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

Dictionary of Computer Vision and Image Processing

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

Set in 9/10pt Garamond by Aptara Inc., New Delhi, India

1 2014

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

Contents

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 ($\mathbf{M}$).

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

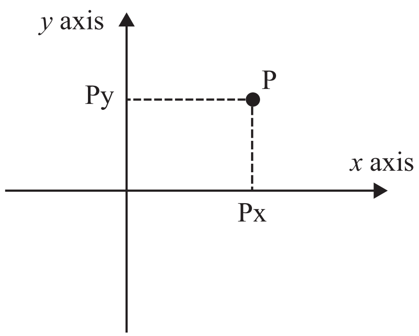
1D: One dimensional, usually in reference to some structure. Examples include: a signal $x(t)$ that is a function of time t; the dimensionality of a single property value; and one degree of freedom in shape variation or motion. [Hec87:2.1]

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 $\vec{x}$, $\|\vec{x}\|_1 = \sum_{i=0}^{n-1} \mid \vec{x}_i \mid$, 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]

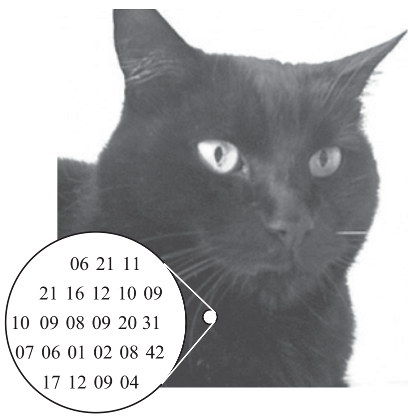


Image values

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

Dictionary of Computer Vision and Image Processing, Second Edition.
R. B. Fisher, T. P. Breckon, K. Dawson-Howe, A. Fitzgibbon, C. Robertson, E. Trucco and C. K. I. Williams.

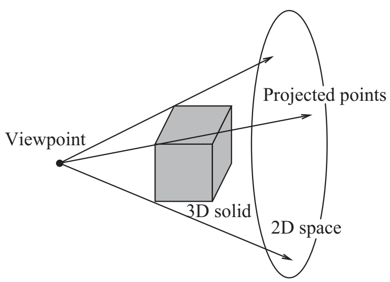
are flatbed scanners and X-ray scanners. [SQ04:4.2.1]

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.



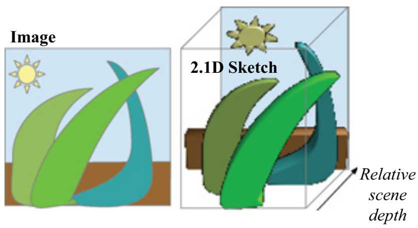
For example, the main steps for a computer graphics projection are as follows: apply normalizing transform to 3D point world coordinates; clip against canonical view volume; project onto projection plane; transform into viewport in 2D device coordinates for display. Commonly used projection functions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

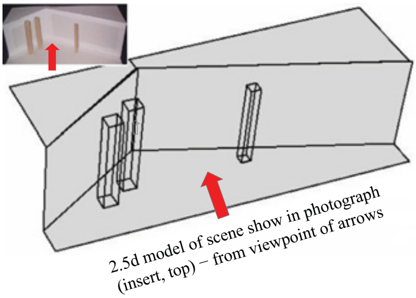
2D view: Planar aspect view or planar projected view (such as an image under perspective projection) such that positions within its spatial representation can be indexed in two dimensions. [SB11:2.3.1]

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]

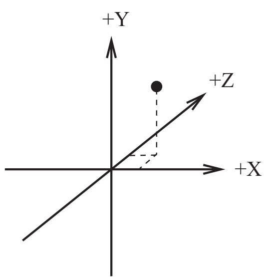
2.5D model: A geometric model representation corresponding to the 2.5D image representation used in the model to (image) data matching problem of model-based recognition: [Mar82] An example model is:



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]

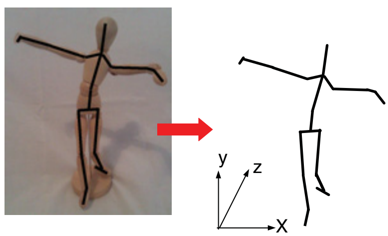
3D reconstruction: The recovery of 3D scene information and organization into a 3D shape via e.g., multi-view geometry: [HZ00:Ch. 10]

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]

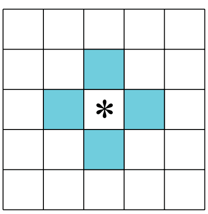
3D surface imaging: Obtaining surface information embedded in a 3D space. See also 3D imaging and 3D volumetric imaging.

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]

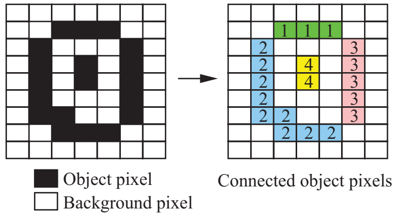
3D vision: A branch of computer vision dealing with characterizing data composed of 3D measurements. This may involve segmentation of the data into individual surfaces that are then used to identify the data as one of several models. Reverse engineering is a specialism in 3D vision. [Dav90:16.2]

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 (*):

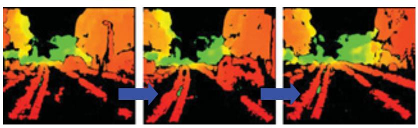


Four groups of pixels joined by 4 connectedness:

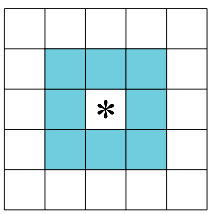


4D approach: An approach or solution to a given problem that utilizes both 3D-spatial and temporal information. See 4D representation (3D-spatial + time).

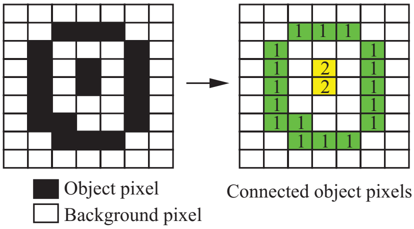
4D representation (3D-spatial + time): A 3D times series data representation whereby 3D scene information is available over a temporal sequence. An example would be a video sequence obtained from stereo vision or some other form of depth sensing: [RG08:Ch. 2]



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]

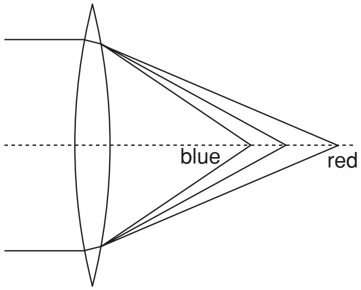
A

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 $\mathbf{KK}^\top$, 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 $\{\vec{m}_1, \ldots, \vec{m}_n\}$ to corresponding data points $\{\vec{d}_1, \ldots, \vec{d}_n\}$ by minimizing the least-squares error

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

Dictionary of Computer Vision and Image Processing, Second Edition.
R. B. Fisher, T. P. Breckon, K. Dawson-Howe, A. Fitzgibbon, C. Robertson, E. Trucco and C. K. I. Williams.

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 $\mathbf{\Omega} = \begin{pmatrix} \mathbf{I}_3 & \vec{0}_3 \\ \vec{0}_3^\top & 0 \end{pmatrix}$. Like the absolute conic, it is defined to be invariant under Euclidean transformations, is rescaled under similarities, takes the form $\mathbf{\Omega} = \begin{pmatrix} \mathbf{A}^\top\mathbf{A} & \vec{0}_3 \\ \vec{0}_3^\top & 0 \end{pmatrix}$ 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]

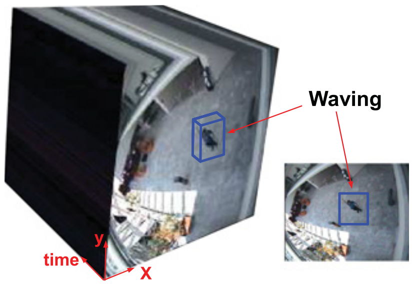
accumulative difference: A means of detecting motion in image sequences. Each frame in the sequence is compared to a reference frame (after registration if necessary) to produce a difference image. Thresholding the difference image gives a binary motion mask. A counter for each pixel location in the 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]

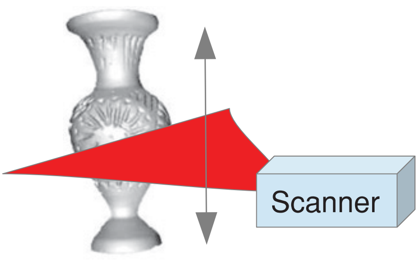
active recognition: An approach to object recognition or scene classification in which the recognition agent or algorithm collects further evidence samples (e.g., more images after moving) until a sufficient level of confidence is obtained to make a decision on identification. See also active learning.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



[Sze10:12.2] The figure shows the shape from structured light method, with the light plane being swept along the object:

active surface: 1) A surface determined using a range sensor; 2) an active shape model that deforms to fit a surface. [WP:Active_surface]

active triangulation: Determination of surface depth by triangulation between a light source at a known position and a camera that observes the effects of the illuminant on the scene. Light stripe ranging is one form of active triangulation. A variant is to use a single scanning laser beam to illuminate the scene and a stereo pair of cameras to compute depth. [WP:3D_scanner#Triangulation]

active vision: An approach to computer vision in which the camera or sensor is moved in a controlled manner, so as to simplify the nature of a problem. For example, rotating a camera with constant angular velocity while maintaining fixation at a point allows absolute calculation of scene point depth, instead of relative depth that depends on the camera speed. See also kinetic depth. [Nal93:10]

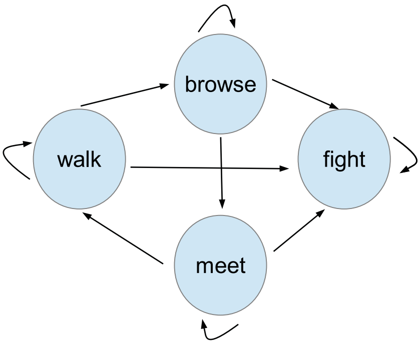
active volume: The volume of interest in a machine vision application. [SZH+10:Ch. 1]

activity: A temporal sequence of actions performed by an entity (e.g., a person, animal or vehicle) indicative of a given task, behavior or intended goal. See activity classification. [Sze10:12.6.4]

activity analysis: Analyzing the behavior of people or objects in a video sequence, for the purpose of identifying the immediate actions occurring or the long-term sequence of actions, e.g., detecting potential intruders in a restricted area. [WP:Occupational_therapy#Activity_analysis]

activity classification: The classification of a given temporal sequence of actions forming a given activity into a discrete set of labels. [Sze10:12.6.4]

activity graph: A graph encoding the activity transition matrix where each node in the graph corresponds to an activity (or stage of an activity) and the arcs among nodes represent the allowable next activities (or stages): [GX11:7.3.2]



activity model: A representation of a given activity used for activity classification via an approach akin to that of model-based object recognition.

activity recognition: See activity classification.

activity representation: See activity model.

activity segmentation: The task of segmenting a video sequence into a series of sub-sequences based on variations in activity performed along that sequence. [GX11:7.2]

activity transition matrix: An NxN matrix, for a set of N different activities, where each entry corresponds to the transition probability between two states and each state is itself an activity being performed within the scene. See also state transition probability: [GX11:7.2]

Activity	Walk	Browse	Meet	Fight
Walk	0.9	0.84	0.63	0.4
Browse	0.3	0.78	0.73	0.2
Meet	0.74	0.79	0.68	0.28
Fight	0.32	0.45	0.23	0.60

acuity: The ability of a vision system to discriminate (or resolve) between closely arranged visual stimuli. This can be measure using a grating, i.e., a pattern of parallel black and white stripes of equal widths. Once the bars become too close, the grating becomes indistinguishable from a uniform image of the same average intensity as the bars. Under optimal lighting, the minimum spacing that a person can resolve is 0.5 min of arc. [Umb98:7.6]

AdaBoost: An Adaptive Boosting approach for ensemble learning whereby the (weak) classifiers are trained in sequence such that the nth classifier is trained over a training set re-weighted to give greater emphasis to training examples upon which the previous $(n-1)$ classifiers performed poorly. See boosting. [Bis06:14.3]

adaptation: See adaptive.

adaptive: The property of an algorithm to adjust its parameters to the data at hand in order to optimize performance. Examples include adaptive contrast enhancement, adaptive filtering and adaptive smoothing. [WP: Adaptive_algorithm]

adaptive behavior model: A behavior model exhibiting adaptive properties that facilitate the online updating of the model used for behavior analysis. Generally follows a three-stage process: model initialization, online anomalous behavior detection and online model updating via unsupervised learning. See also unsupervised behavior modeling. [GX11:8.3]

adaptive bilateral filter: A variant on bilateral filtering used as an image-sharpening operator with simultaneous noise removal. Performs image sharpening by increasing the overall "slope" (i.e., the gradient range) of the edges without producing overshoot or undershoot associated with the unsharp operator. [ZA08]

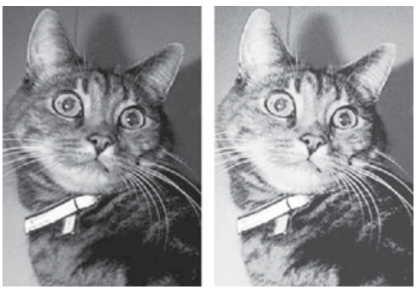
adaptive coding: A scheme for the transmission of signals over unreliable channels, e.g., wireless links. Adaptive coding varies the parameters of the encoding to respond to changes in the channel, e.g., "fading", where the signal-to-noise ratio degrades. [WP: Adaptive_coding]

adaptive contrast enhancement: An image processing operation that applies histogram equalization locally across an image. [WP: Adaptive_histogram_equalization]

adaptive edge detection: Edge detection with adaptive thresholding of the gradient magnitude image. [Nal93:3.1.2]

adaptive filtering: In signal processing, any filtering process in which the parameters of the filter change over time or where the parameters are different at different parts of the signal or image. [WP:Adaptive_filter]

adaptive histogram equalization: A localized method of improving image contrast. A histogram is constructed of the gray levels present. These gray levels are re-mapped so that the histogram is approximately flat. It can be made perfectly flat by dithering: [WP: Adaptive_histogram_equalization]



original   after adaptive histogram equalization

adaptive Hough transform: A Hough transform method that iteratively increases the resolution of the parameter space quantization. It is particularly useful for dealing with high-dimensional parameter spaces. Its disadvantage is that sharp peaks in the histogram can be missed. [NA05:5.6]

adaptive meshing: Methods for creating simplified meshes where elements are made smaller in regions of high detail (rapid changes in surface orientation) and larger in regions of low detail, such as planes. [WP:Adaptive_mesh_refinement]

adaptive pyramid: A method of multi-scale processing where small areas of image having some feature in common (e.g., color) are first extracted into a graph representation. This graph is then manipulated, e.g., by pruning or merging, until the level of desired scale is reached. [JM92]

adaptive reconstruction: Data-driven methods for creating statistically significant data in areas of a 3D data cloud where data may be missing because of sampling problems. [YGK95]

adaptive smoothing: An iterative smoothing algorithm that avoids smoothing over edges. Given an image $I(x, y)$, one iteration of adaptive smoothing proceeds as follows:

1. Compute gradient magnitude image $G(x, y) = |\nabla I(x, y)|$.
2. Make weights image $W(x, y) = e^{-\lambda G(x,y)}$.
3. Smooth the image: $S(x, y) = \frac{\sum_{i=-1}^{1} \sum_{j=-1}^{1} A_{xyij}}{\sum_{i=-1}^{1} \sum_{j=-1}^{1} B_{xyij}}$ where $A_{xyij} = I(x+i, y+j)W(x+i, y+j)$ $B_{xyij} = W(x+i, y+j)$

[WP:Additive_smoothing]

adaptive thresholding: An improved image thresholding technique where the threshold value varies at each pixel. A common technique is to use the average intensity in a neighbourhood to set the threshold: [Dav90:4.4]

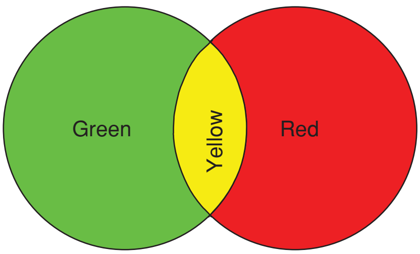
Image, I   Smoothed, S   Thresholded I > S–6

adaptive triangulation: See adaptive meshing.

adaptive visual servoing: See visual servoing. [WP:Visual_Servoing]

adaptive weighting: A scheme for weighting elements in a summation, voting or other formulation such that the relative influence of each element is representative (i.e., adapted to some underlying structure). For example this may be the similarity of pixels within a neighborhood (e.g., an adaptive bilateral filter) or a property changing over time. See also adaptive. [YK06]

additive color: The way in which multiple wavelengths of light can be combined to allow other colors to be perceived (e.g., if equal amounts of green and red light are shone onto a sheet of white paper, the paper will appear to be illuminated with a yellow light source (see below). Contrast this with subtractive color: [Gal90:3.7]



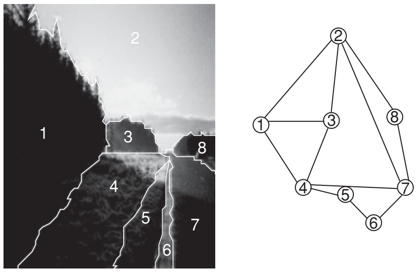
additive noise: Generally, image-independent noise that is added to it by some external process. The recorded image I at pixel (i, j) is then the sum of the true signal S and the noise N.

$$I_{i,j} = S_{i,j} + N_{i,j}$$

The noise added at each pixel (i, j) could be different. [Umb98:3.2]

adjacency: See adjacent.

adjacency graph: A graph that shows the adjacency between structures, such as segmented image regions. The nodes of the graph are the structures and an arc implies adjacency of the two structures connected by the arc. The figure shows the graph associated with the segmented image on the left: [AFF85]



adjacent: Commonly meaning "next to each other", whether in a physical sense of pixel connectivity in an image, image regions sharing some common boundary, nodes in a graph connected by an arc or components in a geometric model sharing some common bounding component. Formally defining "adjacent" can be somewhat heuristic because you may need a way to specify closeness (e.g., on a quantized grid of pixels) or to consider how much shared "boundary" is required before two structures are adjacent. [Nev82:2.1.1]

affective body gesture: See affective gesture.

affective gesture: A gesture made by the body (human or animal) which is indicative of emotional feeling or response. Used in gesture analysis to indicate social interaction. [GX11:5.4]

affective state: The emotional state on an entity (human or animal) relating to emotional feeling or response. Often measured via gesture analysis or facial expression analysis. See affective gesture.

affine: A term first used by Euler. Affine geometry is a study of properties of geometric objects that remain invariant under affine transformations (mappings), including parallelness, cross ratio and adjacency. [WP:Affine_geometry]

affine arc length: For a parametric equation of a curve $\vec{f}(u) = (x(u), y(u))$, arc length is not preserved under an affine transformation. The affine length

$$\tau(u) = \int_0^u (\dot{x}\ddot{y} - \ddot{x}\dot{y})^{\frac{1}{3}}$$

is invariant under affine transformations. [SQ04:8.4]

affine camera: A special case of the projective camera that is obtained by constraining the 3×4 camera parameter matrix $\mathbf{T}$ such that $T_{3,1} = T_{3,2} = T_{3,3} = 0$ and reducing the camera parameter vector from 11 degrees of freedom to 8. [FP03:2.3.1]

affine curvature: A measure of curvature based on the affine arc length, τ. For a parametric equation of a curve $\vec{f}(u) = (x(u), y(u))$, its affine curvature, μ, is

$$\mu(\tau) = x''(\tau)y'''(\tau) - x'''(\tau)y''(\tau)$$

[WP:Affine_curvature]

affine flow: A method of finding the movement of a surface patch by estimating the affine transformation parameters required to transform the patch from its position in one view to another. [Cal05]

affine fundamental matrix: The fundamental matrix which is obtained from a pair of cameras under affine viewing conditions. It is a 3×3 matrix whose upper left 2×2 submatrix is all zero. [HZ00:13.2.1]

affine invariant: An object or shape property that is not changed by (i.e., is invariant under) the application of an affine transformation. [FP03:18.4.1]

affine length: See affine arc length. [WP: Affine_curvature]

affine moment: Four shape measures derived from second and third order moments that remain invariant under affine transformations. They are given by the following equations, where each μ is the associated central moment: [NA05:7.3]

$$I_1 = \frac{\mu_{20}\mu_{02} - \mu_{11}^2}{\mu_{00}^4}$$

$$\begin{aligned}
I_2 = {} & (\mu_{30}^2\mu_{03}^2 - 6\mu_{30}\mu_{21}\mu_{12}\mu_{03} \\
& + 4\mu_{30}\mu_{12}^3 + 4\mu_{21}^3\mu_{03} \\
& - 3\mu_{21}^2\mu_{12}^2)/\mu_{00}^{10} \\
I_3 = {} & (\mu_{20}(\mu_{21}\mu_{03} - \mu_{12}^2) \\
& - \mu_{11}(\mu_{30}\mu_{03} - \mu_{21}\mu_{12}) \\
& + \mu_{02}(\mu_{30}\mu_{12} - \mu_{21}^2))/\mu_{00}^7 \\
I_4 = {} & (\mu_{20}^3\mu_{03}^2 - 6\mu_{20}^2\mu_{11}\mu_{12}\mu_{03} \\
& - 6\mu_{20}^2\mu_{02}\mu_{21}\mu_{03} + 9\mu_{20}^2\mu_{02}\mu_{12}^2 \\
& + 12\mu_{20}\mu_{11}^2\mu_{21}\mu_{03} \\
& + 6\mu_{20}\mu_{11}\mu_{02}\mu_{30}\mu_{03} \\
& - 18\mu_{20}\mu_{11}\mu_{02}\mu_{21}\mu_{12} \\
& - 8\mu_{11}^3\mu_{30}\mu_{03} - 6\mu_{20}\mu_{02}^2\mu_{30}\mu_{12} \\
& + 9\mu_{20}\mu_{02}^2\mu_{21}^2 + 12\mu_{11}^2\mu_{02}\mu_{30}\mu_{12} \\
& - 6\mu_{11}\mu_{02}^2\mu_{30}\mu_{21} + \mu_{02}^3\mu_{30}^2)/\mu_{00}^{11}
\end{aligned}$$

affine quadrifocal tensor: The form taken by the quadrifocal tensor when specialized to the viewing conditions modeled by the affine camera. [HTM99]

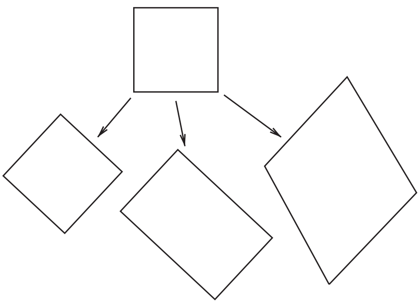
affine reconstruction: A three-dimensional reconstruction where the ambiguity in the choice of basis is affine only. Planes that are parallel in the Euclidean basis are parallel in the affine reconstruction. A projective reconstruction can be upgraded to an affine reconstruction by identification of the plane at infinity, often by locating the absolute conic in the reconstruction. [HZ00:9.4.1]

affine registration: The registration of two or more images, surface meshes or point clouds using an affine transformation. [JV05]

affine stereo: A method of scene reconstruction using two calibrated views of a scene from known viewpoints. It is a simple but very robust approximation to the geometry of stereo vision, to estimate positions, shapes and surface orientations. It can be calibrated very easily by observing just four reference points. Any two views of the same planar surface will be related by an affine transformation that maps one image to the other. This consists of a translation and a tensor, known as the disparity gradient tensor, representing the distortion in image shape. If the standard unit vectors X and Y in one image are the projections of some vectors on the object surface and the linear mapping between images is represented by a 2×3 matrix **A**, then the first two columns of **A** will be the corresponding vectors in the other image. Since the centroid of the plane will map to both image centroids, it can be used to find the surface orientation. [Qua93]

affine transformation: A special set of transformations in Euclidean geometry that preserve some properties of the construct being transformed:

- Points remain collinear: if three points belong to the same straight line, their images under affine transformations also belong to the same line and the middle point remains between the other two points.
- Parallel lines remain parallel and concurrent lines remain concurrent (images of intersecting lines intersect).
- The ratio of lengths of the segments of a given line remains constant.
- The ratio of areas of two triangles remains constant.
- Ellipses remain ellipses; parabolas remain parabolas and hyperbolas remain hyperbolas.
- Barycenters of triangles (and other shapes) map into the corresponding barycenters.



Analytically, affine transformations are represented in the matrix form

$$f(x) = \mathbf{A}x + b$$

where the determinant det(**A**) of the square matrix **A** is not 0. In 2D, the matrix is 2×2; in 3D it is 3×3. [FP03:2.2]

affine trifocal tensor: The form taken by the trifocal tensor when specialized to the viewing conditions modeled by the affine camera. [HTM99]

affinely invariant region: Image patches that automatically deform with changing viewpoint in such a way that they cover identical physical parts of a scene. Since such regions are describable by a set of invariant features they are relatively easy to match between views under changing illumination. [TG00]

affinity matrix: A matrix capturing the similarity of two entities or their relative attraction in a force- or flow-based model. Often referred to in graph cut formulations. See affinity metric. [Sze10:5.4]

affinity metric: A measurement of the similarity between two entities (e.g., features, nodes or images). See similarity metric.

affordance and action recognition: An affordance is an opportunity for an entity to take an action. The recognition of such occurrences thus identifies such action opportunities. See action recognition. [Gib86]

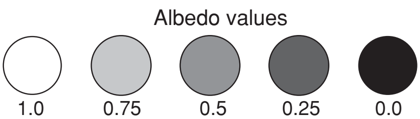
age progression: Refers to work considering the change in visual appearance because of the human aging process. Generally considered in tasks such as face recognition, face detection and face modeling. Recent work considers artificial aging of a sample facial image to produce an aged interpretation of the same: [GZSM07]

agglomerative clustering: A class of iterative clustering algorithms that begin with a large number of clusters and, at each iteration, merge pairs (or tuples) of clusters. Stopping the process at a certain number of iterations gives the final set of clusters. The process can be run until only one cluster remains and the progress of the algorithm can be represented as a dendrogram. [WP: Hierarchical_clustering]

AIC: See Akaike Information Criterion (AIC).

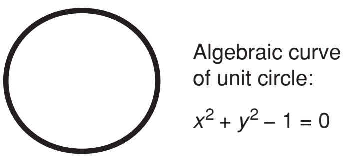
Akaike Information Criterion (AIC): A method for statistical model selection where the best-fit log-likelihood is penalized by the number of adjustable parameters in the model, so as to counter over-fitting. Compare with Bayesian information criterion. [Bis06:1.3]

albedo: Whiteness. Originally a term used in astronomy to describe reflecting power.



If a body reflects 50% of the light falling on it, it is said to have albedo of 0.5. [FP03:4.3.3]

algebraic curve: A simple parameterized curve representation using Euclidean geometry for an object that cannot be represented by linear properties (see figure). Parameterized in $\mathbb{R}^n$ Euclidean space, in the form of $\{\vec{x} : f(\vec{x}) = 0\}$: [Gib98:Ch. 1]

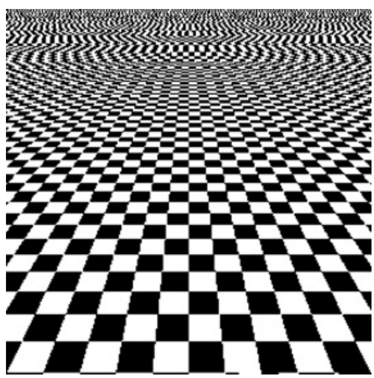


algebraic distance: A linear distance metric commonly used in computer vision applications because of its simple form and standard matrix-based least mean square estimation operations. If a curve or surface is defined implicitly by $f(\vec{x}, \vec{a}) = 0$ (e.g., $\vec{x} \cdot \vec{a} = 0$ for a hyperplane), the algebraic distance of a point $\vec{x}_i$ to the surface is simply $f(\vec{x}_i, \vec{a})$. [FP03:10.1.5]

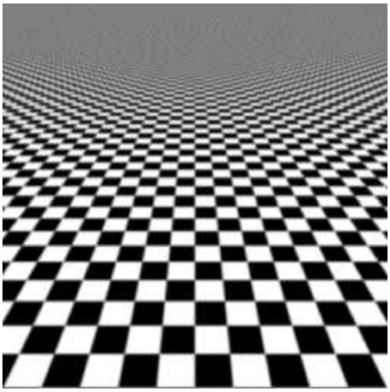
algebraic point set surfaces: A smooth surface model defined from a point cloud representation using localized moving least squares fitting of an algebraic surface (namely an algebraic sphere). [GG07]

algebraic surface: A parameterized surface representation using Euclidean geometry defined in $\mathbb{R}^n$ Euclidean space. Regular 3D surfaces such as planes, spheres, tori and generalized cylinders occur in $\mathbb{R}^3$, in the form of $\{\vec{x} : f(\vec{x}) = 0\}$. [Zar71]

aliasing: The erroneous replacement of high spatial frequency (HF) components by low-frequency ones when a signal is sampled. The affected HF components are those that are higher than the Nyquist frequency, or half the sampling frequency. Examples include the slowing of periodic signals by strobe lighting and corruption of areas of detail in image resizing. If the source signal has no HF components, the effects of aliasing are avoided, so the low-pass filtering of a signal to remove HF components prior to sampling is one form of anti-aliasing. Consider the perspective projection of a checkerboard. The image is obtained by sampling the scene at a set of integer locations. The spatial frequency increases as the plane recedes, producing aliasing artifacts (jagged lines in the foreground, moiré patterns in the background):



Removing high-frequency components (i.e., smoothing) before downsampling mitigates the effect: [FP03:7.4]



alignment: An approach to geometric model matching by registration of a geometric model to the image data. [FP03:18.2]

ALVINN: Autonomous Land Vehicle In a Neural Network; an early attempt, at Carnegie Mellon University, to learn a complex behavior (maneuvering a vehicle) by observing humans. [Pom89]

ambient light: Illumination by diffuse reflections from all surfaces within a scene (including the sky, which acts as an external distant surface). In other words, light that comes from all directions, such as the sky on a cloudy day. Ambient light ensures that all surfaces are illuminated, including those not directly facing light sources. [FP03:5.3.3]

ambient space: Refers to the dimensional space surrounding a given mathematical object in general terms. For example, a line can be studied in isolation or within a 2D space – in which case, the ambient space is a plane. Similarly a sphere may be studied in 3D ambient space. This is of particular relevance if the ambient space is nonlinear or skewed (e.g., a magnetic field). [SMC05]

AMBLER: An autonomous active vision system using both structured light and sonar, developed by NASA and Carnegie Mellon University. It is supported by a 12-legged robot and is intended for planetary exploration. [BHK+89]

amplifier noise: Spurious additive noise signal generated by the electronics in a sampling device. The standard model for this type of noise is Gaussian. It is independent of the signal. In color cameras, where more amplification is used in the blue color channel than in the green or red channels, there tends to be more noise in the blue channel. In well-designed electronics, amplifier noise is generally negligible. [WP: Image_noise#Amplifier_noise_.28 Gaussian_noise.29]

analog/mixed analog–digital image processing: The processing of images as analog signals (e.g., by optical image processing) prior to or without any form of image digitization. Largely superseded by digital image processing. [RK82]

analytic curve finding: A method of detecting parametric curves by transforming data into a feature space that is then searched for the hypothesized curve parameters. An example is line finding using the Hough transform. [XOK90]

anamorphic lens: A lens having one or more cylindrical surfaces. Anamorphic lenses are used in photography to produce images that are compressed in one dimension. Images can later be restored to true form using a reversing anamorphic lens set. This form of lens is used in wide-screen movie photography. [WP:Anamorphic_lens]

anatomical map: A biological model usable for alignment with, or region labeling of, a corresponding image dataset. For example, one could use a model of the brain's functional regions to assist in the identification of brain structures in an NMR dataset. [GHC+00]

AND operator: A Boolean logic operator that combines two input binary images:

p	q	$p\&q$
0	0	0
0	1	0
1	0	0
1	1	1

This approach is used to select image regions by applying the AND logic at each pair of corresponding pixels. The rightmost image below is the result of ANDing the two leftmost images: [SB11:3.2.2]

angiography: A method for imaging blood vessels by introducing a dye that is opaque when photographed by X-ray. Also the study of images obtained in this way. [WP:Angiography]

angularity ratio: Given two figures, X and Y, $\alpha_i(X)$ and $\beta_j(Y)$ are angles subtending convex parts of the contour of the figure X and $\gamma_k(X)$ are angles subtending plane parts of the contour of figure X; the angularity ratios are:

$$\sum_i \frac{\alpha_i(X)}{360^\circ}$$

and

$$\frac{\sum_i \beta_j(X)}{\sum_k \gamma_k(X)}$$

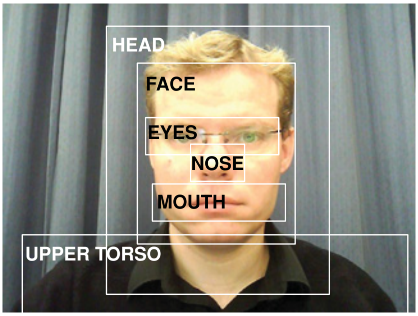
[Lee64]

anisotropic diffusion: An edge-preserving smoothing filter commonly used for noise removal. Also called Perona-Malik diffusion. See also bilateral filtering. [Sze10:3.3]

anisotropic filtering: Any filtering technique where the filter parameters vary over the image or signal being filtered. [WP:Anisotropic_filtering]

anisotropic structure tensor: A matrix representing non-uniform, second-order (i.e., gradient/edge) information of an image or function neighborhood. Commonly used in corner detection and anisotropic filtering approaches. [Sze10:3.3]

annotation: A general term referring to the labeling of imagery either with regard to manual ground truth labeling or automatic image labeling of the output of a scene understanding, semantic scene segmentation or augmented reality approach: [Sze10:14.6, 13.1.2]



anomalous behavior detection: Special case of surveillance where human movement is analyzed. Used in particular to detect intruders or behavior likely to precede or indicate crime. [WP:Anomaly_detection]

anomaly detection: The automated detection of an unexpected event, behavior or object within a given environment based on comparison with a model of what is normally expected within the same. Often considered as an unsupervised learning task and commonly applied in visual industrial inspection and automated visual surveillance. [Bar12:13.1.3]

antimode: The minimum between two maxima. One method of threshold selection is done by determining the antimode in a bimodal histogram. [WP: Bimodal_distribution#Terminology]