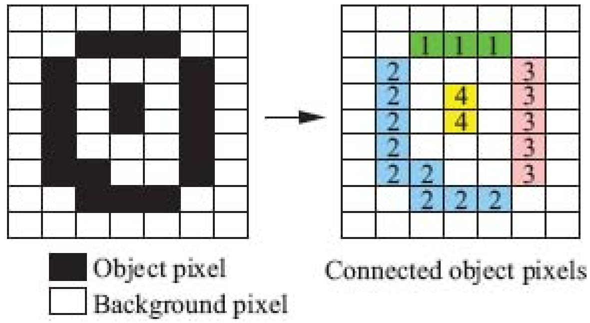
3D vision: A branch of computer vision dealing with characterizing data composed of 3D measurements. This may involve segmentation of the data into individual surfaces that are then used to identify the data as one of several models. Reverse engineering is a specialism in 3D vision. [Dav90:16.2]

3D volumetric imaging: Obtaining measurements of scene properties at all points in a 3D space, including the insides of objects. This is used for inspection but more commonly for medical imaging. Techniques include nuclear magnetic resonance, computerized tomography, positron emission tomography and single photon emission computed tomography. See also 3D imaging and 3D surface imaging.

4 connectedness: A type of image connectedness in which each rectangular pixel is considered to be connected to the four neighboring pixels that share a common crack edge. See also 8 connectedness: [SQ04:4.5] Four pixels connected to a central pixel (*):

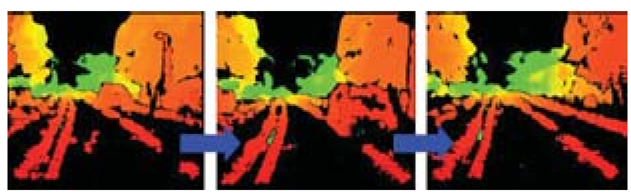


Four groups of pixels joined by 4 connectedness:

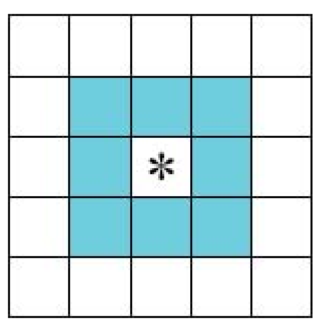


4D approach: An approach or solution to a given problem that utilizes both 3D-spatial and temporal information. See 4D representation (3D-spatial + time).

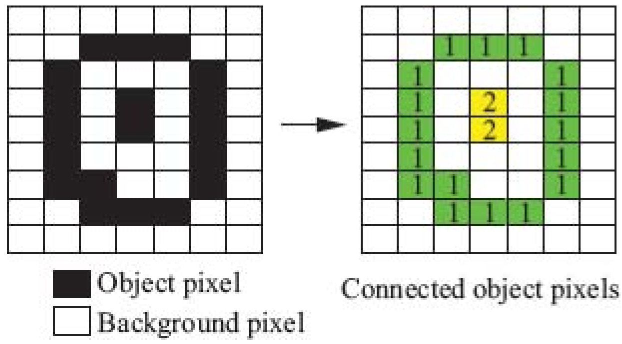
4D representation (3D-spatial + time): A 3D times series data representation whereby 3D scene information is available over a temporal sequence. An example would be a video sequence obtained from stereo vision or some other form of depth sensing: [RG08:Ch. 2]



8 connectedness: A type of image connectedness in which each rectangular pixel is considered to be connected to all eight neighboring pixels. See also 4 connectedness: [SQ04:4.5] Eight pixels connected to a central pixel (*):



Two groups of pixels joined by 8 connectedness:



8-point algorithm: An approach for the recovery of the fundamental matrix using a set of eight feature point correspondences for stereo camera calibration. [HZ00:11.2]

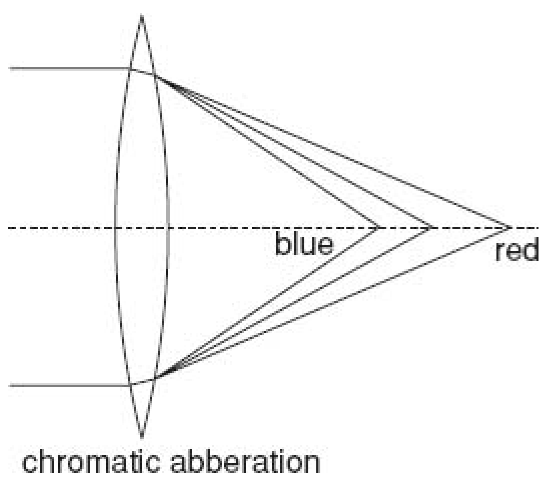
A

A*: A search technique that performs best-first searching based on an evaluation function that combines the cost so far and the estimated cost to the goal. [WP:A*_search_algorithm]

***a posteriori* probability**: Literally, “after” probability. It is the probability $p(s|e)$ that some situation s holds after some evidence e has been observed. This contrasts with the *a priori* probability, $p(s)$, the probability of s before any evidence is observed. Bayes’ rule is often used to compute the *a posteriori* probability from the *a priori* probability and the evidence. See also posterior distribution. [JKS95:15.5]

***a priori* probability**: A probability distribution that encodes an agent’s beliefs about some uncertain quantity before some evidence or data is taken into account. See also prior distribution. [Bis06:1.2.3]

aberration: Problem exhibited by a lens or a mirror whereby unexpected results are obtained. Two types of aberration are commonly encountered: chromatic aberration, where different frequencies of light focus at different positions:



chromatic abberation

and spherical aberration, where light passing through the edges of a lens (or mirror) focuses at slightly different positions. [FP03:1.2.3]

absolute conic: The conic in 3D projective space that is the intersection of the unit (or any) sphere with the plane at infinity. It consists only of complex points. Its importance in computer vision is because of its role in the problem of autocalibration: the image of the absolute conic (IAC), a 2D conic, is represented by a 3× 3 matrix ω that is the inverse of the matrix $\mathbf{K}\,\mathbf{K}^{\top}$, where $\mathbf{K}$ is the matrix of the internal parameters for camera calibration. Subsequently, identifying ω allows the camera calibration to be computed. [FP03:13.6]

absolute coordinates: Generally used in contrast to *local* or *relative* coordinates. A coordinate system that is referenced to some external datum. For example, a pixel in a satellite image might be at (100, 200) in image coordinates, but at (51:48:05N, 8:17:54W) in georeferenced absolute coordinates. [JKS95:1.4.2]

absolute orientation: In photogrammetry, the problem of registration of two corresponding sets of 3D points. Used to register a photogrammetric reconstruction to some absolute coordinate system. Often expressed as the problem of determining the rotation $\mathbf{R}$, translation $\vec{t}$ and scale s that best transforms a set of *model* points $\{\vec{m}_1, \dots, \vec{m}_n\}$ to corresponding data points $\{\vec{d}_1, \dots, \vec{d}_n\}$ by minimizing the least-squares error

$$\epsilon(R, \vec{t}, s) = \sum_{i=1}^{n} \|\vec{d}_i - s(\mathbf{R}\vec{m}_i + \vec{t})\|^2$$

to which a solution may be found by using singular value decomposition. [JKS95:1.4.2]

absolute point: A 3D point defining the origin of a coordinate system. [WP:Cartesian_coordinate_system]

absolute quadric: The symmetric 4 × 4 rank 3 matrix $\Omega = \begin{pmatrix} \mathbf{I}_3 & \vec{0}_3 \\ \vec{0}_3^\top & 0 \end{pmatrix}$. Like the absolute conic, it is defined to be invariant under Euclidean transformations, is rescaled under similarities, takes the form $\Omega = \begin{pmatrix} \mathbf{A}^\top\mathbf{A} & \vec{0}_3 \\ \vec{0}_3^\top & 0 \end{pmatrix}$ under affine transforms and becomes an arbitrary 4 × 4 rank 3 matrix under projective transforms. [FP03:13.6]

absorption: Attenuation of light caused by passing through an optical system or being incident on an object surface. [Hec87:3.5]

accumulation method: A method of accumulating evidence in histogram form, then searching for peaks, which correspond to hypotheses. See also Hough transform and generalized Hough transform. [Low91:9.3]

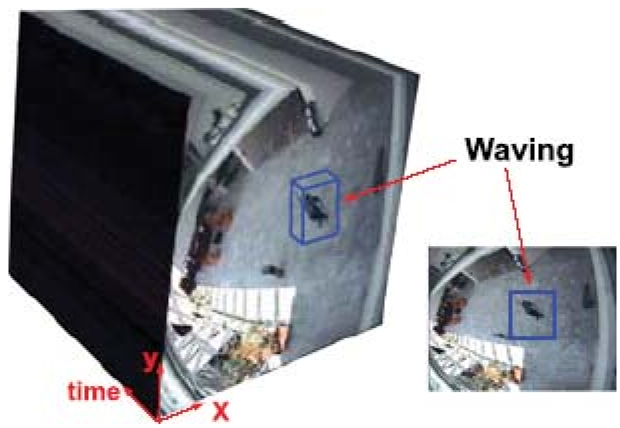
accumulative difference: A means of detecting motion in image sequences. Each frame in the sequence is compared to a reference frame (after registration if necessary) to produce a difference image. Thresholding the difference image gives a binary motion mask. A counter for each pixel location in the accumulative image is incremented every time the difference between the reference image and the current image exceeds some threshold. Used for change detection. [JKS95:14.1.1]

accuracy: The error of a value away from the true value. Contrast this with precision. [WP:Accuracy_and_precision]

acoustic sonar: Sound Navigation And Ranging. A device that is used primarily for the detection and location of objects (e.g., underwater or in air, as in mobile robotics, or internal to a human body, as in medical ultrasound) by reflecting and intercepting acoustic waves. It operates with acoustic waves in a way analogous to that of radar, using both the time of flight and Doppler effects, giving the radial component of relative position and velocity. [WP:Sonar]

ACRONYM: A vision system developed by Brooks that attempted to recognize three-dimensional objects from two-dimensional images, using generalized cylinder primitives to represent both stored model and objects extracted from the image. [Nev82:10.2]

action cuboid: The 3D spatio-temporal space in which an action detection may be localized in a video sequence:



Analogous to a window (or region of interest) in which an object detection may be localized within a 2D image. [GX11:6.4]

action detection: An approach to the automated detection of a given human, vehicle or animal activity (action) from imagery. Most commonly carried out as a video analysis task due to the temporal nature of actions. [Sze10:12.6.4]

action localization: An approach to in-image or in-scene positional localization of a given human, vehicle or animal activity. See also action detection. [Sze10:12.6.4]

action model: A pre-defined or learned model of a given human action which is matched against a given unseen action instance to perform action recognition or action detection. Akin to the use of models in model-based object recognition. [NWF08]

action recognition: Similar to action detection but further considering the classification of actions (e.g., walking, running, kicking, lifting, stretching). See also activity recognition and behavior classification, of which action recognition is often a sub-task, i.e., an activity or behavior is considered as a series of actions:

Commonly the terms action, activity and behavior are used inter-changeably in the literature. [Sze10:12.6.4]

action representation: A model-based approach whereby an action is represented as a spatio-temporal feature vector over a given video sequence. [GX11:Ch. 6]

action unit: The smallest atom or measurement of action within an action sequence or action representation removed from the raw measurement of pixel movement itself (e.g., optical flow). [LJ11:18.2.2]

active appearance model: A generalization of the widely used active shape model approach that includes all of the information in the image region covered by the target object, rather than just that near modeled edges. The active appearance model has a statistical model of the shape and gray-level appearance of the object of interest. This statistical model generalizes to cover most valid examples. Matching to an image involves finding model parameters that minimize the difference between the image and a synthesized model example, projected into the image. [NA05:6.5]

active blob: A region-based approach to the tracking of non-rigid motion in which an active shape model is used. The model is based on an initial region that is divided using Delaunay triangulation and then each patch is tracked from frame to frame (note that the patches can deform). [SI98]

active calibration: An approach to camera calibration that uses naturally occurring features within the scene with active motion of the camera to perform calibration. By contrast, traditional approaches assume a static camera and a predefined calibration object with fixed features. [Bas95]

active contour model: A technique used in model-based vision where object boundaries are detected using a deformable curve representation such as a snake. The term “active” refers to the ability of the snake to deform shape to better match the image data. See also active shape model. [SQ04:8.5]

active contour tracking: A technique used in model-based vision for tracking object boundaries in a video sequence using active contour models. [LL93]

active illumination: A system of lighting where intensity, orientation or pattern may be continuously controlled and altered. This kind of system may be used to generate structured light. [CS09:1.2]

active learning: A machine-learning approach in which the learning agent can actively query the environment for data examples. For example, a classification approach may recognize that it is less reliable over a certain sub-region of the input example space and thus request more training examples that characterize inputs for that sub-region. Considered to be a supervised learning approach. [Bar12:13.1.5]

active net: An active shape model that parameterizes a triangulated mesh. [TY89]

active recognition: An approach to object recognition or scene classification in which the recognition agent or algorithm collects further evidence samples (e.g., more images after moving) until a sufficient level of confidence is obtained to make a decision on identification. See also active learning. [RSB04]

active sensing: 1) A sensing activity carried out in an active or purposive way, e.g., where a camera is moved in space to acquire multiple or optimal views of an object (see also active vision, purposive vision and sensor planning).

2) A sensing activity implying the projection of a pattern of energy, e.g., a laser line, onto the scene (see also laser stripe triangulation and structured light triangulation). [FP03:21.1]

active shape model: Statistical model of the shape of an object that can deform to fit a new example of the object. The shapes are constrained by a statistical shape model so that they may vary only in ways seen in a training set. The models are usually formed using principal component analysis to identify the dominant modes of shape variation in observed examples of the shape. Model shapes are formed by linear combinations of the dominant modes. [WP:Active_shape_model]

active stereo: An alternative approach to traditional binocular stereo. One of the cameras is replaced with a structured light projector, which projects light onto the object of interest. If the camera calibration is known, the triangulation for computing the 3D coordinates of object points simply involves finding the intersection of a ray and known structures in the light field. [CS09:1.2]

active structure from X: The recovery of scene depth (i.e., 3D structure) via an active sensing technique, such as shape from X techniques plus motion.