

CVonline: Image Databases



This is a collated list of image and video databases that people have found useful for computer vision research and algorithm evaluation.

An important article [How Good Is My Test Data? Introducing Safety Analysis for Computer Vision](#) (by Zendel, Murschitz, Humenberger, and Herzner) introduces a methodology for ensuring that your dataset has sufficient variety that algorithm results on the dataset are representative of the results that one could expect in a real setting. In particular, the team have produced a [Checklist](#) of potential hazards (imaging situations) that may cause algorithms to have problems. Ideally, test datasets should have examples of the relevant hazards.

Index by Topic

1. [Action Databases](#)
2. [Agriculture](#)
3. [Attribute recognition](#)
4. [Autonomous Driving](#)
5. [Biological/Medical](#)
6. [Camera calibration](#)
7. [Face and Eye/Iris Databases](#)
8. [Fingerprints](#)
9. [General Images](#)
10. [General RGBD and depth datasets](#)
11. [General Videos](#)
12. [Hand, Hand Grasp, Hand Action and Gesture Databases](#)
13. [Image, Video and Shape Database Retrieval](#)
14. [Object Databases](#)
15. [People \(static and dynamic\), human body pose](#)
16. [People Detection and Tracking Databases](#) (See also [Surveillance](#))
17. [Remote Sensing](#)
18. [Robotics](#)
19. [Scenes or Places, Scene Segmentation or Classification](#)
20. [Segmentation](#)
21. [Simultaneous Localization and Mapping](#)
22. [Surveillance and Tracking](#) (See also [People](#))
23. [Textures](#)
24. [Urban Datasets](#)
25. [Vision and Natural Language](#)
26. [Other Collection Pages](#)

27. [Miscellaneous Topics](#)

Two other helpful sites are:

1. [YACVID](#) - a tagged index to some computer vision datasets
 2. [Academic Torrents - computer vision](#) - a set of 30+ large datasets available in BitTorrent form
-

Action Databases

See also: [Action Recognition's](#) dataset summary with league tables (Gall, Kuehne, Bhattarai).

1. [20bn-Something-Something](#) - densely-labeled video clips that show humans performing predefined basic actions with everyday objects (Twenty Billion Neurons GmbH)
2. [3D online action dataset](#) - There are seven action categories (Microsoft and Nanyang Technological University)
3. [50 Salads](#) - fully annotated 4.5 hour dataset of RGB-D video + accelerometer data, capturing 25 people preparing two mixed salads each (Dundee University, Sebastian Stein)
4. [A first-person vision dataset of office activities \(FPVO\)](#) - FPVO contains first-person video segments of office activities collected using 12 participants. (G. Abebe, A. Catala, A. Cavallaro)
5. [ActivityNet](#) - A Large-Scale Video Benchmark for Human Activity Understanding (200 classes, 100 videos per class, 648 video hours) (Heilbron, Escorcia, Ghanem and Niebles)
6. [Action Detection in Videos](#) - MERL Shopping Dataset consists of 106 videos, each of which is a sequence about 2 minutes long (Michael Jones, Tim Marks)
7. [Actor and Action Dataset](#) - 3782 videos, seven classes of actors performing eight different actions (Xu, Hsieh, Xiong, Corso)
8. [An analyzed collation of various labeled video datasets for action recognition](#) (Kevin Murphy)
9. [ASLAN Action similarity labeling challenge](#) database (Orit Kliper-Gross)
10. [Attribute Learning for Understanding Unstructured Social Activity](#) - Database of videos containing 10 categories of unstructured social events to recognise, also annotated with 69 attributes. (Y. Fu Fudan/QMUL, T. Hospedales Edinburgh/QMUL)
11. [Audio-Visual Event \(AVE\) dataset](#)- AVE dataset contains 4143 YouTube videos covering 28 event categories and videos in AVE dataset are temporally labeled with audio-visual event boundaries. (Yapeng Tian, Jing Shi, Bochen Li, Zhiyao Duan, and Chenliang Xu)

12. [AVA: A Video Dataset of Atomic Visual Action](#) - 80 atomic visual actions in 430 15-minute movie clips. (Google Machine Perception Research Group)
13. [BBDB](#) - Baseball Database (BBDB) is a large-scale baseball video dataset that contains 4200 hours of full baseball game videos with 400,000 temporally annotated activity segments. (Shim, Minh, Young Hwi, Kyungmin, Kim, Seon Joo)
14. [BEHAVE Interacting Person Video Data with markup](#) (Scott Blunsden, Bob Fisher, Aroosha Laghaee)
15. [BU-action Datasets](#) - Three image action datasets (BU101, BU101-unfiltered, BU203-unfiltered) that have 1:1 correspondence with classes of the video datasets UCF101 and ActivityNet. (S. Ma, S. A. Bargal, J. Zhang, L. Sigal, S. Sclaroff.)
16. [Berkeley MHAD: A Comprehensive Multimodal Human Action Database](#) (Ferda Ofli)
17. [Berkeley Multimodal Human Action Database](#) - five different modalities to expand the fields of application (University of California at Berkeley and Johns Hopkins University)
18. [Breakfast dataset](#) - It's a dataset with 1712 video clips showing 10 kitchen activities, which are hand segmented into 48 atomic action classes . (H. Kuehne, A. B. Arslan and T. Serre)
19. [Bristol Egocentric Object Interactions Dataset](#) - Contains videos shot from a first-person (egocentric) point of view of 3-5 users performing tasks in six different locations (Dima Damen, Teesid Leelaswassuk and Walterio Mayol-Cuevas, Bristol University)
20. [Brown Breakfast Actions Dataset](#) - 70 hours, 4 million frames of 10 different breakfast preparation activities (Kuehne, Arslan and Serre)
21. [CAD-120 dataset](#) - focuses on high level activities and object interactions (Cornell University)
22. [CAD-60 dataset](#) - The CAD-60 and CAD-120 data sets comprise of RGB-D video sequences of humans performing activities (Cornell University)
23. [CVBASE06: annotated sports videos](#) (Janez Pers)
24. [Charades Dataset](#) - 10,000 videos from 267 volunteers, each annotated with multiple activities, captions, objects, and temporal localizations. (Sigurdsson, Varol, Wang, Laptev, Farhadi, Gupta)
25. [Composable activities dataset](#) - Different combinations of 26 atomic actions formed 16 activity classes which were performed by 14 subjects and annotations were provided (Pontificia Universidad Catolica de Chile and Universidad del Norte)
26. [Continuous Multimodal Multi-view Dataset of Human Fall](#) - The dataset consists of both normal daily activities and simulated falls for evaluating human fall detection. (Thanh-Hai Tran)
27. [Cornell Activity Datasets CAD 60, CAD 120](#) (Cornell Robot Learning Lab)
28. [DMLSmartActions dataset](#) - Sixteen subjects performed 12 different actions in a natural manner. (University of British Columbia)

29. [DemCare dataset](#) - DemCare dataset consists of a set of diverse data collection from different sensors and is useful for human activity recognition from wearable/depth and static IP camera, speech recognition for Alzheimer's disease detection and physiological data for gait analysis and abnormality detection. (K. Avgerinakis, A.Karakostas, S.Vrochidis, I. Kompatsiaris)
30. [Depth-included Human Action video dataset](#) - It contains 23 different actions (CITI in Academia Sinica)
31. [DogCentric Activity Dataset](#) - first-person videos taken from a camera mounted on top of a *dog* (Michael Ryoo)
32. [Edinburgh ceilidh overhead video data](#) - 16 ground-truthed dances viewed from overhead, where the 10 dancers follow a structured dance pattern (2 different dances). The dataset is useful for highly structured behavior understanding (Aizeboje, Fisher)
33. [EPIC-KITCHENS](#) - egocentric video recorded by 32 participants in their native kitchen environments, non-scripted daily activities, 11.5M frames, 39.6K frame-level action segments and 454.2K object bounding boxes (Damen, Doughty, Fidler, et al)
34. [EPFL crepe cooking videos](#) - 6 types of structured cooking activity (12 videos in 1920x1080 resolution (Lee, Ognibene, Chang, Kim and Demiris)
35. [ETS Hockey Game Event Data Set](#) - This data set contains footage of two hockey games captured using fixed cameras. (M.-A. Carbonneau, A. J. Raymond, E. Granger, and G. Gagnon)
36. [FCVID: Fudan-Columbia Video Dataset](#) - 91,223 Web videos annotated manually according to 239 categories (Jiang, Wu, Wang, Xue, Chang)
37. [SoccerNet](#) - Scalable dataset for action spotting in soccer videos: 500 soccer games fully annotated with main actions (goal, cards, subs) and more than 13K soccer games annotated with 500K commentaries for event captioning and game summarization. (Silvio Giancola, Mohieddine Amine, Tarek Dghaily, Bernard Ghanem)
38. [G3D](#) - synchronised video, depth and skeleton data for 20 gaming actions captured with Microsoft Kinect (Victoria Bloom)
39. [G3Di](#) - This dataset contains 12 subjects split into 6 pairs (Kingston University)
40. [Gaming 3D dataset](#) - real-time action recognition in gaming scenario (Kingston University)
41. [Georgia Tech Egocentric Activities - Gaze\(+\)](#) - videos of where people look at and their gaze location (Fathi, Li, Rehg)
42. [HMDB: A Large Human Motion Database](#) (Serre Lab)
43. [Hollywood 3D dataset](#) - 650 3D video clips, across 14 action classes (Hadfield and Bowden)
44. [Human Actions and Scenes Dataset](#) (Marcin Marszalek, Ivan Laptev, Cordelia Schmid)
45. [Human Searches](#) Search sequences of human annotators that were tasked

- to spot actions in AVA and THUMOS14 datasets. (Alwassel, H., Caba Heilbron, F., Ghanem, B.)
46. [Hollywood Extended](#) - 937 video clips with a total of 787720 frames containing sequences of 16 different actions from 69 Hollywood movies. (Bojanowski, Lajugie, Bach, Laptev, Ponce, Schmid, and Sivic)
 47. [HumanEva](#): Synchronized Video and Motion Capture Dataset for Evaluation of Articulated Human Motion (Brown University)
 48. [I-LIDS video event image dataset \(Imagery library for intelligent detection systems\)](#) (Paul Hosner)
 49. [I3DPost Multi-View Human Action Datasets](#) (Hansung Kim)
 50. [IAS-lab Action dataset](#) - contain sufficient variety of actions and number of people performing the actions (IAS Lab at the University of Padua)
 51. [ICS-FORTH MHAD101 Action Co-segmentation](#) - 101 pairs of long-term action sequences that share one or multiple common actions to be co-segmented, contains both 3d skeletal and video related frame-based features (Universtiy of Crete and FORTH-ICS, K. Papoutsakis)
 52. [IIIT Extreme Sports](#) - 160 first person (egocentric) sport videos from YouTube with frame level annotations of 18 action classes. (Suriya Singh, Chetan Arora, and C. V. Jawahar. Trajectory Aligned)
 53. [INRIA Xmas Motion Acquisition Sequences \(IXMAS\)](#) (INRIA)
 54. [InfAR Dataset](#) -Infrared Action Recognition at Different Times Neurocomputing(Chenqiang Gao, Yinhe Du, Jiang Liu, Jing Lv, Luyu Yang, Deyu Meng, Alexander G. Hauptmann)
 55. [JHMDB: Joints for the HMDB dataset](#) (J-HMDB) based on 928 clips from HMDB51 comprising 21 action categories (Jhuang, Gall, Zuffi, Schmid and Black)
 56. [JPL First-Person Interaction dataset](#) - 7 types of human activity videos taken from a first-person viewpoint (Michael S. Ryoo, JPL)
 57. [Jena Action Recognition Dataset](#) - Aibo dog actions (Korner and Denzler)
 58. [K3Da - Kinect 3D Active dataset](#) - K3Da (Kinect 3D active) is a realistic clinically relevant human action dataset containing skeleton, depth data and associated participant information (D. Leightley, M. H. Yap, J. Coulson, Y. Barnouin and J. S. McPhee)
 59. [Kinetics Human Action Video Dataset](#) - 300,000 video clips, 400 human action classe, 10 second clips, single action per clip (Kay, Carreira, et al)
 60. [KIT Robo-Kitchen Activity Data Set](#) - 540 clips of 17 people performing 12 complex kitchen activities.(L. Rybok, S. Friedberger, U. D. Hanebeck, R. Stiefelhagen)
 61. [KTH human action recognition database](#) (KTH CVAP lab)
 62. [Karlsruhe Motion, Intention, and Activity Data set \(MINTA\)](#) - 7 types of activities of daily living including fully motion primitive segments.(D. Gehrig, P. Krauthausen, L. Rybok, H. Kuehne, U. D. Hanebeck, T. Schultz, R. Stiefelhagen)
 63. [LIRIS Human Activities Dataset](#) - contains (gray/rgb/depth) videos showing

- people performing various activities (Christian Wolf, et al, French National Center for Scientific Research)
64. [MEXaction2 action detection and localization dataset](#) - To support the development and evaluation of methods for 'spotting' instances of short actions in a relatively large video database: 77 hours, 117 videos (Michel Crucianu and Jenny Benois-Pineau)
 65. [MLB-YouTube](#) - Dataset for activity recognition in baseball videos (AJ Piergiovanni, Michael Ryoo)
 66. [Moments in Time Dataset](#) - Moments in Time Dataset 1M 3-second videos annotated with action type, the largest dataset of its kind for action recognition and understanding in video. (Monfort, Oliva, et al.)
 67. [MPII Cooking Activities Dataset](#) for fine-grained cooking activity recognition, which also includes the continuous pose estimation challenge (Rohrbach, Amin, Andriluka and Schiele)
 68. [MPII Cooking 2 Dataset](#) - A large dataset of fine-grained cooking activities, an extension of the MPII Cooking Activities Dataset. (Rohrbach, Rohrbach, Regneri, Amin, Andriluka, Pinkal, Schiele)
 69. [MSR-Action3D](#) - benchmark RGB-D action dataset (Microsoft Research Redmond and University of Wollongong)
 70. [MSRAActionPair dataset](#) - : Histogram of Oriented 4D Normals for Activity Recognition from Depth Sequences (University of Central Florida and Microsoft)
 71. [MSRC-12 Kinect gesture data set](#) - 594 sequences and 719,359 frames from people performing 12 gestures (Microsoft Research Cambridge)
 72. [MSRC-12 dataset](#) - sequences of human movements, represented as body-part locations, and the associated gesture (Microsoft Research Cambridge and University of Cambridge)
 73. [MSRDailyActivity3D Dataset](#) - There are 16 activities (Microsoft and the Northwestern University)
 74. [ManiAc](#) RGB-D action dataset: different manipulation actions, 15 different versions, 30 different objects manipulated, 20 long and complex chained manipulation sequences (Eren Aksoy)
 75. [Mivia dataset](#) - It consists of 7 high-level actions performed by 14 subjects. (Mivia Lab at the University of Salerno)
 76. [MuHAVi](#) - Multicamera Human Action Video Data (Hossein Ragheb)
 77. [Multi-modal action detection \(MAD\) Dataset](#) - It contains 35 sequential actions performed by 20 subjects. (CarnegieMellon University)
 78. [Multiview 3D Event dataset](#) - This dataset includes 8 categories of events performed by 8 subjects (University of California at Los Angeles)
 79. [Nagoya University Extremely Low-resolution FIR Image Action Dataset](#) - Action recognition dataset captured by a 16x16 low-resolution FIR sensor. (Nagoya University)
 80. [NTU RGB+D Action Recognition Dataset](#) - NTU RGB+D is a large scale dataset for human action recognition(Amir Shahroudy)

81. [Northwestern-UCLA Multiview Action 3D](#) - There are 10 action categories: (Northwestern University and University of California at Los Angeles)
82. [Office Activity Dataset](#) - It consists of skeleton data acquired by Kinect 2.0 from different subjects performing common office activities. (A. Franco, A. Magnani, D. Maiop)
83. [Oxford TV based human interactions](#) (Oxford Visual Geometry Group)
84. [Parliament](#) - The Parliament dataset is a collection of 228 video sequences, depicting political speeches in the Greek parliament. (Michalis Vrigkas, Christophoros Nikou, Ioannins A. kakadiaris)
85. [Procedural Human Action Videos](#) - This dataset contains about 40,000 videos for human action recognition that had been generated using a 3D game engine. The dataset contains about 6 million frames which can be used to train and evaluate models not only action recognition but also models for depth map estimation, optical flow, instance segmentation, semantic segmentation, 3D and 2D pose estimation, and attribute learning. (Cesar Roberto de Souza)
86. [RGB-D activity dataset](#) - Each video in the dataset contains 2-7 actions involving interaction with different objects. (Cornell University and Stanford University)
87. [RGBD-Action-Completion-2016](#) - This dataset includes 414 complete/incomplete object interaction sequences, spanning six actions and presenting RGB, depth and skeleton data. (Farnoosh Heidarinvinch, Majid Mirmehdi, Dima Damen)
88. [RGB-D-based Action Recognition Datasets](#) - Paper that includes the list and links of different rgb-d action recognition datasets. (Jing Zhang, Wanqing Li, Philip O. Ogunbona, Pichao Wang, Chang Tang)
89. [RGBD-SAR Dataset](#) - RGBD-SAR Dataset (University of Electronic Science and Technology of China and Microsoft)
90. [Rochester Activities of Daily Living Dataset](#) (Ross Messing)
91. [SBU Kinect Interaction Dataset](#) - It contains eight types of interactions (Stony Brook University)
92. [SBU-Kinect-Interaction dataset v2.0](#) - It comprises of RGB-D video sequences of humans performing interaction activities (Kiwon Yun etc.)
93. [SDHA Semantic Description of Human Activities 2010 contest - Human Interactions](#) (Michael S. Ryoo, J. K. Aggarwal, Amit K. Roy-Chowdhury)
94. [SDHA Semantic Description of Human Activities 2010 contest - aerial views](#) (Michael S. Ryoo, J. K. Aggarwal, Amit K. Roy-Chowdhury)
95. [SFU Volleyball Group Activity Recognition](#) - 2 levels annotations dataset (9 players' actions and 8 scene's activity) for volleyball videos.(M. Ibrahim, S. Muralidharan, Z. Deng, A. Vahdat, and G. Mori / Simon Fraser University)
96. [SYSU 3D Human-Object Interaction Dataset](#) - Forty subjects perform 12 distinct activities (Sun Yat-sen University)
97. [ShakeFive Dataset](#) - contains only two actions, namely hand shake and high five. (Universiteit Utrecht)

98. [ShakeFive2](#) - A dyadic human interaction dataset with limb level annotations on 8 classes in 153 HD videos(Coert van Gemeren, Ronald Poppe, Remco Veltkamp)
99. [Sports Videos in the Wild \(SVW\)](#) - SVW is comprised of 4200 videos captured solely with smartphones by users of Coach Eye smartphone app, a leading app for sports training developed by TechSmith corporation.(Seyed Morteza Safdarnejad, Xiaoming Liu)
100. [Stanford Sport Events dataset](#) (Jia Li)
101. [The Leeds Activity Dataset--Breakfast \(LAD--Breakfast\)](#) - It is composed of 15 annotated videos, representing five different people having breakfast or other simple meal; (John Folkesson et al.)
102. [THU-READ\(Tsinghua University RGB-D Egocentric Action Dataset\)](#) - THU-READ is a large-scale dataset for action recognition in RGBD videos with pixel-level hand annotation. (Yansong Tang, Yi Tian, Jiwen Lu, Jianjiang Feng, Jie Zhou)
103. [THUMOS](#) - Action Recognition in Temporally Untrimmed Videos! - 430 hours of video data and 45 million frames (Gorban, Idrees, Jiang, Zamir, Laptev Shah, Sukthanka)
104. [TUM Kitchen Data Set of Everyday Manipulation Activities](#) (Moritz Tenorth, Jan Bandouch)
105. [TV Human Interaction Dataset](#) (Alonso Patron-Perez)
106. [The Falling Detection dataset](#) - Six subjects in two sceneries performed a series of actions continuously (University of Texas)
107. [The TJU dataset](#) - contains 22 actions performed by 20 subjects in two different environments; a total of 1760 sequences. (Tianjin University)
108. [UCF-iPhone Data Set](#) - 9 Aerobic actions were recorded from (6-9) subjects using the Inertial Measurement Unit (IMU) on an Apple iPhone 4 smartphone. (Corey McCall, Kishore Reddy and Mubarak Shah)
109. [The UPCV action dataset](#) - The dataset consists of 10 actions performed by 20 subjects twice. (University of Patras)
110. [UC-3D Motion Database](#) - Available data types encompass high resolution Motion Capture, acquired with MVN Suit from Xsens and Microsoft Kinect RGB and depth images. (Institute of Systems and Robotics, Coimbra, Portugal)
111. [UCF 101 action dataset](#) 101 action classes, over 13k clips and 27 hours of video data (Univ of Central Florida)
112. [UCF-Crime Dataset: Real-world Anomaly Detection in Surveillance Videos](#) - A large-scale dataset for real-world anomaly detection in surveillance videos. It consists of 1900 long and untrimmed real-world surveillance videos (of 128 hours), with 13 realistic anomalies such as fighting, road accident, burglary, robbery, etc. as well as normal activities. (Center for Research in Computer Vision, University of Central Florida)
113. [UCFKinect](#) - The dataset is composed of 16 actions (University of Central Florida Orlando)

114. [UCLA Human-Human-Object Interaction \(HHOI\) Dataset Vn1](#) - Human interactions in RGB-D videos (Shu, Ryoo, and Zhu)
115. [UCLA Human-Human-Object Interaction \(HHOI\) Dataset Vn2](#) - Human interactions in RGB-D videos (version 2) (Shu, Gao, Ryoo, and Zhu)
116. [UCR Videoweb Multi-camera Wide-Area Activities Dataset](#) (Amit K. Roy-Chowdhury)
117. [UTD-MHAD](#) - Eight subjects performed 27 actions four times. (University of Texas at Dallas)
118. [UTKinect dataset](#) - Ten types of human actions were performed twice by 10 subjects (University of Texas)
119. [UWA3D Multiview Activity Dataset](#) - Thirty activities were performed by 10 individuals (University of Western Australia)
120. [Univ of Central Florida - 50 Action Category Recognition in Realistic Videos \(3 GB\)](#) (Kishore Reddy)
121. [Univ of Central Florida - ARG](#) Aerial camera, Rooftop camera and Ground camera (UCF Computer Vision Lab)
122. [Univ of Central Florida - Feature Films Action Dataset](#) (Univ of Central Florida)
123. [Univ of Central Florida - Sports Action Dataset](#) (Univ of Central Florida)
124. [Univ of Central Florida - YouTube Action Dataset \(sports\)](#) (Univ of Central Florida)
125. [Unsegmented Sports News Videos](#) - Database of 74 sports news videos tagged with 10 categories of sports. Designed to test multi-label video tagging. (T. Hospedales, Edinburgh/QMUL)
126. [Utrecht Multi-Person Motion Benchmark \(UMPM\)](#) - a collection of video recordings of people together with a ground truth based on motion capture data.(N.P. van der Aa, X. Luo, G.J. Giezeman, R.T. Tan, R.C. Veltkamp.)
127. [VIRAT Video Dataset](#) - event recognition from two broad categories of activities (single-object and two-objects) which involve both human and vehicles. (Sangmin Oh et al)
128. [Verona Social interaction dataset](#) (Marco Cristani)
129. [ViHASi: Virtual Human Action Silhouette Data](#) (userID: VIHASI password: virtual\$virtual) (Hossein Ragheb, Kingston University)
130. [Videoweb \(multicamera\) Activities Dataset](#) (B. Bhanu, G. Denina, C. Ding, A. Ivers, A. Kamal, C. Ravishankar, A. Roy-Chowdhury, B. Varda)
131. [WVU Multi-view action recognition dataset](#) (Univ. of West Virginia)
132. [WorkoutSU-10](#) Kinect dataset for exercise actions (Ceyhun Akgul)
133. [WorkoutSU-10 dataset](#) - contains exercise actions selected by professional trainers for therapeutic purposes. (Sabanci University)
134. [Wrist-mounted camera video dataset](#) - object manipulation (Ohnishi, Kanehira, Kanezaki, Harada)
135. [YouCook](#) - 88 open-source YouTube cooking videos with annotations (Jason Corso)
136. [YouTube-8M Dataset](#) -A Large and Diverse Labeled Video Dataset for Video

Understanding Research(Google Inc.)

Agriculture

1. [Aberystwyth Leaf Evaluation Dataset](#) - Timelapse plant images with hand marked up leaf-level segmentations for some time steps, and biological data from plant sacrifice. (Bell, Jonathan; Dee, Hannah M.)
2. [Fieldsafe](#) - A multi-modal dataset for obstacle detection in agriculture. (Aarhus University)
3. [KOMATSUNA dataset](#) - The datasets is designed for instance segmentation, tracking and reconstruction for leaves using both sequential multi-view RGB images and depth images. (Hideaki Uchiyama, Kyushu University)
4. [Leaf counting dataset](#) - Dataset for estimating the growth stage of small plants. (Aarhus University)
5. [Leaf Segmentation Challenge](#) Tobacco and arabidopsis plant images (Hanno Scharr, Massimo Minervini, Andreas Fischbach, Sotirios A. Tsaftaris)
6. [Multi-species fruit flower detection](#) - This dataset consists of four sets of flower images, from three different tree species: apple, peach, and pear, and accompanying ground truth images. (Philippe A. Dias, Amy Tabb, Henry Medeiros)
7. [Plant Phenotyping Datasets](#) - plant data suitable for plant and leaf detection, segmentation, tracking, and species recognition (M. Minervini, A. Fischbach, H. Scharr, S. A. Tsaftaris)
8. [Plant seedlings dataset](#) - High-resolution images of 12 weed species. (Aarhus University)

Attribute recognition

1. [Attribute Learning for Understanding Unstructured Social Activity](#) - Database of videos containing 10 categories of unstructured social events to recognise, also annotated with 69 attributes. (Y. Fu Fudan/QMUL, T. Hospedales Edinburgh/QMUL)
2. [Animals with Attributes 2](#) - 37322 (freely licensed) images of 50 animal classes with 85 per-class binary attributes. (Christoph H. Lampert, IST Austria)
3. [Birds](#) This database contains 600 images (100 samples each) of six different classes of birds. (Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce)
4. [Butterflies](#) This database contains 619 images of seven different classes of butterflies. (Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce)
5. [CALVIN research group datasets](#) - object detection with eye tracking, imagenet bounding boxes, synchronised activities, stickman and body poses, youtube objects, faces, horses, toys, visual attributes, shape classes (CALVIN ggroup)

6. [CelebA](#) - Large-scale CelebFaces Attributes Dataset(Ziwei Liu, Ping Luo, Xiaogang Wang, Xiaoou Tang)
7. [DukeMTMC-attribute](#) - 23 pedestrian attributes for DukeMTMC-reID (Lin, Zheng, Zheng, Wu and Yang)
8. [EMOTIC \(EMOTions in Context\)](#) - Images of people (34357) embedded in their natural environments, annotated with 2 distinct emotion representation. (Ronak kosti, Agata Lapedriza, Jose Alvarez, Adria Recasens)
9. [HAT](#) database of 27 human attributes (Gaurav Sharma, Frederic Jurie)
10. [LFW-10 dataset for learning relative attributes](#) - A dataset of 10,000 pairs of face images with instance-level annotations for 10 attributes.(CVIT, IIIT Hyderabad.)
11. [Market-1501-attribute](#) - 27 visual attributes for 1501 shoppers. (Lin, Zheng, Zheng, Wu and Yang)
12. [Multi-Class Weather Dataset](#) - Our multi-class benchmark dataset contains 65,000 images from 6 common categories for sunny, cloudy, rainy, snowy, haze and thunder weather. This dataset benefits weather classification and attribute recognition. (Di Lin)
13. [Person Recognition in Personal Photo Collections](#) - we introduced three harder splits for evaluation and long-term attribute annotations and per-photo timestamp metadata.(Oh, Seong Joon and Benenson, Rodrigo and Fritz, Mario and Schiele, Bernt)
14. [UT-Zappos50K Shoes](#) - Large scale shoe dataset consisting of 50,000 catalog images and over 50,000 pairwise relative attribute labels on 11 fine-grained attributes (Aron Yu, Mark Stephenson, Kristen Grauman, UT Austin)
15. [Visual Attributes Dataset](#) visual attribute annotations for over 500 object classes (animate and inanimate) which are all represented in ImageNet. Each object class is annotated with visual attributes based on a taxonomy of 636 attributes (e.g., has fur, made of metal, is round).
16. [The Visual Privacy \(VISPR\) Dataset](#) - Privacy Multilabel Dataset (22k images, 68 privacy attributes) (Orekondy, Schiele, Fritz)
17. [WIDER Attribute Dataset](#) - WIDER Attribute is a large-scale human attribute dataset, with 13789 images belonging to 30 scene categories, and 57524 human bounding boxes each annotated with 14 binary attributes.(Li, Yining and Huang, Chen and Loy, Chen Change and Tang, Xiaoou)

Autonomous Driving

1. [AMUSE](#) -The automotive multi-sensor (AMUSE) dataset taken in real traffic scenes during multiple test drives. (Philipp Koschorrek etc.)
2. [Autonomous Driving](#) - Semantic segmentation, pedestrian detection, virtual-world data, far infrared, stereo,driver monitoring. (CVC research

- center and the UAB and UPC universities)
3. [Ford Campus Vision and Lidar Data Set](#) - time-registered data from professional (Applanix POS LV) and consumer (Xsens MTI-G) Inertial Measuring Unit (IMU), Velodyne 3D-lidar scanner, two push-broom forward looking Riegl lidars, and a Point Grey Ladybug3 omnidirectional camera system (Pandey, McBride, Eustice)
 4. [FRIDA \(Foggy Road Image DAtabase\) Image Database](#) - images for performance evaluation of visibility and contrast restoration algorithms. FRIDA: 90 synthetic images of 18 urban road scenes. FRIDA2: 330 synthetic images of 66 diverse road scenes, with viewpoint closed to that of the vehicle's driver. (Tarel, Cord, Halmaoui, Gruyer, Hautiere)
 5. [House3D](#) - House3D is a virtual 3D environment which consists of thousands of indoor scenes equipped with a diverse set of scene types, layouts and objects sourced from the SUNCG dataset. It consists of over 45k indoor 3D scenes, ranging from studios to two-storied houses with swimming pools and fitness rooms. All 3D objects are fully annotated with category labels. Agents in the environment have access to observations of multiple modalities, including RGB images, depth, segmentation masks and top-down 2D map views. The renderer runs at thousands frames per second, making it suitable for large-scale RL training. (Yi Wu, Yuxin Wu, Georgia Gkioxari, Yuandong Tian, facebook research)
 6. [India Driving Dataset \(IDD\)](#) - unstructured driving conditions from India with 50,000 frames (10,000 semantic, and 40,000 coarse annotations) for training autonomous cars to see using object detection, scene-level and instance-level semantic segmentation (CVIT, IIIT Hyderabad and Intel)
 7. [Joint Attention in Autonomous Driving \(JAAD\)](#) - The dataset includes instances of pedestrians and cars intended primarily for the purpose of behavioural studies and detection in the context of autonomous driving. (Iuliia Kotseruba, Amir Rasouli and John K. Tsotsos)
 8. [LISA Vehicle Detection Dataset](#) - colour first person driving video under various lighting and traffic conditions (Sivaraman, Trivedi)
 9. [Lost and Found Dataset](#) - The Lost and Found Dataset addresses the problem of detecting unexpected small road hazards (often caused by lost cargo) for autonomous driving applications. (Sebastian Ramos, Peter Pinggera, Stefan Gehrig, Uwe Franke, Rudolf Mester, Carsten Rother)
 10. [nuTonomy scenes dataset \(nuScenes\)](#) - The nuScenes dataset is a large-scale autonomous driving dataset. It features: Full sensor suite (1x LIDAR, 5x RADAR, 6x camera, IMU, GPS), 1000 scenes of 20s each, 1,440,000 camera images, 400,000 lidar sweeps, two diverse cities: Boston and Singapore, left versus right hand traffic, detailed map information, manual annotations for 25 object classes, 1.1M 3D bounding boxes annotated at 2Hz, attributes such as visibility, activity and pose. (Caesar et al)
 11. [RESIDE \(Realistic Single Image DEhazing\)](#) - The current largest-scale benchmark consisting of both synthetic and real-world hazy images, for

image dehazing research. RESIDE highlights diverse data sources and image contents, and serves various training or evaluation purposes. (Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, Zhangyang Wang)

12. [SYNTHIA](#) - Large set (~half million) of virtual-world images for training autonomous cars to see. (ADAS Group at Computer Vision Center)
13. [The Multi Vehicle Stereo Event Camera Dataset](#) - Multiple sequences containing a stereo pair of DAVIS 346b event cameras with ground truth poses, depth maps and optical flow. (lex Zihao Zhu, Dinesh Thakur, Tolga Ozaslan, Bernd Pfrommer, Vijay Kumar, Kostas Daniilidis)
14. [The SYNTHetic collection of Imagery and Annotations](#) - The purpose of aiding semantic segmentation and related scene understanding problems in the context of driving scenarios. (Computer vision center,UAB)
15. [TRoM: Tsinghua Road Markings](#) - This is a dataset which contributes to the area of road marking segmentation for Automated Driving and ADAS. (Xiaolong Liu, Zhidong Deng, Lele Cao, Hongchao Lu)
16. [University of Michigan North Campus Long-Term Vision and LIDAR Dataset](#) - 27 sessions spaced approximately biweekly over the course of 15 months, indoors and outdoors, varying trajectories, different times of the day across all four seasons. Includes: moving obstacles (e.g., pedestrians, bicyclists, and cars), changing lighting, varying viewpoint, seasonal and weather changes (e.g., falling leaves and snow), and long-term structural changes caused by construction. Includes ground-truth pose. (Carlevaris-Bianco, Ushani, Eustice)

Biological/Medical

1. [2008 MICCAI MS Lesion Segmentation Challenge](#) (National Institutes of Health Blueprint for Neuroscience Research)
2. [ASU DR-AutoCC Data](#) - a Multiple-Instance Learning feature space for a diabetic retinopathy classification dataset (Ragav Venkatesan, Parag Chandakkar, Baoxin Li - Arizona State University)
3. [Aberystwyth Leaf Evaluation Dataset](#) - Timelapse plant images with hand marked up leaf-level segmentations for some time steps, and biological data from plant sacrifice. (Bell, Jonathan; Dee, Hannah M.)
4. [Annotated Spine CT Database](#) for Benchmarking of Vertebrae Localization, 125 patients, 242 scans (Ben Glocker)
5. [BRATS](#) - the identification and segmentation of tumor structures in multiparametric magnetic resonance images of the brain (TU Munchen etc.)
6. [Breast Ultrasound Dataset B](#) - 2D Breast Ultrasound Images with 53 malignant lesions and 110 benign lesions. (UDIAT Diagnostic Centre, M.H. Yap, R. Marti)

7. [Calgary-Campinas Public Brain MR Dataset](#): T1-weighted brain MRI volumes acquired in 359 subjects on scanners from three different vendors (GE, Philips, and Siemens) and at two magnetic field strengths (1.5 T and 3 T). The scans correspond to older adult subjects. (Souza, Roberto, Oeslle Lucena, Julia Garrafa, David Gobbi, Marina Saluzzi, Simone Appenzeller, Leticia Rittner, Richard Frayne, and Roberto Lotufo)
8. [Cholec80](#): 80 gallbladder laparoscopic videos annotated with phase and tool information. (Andru Putra Twinanda)
9. [CRCHistoPhenotypes - Labeled Cell Nuclei Data](#) - colorectal cancer?histology images?consisting of nearly 30,000 dotted nuclei with over 22,000 labeled with the cell type (Rajpoot + Sirinukunwattana)
10. [Cavy Action Dataset](#) - 16 sequences with 640 x 480 resolutions recorded at 7.5 frames per second (fps) with approximately 31621506 frames in total (272 GB) of interacting cavies (guinea pig) (Al-Raziqi and Denzler)
11. [Cell Tracking Challenge Datasets](#) - 2D/3D time-lapse video sequences with ground truth(Ma et al., Bioinformatics 30:1609-1617, 2014)
12. [Computed Tomography Emphysema Database](#) (Lauge Sorensen)
13. [COPD Machine Learning Dataset](#) - A collection of feature datasets derived from lung computed tomography (CT) images, which can be used in diagnosis of chronic obstructive pulmonary disease (COPD). The images in this database are weakly labeled, i.e. per image, a diagnosis(COPD or no COPD) is given, but it is not known which parts of the lungs are affected. Furthermore, the images were acquired at different sites and with different scanners. These problems are related to two learning scenarios in machine learning, namely multiple instance learning or weakly supervised learning, and transfer learning or domain adaptation. (Veronika Cheplygina, Isabel Pino Pena, Jesper Holst Pedersen, David A. Lynch, Lauge S., Marleen de Bruijne)
14. [CREMI: MICCAI 2016 Challenge](#) - 6 volumes of electron microscopy of neural tissue,neuron and synapse segmentation, synaptic partner annotation. (Jan Funke, Stephan Saalfeld, Srini Turaga, Davi Bock, Eric Perlman)
15. [CRIM13 Caltech Resident-Intruder Mouse dataset](#) - 237 10 minute videos (25 fps) annotated with actions (13 classes) (Burgos-Artizzu, Doll??r, Lin, Anderson and Perona)
16. [DIADEM: Digital Reconstruction of Axonal and Dendritic Morphology Competition](#) (Allen Institute for Brain Science et al)
17. [DIARETDB1 - Standard Diabetic Retinopathy Database](#) (Lappeenranta Univ of Technology)
18. [DRIVE: Digital Retinal Images for Vessel Extraction](#) (Univ of Utrecht)
19. [DeformIt 2.0](#) - Image Data Augmentation Tool: Simulate novel images with ground truth segmentations from a single image-segmentation pair (Brian Booth and Ghassan Hamarneh)
20. [Deformable Image Registration Lab dataset](#) - for objective and rigorous

- evaluation of deformable image registration (DIR) spatial accuracy performance. (Richard Castillo et al.)
21. [DERMOFIT Skin Cancer Dataset](#) - 1300 lesions from 10 classes captured under identical controlled conditions. Lesion segmentation masks are included (Fisher, Rees, Aldridge, Ballerini, et al)
 22. [Dermoscopy images](#) (Eric Ehram)
 23. [EPT29](#). This database contains 4842 images of 1613 specimens of 29 taxa of EPTs: (Tom etc.)
 24. [EATMINT \(Emotional Awareness Tools for Mediated INteraction\) database](#) - The EATMINT database contains multi-modal and multi-user recordings of affect and social behaviors in a collaborative setting. (Guillaume Chanel, Gaelle Molinari, Thierry Pun, Mireille Betrancourt)
 25. [FIRE Fundus Image Registration Dataset](#) - 134 retinal image pairs and ground truth for registration. (FORTH-ICS)
 26. [Histology Image Collection Library \(HICL\)](#) - The HICL is a compilation of 3870 histopathological images (so far) from various diseases, such as brain cancer, breast cancer and HPV (Human Papilloma Virus)-Cervical cancer. (Medical Image and Signal Processing (MEDISP) Lab., Department of Biomedical Engineering, School of Engineering, University of West Attica)
 27. [Honeybee segmentation dataset](#) - It is a dataset containing positions and orientation angles of hundreds of bees on a 2D surface of honey comb. (Bozek K, Hebert L, Mikheyev AS, Stephen GJ)
 28. [IIT MBADA mice](#) - Mice behavioral data. FLIR A315, spacial resolution of 320x240px at 30fps, 50x50cm open arena, two experts for three different mice pairs, mice identities. (Italian Inst. of Technology, PAVIS lab)
 29. [Indian Diabetic Retinopathy Image Dataset](#) - This dataset consists of retinal fundus images annotated at pixel-level for lesions associated with Diabetic Retinopathy. Also, it provides the disease severity of diabetic retinopathy and diabetic macular edema. This dataset is useful for development and evaluation of image analysis algorithms for early detection of diabetic retinopathy. (Prasanna Porwal, Samiksha Pachade, Ravi Kamble, Manesh Kokare, Girish Deshmukh, Vivek Sahasrabudhe, Fabrice Meriaudeau)
 30. [IRMA \(Image retrieval in medical applications\)](#) - This collection compiles anonymous radiographs (Deserno TM, Ott B)
 31. [KID](#) - A capsule endoscopy database for medical decision support (Anastasios Koulaouzidis and Dimitris Iakovidis)
 32. [Leaf Segmentation Challenge](#) Tobacco and arabidopsis plant images (Hanno Scharr, Massimo Minervini, Andreas Fischbach, Sotirios A. Tsaftaris)
 33. [LITS Liver Tumor Segmentation](#) - 130 3D CT scans with segmentations of the liver and liver tumor. Public benchmark with leaderboard at Codalab.org (Patrick Christ)
 34. [Medical image database](#) - Database of ultrasound images of breast abnormalities with the ground truth. (Prof. Stanislav Makhanov, biomedsiit.com)

35. [MIT CBCL Automated Mouse Behavior Recognition datasets](#) (Nicholas Edelman)
36. [MUCIC: Masaryk University Cell Image Collection](#) - 2D/3D synthetic images of cells/tissues for benchmarking(Masaryk University)
37. [MiniMammographic Database](#) (Mammographic Image Analysis Society)
38. [Moth fine-grained recognition](#) - 675 similar classes, 5344 images (Erik Rodner et al)
39. [Mouse Embryo Tracking Database](#) - cell division event detection (Marcelo Cicconet, Kris Gunsalus)
40. [OASIS](#) - Open Access Series of Imaging Studies - 500+ MRI data sets of the brain (Washington University, Harvard University, Biomedical Informatics Research Network)
41. [Plant Phenotyping Datasets](#) - plant data suitable for plant and leaf detection, segmentation, tracking, and species recognition (M. Minervini, A. Fischbach, H. Scharr, S. A. Tsiftaris)
42. [RatSI: Rat Social Interaction Dataset](#) - 9 fully annotated (11 class) videos (15 minute, 25 FPS) of two rats interacting socially in a cage (Malte Lorbach, Noldus Information Technology)
43. [Retinal fundus images - Ground truth of vascular bifurcations and crossovers](#) (Univ of Groningen)
44. [SCORHE](#) - 1, 2 and 3 mouse behavior videos, 9 behaviors, (Ghadi H. Salem, et al, NIH)
45. [STructured Analysis of the Retina](#) - DESCRIPTION(400+ retinal images, with ground truth segmentations and medical annotations)
46. [Spine and Cardiac data](#) (Digital Imaging Group of London Ontario, Shuo Li)
47. [Stonefly9](#)This database contains 3826 images of 773 specimens of 9 taxa of Stoneflies (Tom etc.)
48. [Synthetic Migrating Cells](#) -Six artificial migrating cells (neutrophils) over 98 time frames, various levels of Gaussian/Poisson noise and different paths characteristics with ground truth. (Dr Constantino Carlos Reyes-Aldasoro et al.)
49. [UBFC-RPPG Dataset](#) - remote photoplethysmography (rPPG) video data and ground truth acquired with a CMS50E transmissive pulse oximeter (Bobbia, Macwan, Benezeth, Mansouri, Dubois)
50. [Uni Bremen Open, Abdominal Surgery RGB Dataset](#) - Recording of a complete, open, abdominal surgery using a Kinect v2 that was mounted directly above the patient looking down at patient and staff. (Joern Teuber, Gabriel Zachmann, University of Bremen)
51. [Univ of Central Florida - DDSM: Digital Database for Screening Mammography](#) (Univ of Central Florida)
52. [VascuSynth](#) - 120 3D vascular tree like structures with ground truth (Mengliu Zhao, Ghassan Hamarneh)
53. [VascuSynth](#) - Vascular Synthesizer generates vascular trees in 3D volumes. (Ghassan Hamarneh, Preet Jassi, Mengliu Zhao)

54. [York Cardiac MRI dataset](#) (Alexander Andreopoulos)

Camera calibration

1. [Catadioptric camera calibration images](#) (Yalin Bastanlar)
2. [GoPro-Gyro Dataset](#) - This dataset consists of a number of wide-angle rolling shutter video sequences with corresponding gyroscope measurements (Hannes etc.)
3. [LO-RANSAC](#) - LO-RANSAC library for estimation of homography and epipolar geometry (K. Lebeda, J. Matas and O. Chum)

Face and Eye/Iris Databases

1. [2D-3D face dataset](#) - This dataset includes pairs of 2D face image and its corresponding 3D face geometry model with geometry details. (Yudong Guo, Juyong Zhang, Jianfei Cai, Boyi Jiang, Jianmin Zheng)
2. [300 Videos in the Wild \(300-VW\)](#) - 68 Facial Landmark Tracking (Chrysos, Antonakos, Zafeiriou, Snape, Shen, Kossaifi, Tzimiropoulos, Pantic)
3. [3D Mask Attack Database \(3DMAD\)](#) - 76500 frames of 17 persons using Kinect RGBD with eye positions (Sebastien Marcel)
4. [3D facial expression](#) - Binghamton University 3D Static and Dynamic Facial Expression Databases (Lijun Yin, Jeff Cohn, and teammates)
5. [AginG Faces in the Wild v2](#) Database description: AGFW-v2 consists of 36,299 facial images divided into 11 age groups with a span of five years between groups. On average, there are 3,300 images per group. Facial images in AGFW-v2 are not public figures and less likely to have significant make-up or facial modifications, helping embed accurate aging effects during the learning process. (Chi Nhan Duong, Khoa Luu, Kha Gia Quach, Tien D. Bui)
6. [Audio-visual database for face and speaker recognition](#) (Mobile Biometry MOBIO <http://www.mobioproject.org/>)
7. [BANCA face and voice database](#) (Univ of Surrey)
8. [Binghamton Univ 3D static and dynamic facial expression database](#) (Lijun Yin, Peter Gerhardstein and teammates)
9. [Binghamton-Pittsburgh 4D Spontaneous Facial Expression Database](#) - consist of 2D spontaneous facial expression videos and FACS codes. (Lijun Yin et al.)
10. [BioID face database](#) (BioID group)
11. [BioVid Heat Pain Database](#) - This video (and biomedical signal) dataset contains facial and physiopsychological reactions of 87 study participants who were subjected to experimentally induced heat pain. (University of Magdeburg (Neuro-Information Technology group) and University of Ulm (Emotion Lab))

12. [Biometric databases](#) - biometric databases related to iris recognition (Adam Czajka)
13. [Biwi 3D Audiovisual Corpus of Affective Communication](#) - 1000 high quality, dynamic 3D scans of faces, recorded while pronouncing a set of English sentences.
14. [Bosphorus 3D/2D Database of FACS annotated facial expressions, of head poses and of face occlusions](#) (Bogazici University)
15. [Caricature/Photomates dataset](#) - a dataset with frontal faces and corresponding Caricature line drawings (Tayfun Akgul)
16. [CASIA-IrisV3](#) (Chinese Academy of Sciences, T. N. Tan, Z. Sun)
17. [CASIR Gaze Estimation Database](#) - RGB and depth images (from Kinect V1.0) and ground truth values of facial features corresponding to experiments for gaze estimation benchmarking: (Filipe Ferreira etc.)
18. [CMU Facial Expression Database](#) (CMU/MIT)
19. [The CMU Multi-PIE Face Database](#) - more than 750,000 images of 337 people recorded in up to four sessions over the span of five months. (Jeff Cohn et al.)
20. [CMU Pose, Illumination, and Expression \(PIE\) Database](#) (Simon Baker)
21. [CMU/MIT Frontal Faces](#) (CMU/MIT)
22. [CMU/MIT Frontal Faces](#) (CMU/MIT)
23. [CSSE Frontal intensity and range images of faces](#) (Ajmal Mian)
24. [CelebA](#) - Large-scale CelebFaces Attributes Dataset(Ziwei Liu, Ping Luo, Xiaogang Wang, Xiaoou Tang)
25. [Cohn-Kanade AU-Coded Expression Database](#) - 500+ expression sequences of 100+ subjects, coded by activated Action Units (Affect Analysis Group, Univ. of Pittsburgh)
26. [Cohn-Kanade AU-Coded Expression Database](#) - for research in automatic facial image analysis and synthesis and for perceptual studies (Jeff Cohn et al.)
27. [Columbia Gaze Data Set](#) - 5,880 images of 56 people over 5 head poses and 21 gaze directions (Brian A. Smith, Qi Yin, Steven K. Feiner, Shree K. Nayar)
28. [Computer Vision Laboratory Face Database \(CVL Face Database\)](#) - Database contains 798 images of 114 persons, with 7 images per person and is freely available for research purposes.(Peter Peer etc.)
29. [Deep future gaze](#) - This dataset consists of 57 sequences on search and retrieval tasks performed by 55 subjects. Each video clip lasts for around 15 minutes with the frame rate 10 fps and frame resolution 480 by 640. Each subject is asked to search for a list of 22 items (including lanyard, laptop) and move them to the packing location (dining table). (National University of Singapore, Institute for Infocomm Research)
30. [DISFA+:Extended Denver Intensity of Spontaneous Facial Action Database](#) - an extension of DISFA (M.H. Mahoor)
31. [DISFA:Denver Intensity of Spontaneous Facial Action Database](#) - a non-

- posed facial expression database for those who are interested in developing computer algorithms for automatic action unit detection and their intensities described by FACS. (M.H. Mahoor)
32. [DHF1K](#) - 1000 elaborately selected video sequences with fixation annotations from 17 viewers. (Prof. Jianbing Shen)
 33. [EURECOM Facial Cosmetics Database](#) - 389 images, 50 persons with/without make-up, annotations about the amount and location of applied makeup.(Jean-Luc DUGELAY et al)
 34. [EURECOM Kinect Face Database](#) - 52 people, 2 sessions, 9 variations, 6 facial landmarks.(Jean-Luc DUGELAY et al)
 35. [EYEDIAP dataset](#) - The EYEDIAP dataset was designed to train and evaluate gaze estimation algorithms from RGB and RGB-D data.It contains a diversity of participants, head poses, gaze targets and sensing conditions. (Kenneth Funes and Jean-Marc Odobez)
 36. [Face2BMI Dataset](#) The Face2BMI dataset contains 2103 pairs of faces, with corresponding gender, height and previous and current body weights, which allows for training computer vision models that can predict body-mass index (BMI) from profile pictures. (Enes Kocabey, Ferda Ofli, Yusuf Aytar, Javier Marin, Antonio Torralba, Ingmar Weber)
 37. [FDDB: Face Detection Data set and Benchmark - studying unconstrained face detection](#) (University of Massachusetts Computer Vision Laboratory)
 38. [FG-Net Aging Database of faces at different ages](#) (Face and Gesture Recognition Research Network)
 39. [Face Recognition Grand Challenge datasets](#) (FRVT - Face Recognition Vendor Test)
 40. [FMTV](#) - Laval Face Motion and Time-Lapse Video Database. 238 thermal/video subjects with a wide range of poses and facial expressions acquired over 4 years (Ghiass, Bendada, Maldague)
 41. [Face Super-Resolution Dataset](#) - Ground truth HR-LR face images captured with a dual-camera setup (Chengchao Qu etc.)
 42. [FaceScrub](#) - A Dataset With Over 100,000 Face Images of 530 People (50:50 male and female) (H.-W. Ng, S. Winkler)
 43. [FaceTracer Database - 15,000 faces](#) (Neeraj Kumar, P. N. Belhumeur, and S. K. Nayar)
 44. [Facial Expression Dataset](#) - This dataset consists of 242 facial videos (168,359 frames) recorded in real world conditions. (Daniel McDuff et al.)
 45. [Florence 2D/3D Hybrid Face Dataset](#) - bridges the gap between 2D, appearance-based recognition techniques, and fully 3D approaches (Bagdanov, Del Bimbo, and Masi)
 46. [Facial Recognition Technology \(FERET\) Database](#) (USA National Institute of Standards and Technology)
 47. [Gi4E Database](#) - eye-tracking database with 1300+ images acquired with a standard webcam, corresponding to different subjects gazing at different points on a screen, including ground-truth 2D iris and corner points

- (Villanueva, Ponz, Sesma-Sanchez, Mikel Porta, and Cabeza)
48. [Google Facial Expression Comparison dataset](#) - a large-scale facial expression dataset consisting of face image triplets along with human annotations that specify which two faces in each triplet form the most similar pair in terms of facial expression, which is different from datasets that focus mainly on discrete emotion classification or action unit detection (Vemulapalli, Agarwala)
 49. [Hannah and her sisters database](#) - a dense audio-visual person-oriented ground-truth annotation of faces, speech segments, shot boundaries (Patrick Perez, Technicolor)
 50. [Headspace dataset](#) - The Headspace dataset is a set of 3D images of the full human head, consisting of 1519 subjects wearing tight fitting latex caps to reduce the effect of hairstyles. (Christian Duncan, Rachel Armstrong, Alder Hey Craniofacial Unit, Liverpool, UK)
 51. [Hong Kong Face Sketch Database](#)
 52. [IDIAP Head Pose Database \(IHPD\)](#) - The dataset contains a set of meeting videos along with the head groundtruth of individual participants (around 128min)(Sileye Ba and Jean-Marc Odobez)
 53. [IMDB-WIKI](#) - 500k+ face images with age and gender labels (Rasmus Rothe, Radu Timofte, Luc Van Gool)
 54. [Indian Movie Face database \(IMFDB\)](#) - a large unconstrained face database consisting of 34512 images of 100 Indian actors collected from more than 100 videos (Vijay Kumar and C V Jawahar)
 55. [Iranian Face Database](#) - IFDB is the first image database in middle-east, contains color facial images with age, pose, and expression whose subjects are in the range of 2-85. (Mohammad Mahdi Dehshibi)
 56. [Japanese Female Facial Expression \(JAFFE\) Database](#) (Michael J. Lyons)
 57. [LFW: Labeled Faces in the Wild](#) - unconstrained face recognition
 58. [LS3D-W](#) - a large-scale 3D face alignment dataset annotated with 68 points containing faces captured in a "in-the-wild" setting. (Adrian Bulat, Georgios Tzimiropoulos)
 59. [MAFA: MAsked FAcEs](#) - 30,811 images with 35,806 labeled MAsked FAcEs, six main attributes of each masked face. (Shiming Ge, Jia Li, Qiting Ye, Zhao Luo)
 60. [Makeup Induced Face Spoofing \(MIFS\)](#) - 107 makeup-transformations attempting to spoof a target identity. Also other datasets. (Antitza Dantcheva)
 61. [Mexculture142](#) - Mexican Cultural heritage objects and eye-tracker gaze fixations (Montoya Obeso, Benois-Pineau, Garcia-Vazquez, Ramirez Acosta)
 62. [MIT CBCL Face Recognition Database](#) (Center for Biological and Computational Learning)
 63. [MIT Collation of Face Databases](#) (Ethan Meyers)
 64. [MIT eye tracking database \(1003 images\)](#) (Judd et al)
 65. [MMI Facial Expression Database](#) - 2900 videos and high-resolution still

- images of 75 subjects, annotated for FACS AUs.
66. [MORPH \(Craniofacial Longitudinal Morphological Face Database\)](#) (University of North Carolina Wilmington)
 67. [MPIIGaze dataset](#) - 213,659 samples with eye images and gaze target under different illumination conditions and nature head movement, collected from 15 participants with their laptop during daily using. (Xucong Zhang, Yusuke Sugano, Mario Fritz, Andreas Bulling.)
 68. [Manchester Annotated Talking Face Video Dataset](#) (Timothy Cootes)
 69. [MegaFace](#) - 1 million faces in bounding boxes (Kemelmacher-Shlizerman, Seitz, Nech, Miller, Brossard)
 70. [Music video dataset](#) - 8 music videos from YouTube for developing multi-face tracking algorithms in unconstrained environments (Shun Zhang, Jia-Bin Huang, Ming-Hsuan Yang)
 71. [NIST Face Recognition Grand Challenge \(FRGC\)](#) (NIST)
 72. [NIST mugshot identification database](#) (USA National Institute of Standards and Technology)
 73. [NRC-IIT Facial Video Database](#) - this database contains pairs of short video clips each showing a face of a computer user sitting in front of the monitor exhibiting a wide range of facial expressions and orientations (Dmitry Gorodnichy)
 74. [Notre Dame Iris Image Dataset](#) (Patrick J. Flynn)
 75. [Notre Dame face, IR face, 3D face, expression, crowd, and eye biometric datasets](#) (Notre Dame)
 76. [ORL face database: 40 people with 10 views](#) (ATT Cambridge Labs)
 77. [OUI-Adience Faces](#) - unfiltered faces for gender and age classification plus 3D faces (OUI)
 78. [Oxford: faces, flowers, multi-view, buildings, object categories, motion segmentation, affine covariant regions, misc](#) (Oxford Visual Geometry Group)
 79. [Pandora](#) - POSEidon: Face-from-Depth for Driver Pose (Borghini, Venturelli, Vezzani, Cucchiara)
 80. [PubFig: Public Figures Face Database](#) (Neeraj Kumar, Alexander C. Berg, Peter N. Belhumeur, and Shree K. Nayar)
 81. [QMUL-SurvFace](#) - A large-scale face recognition benchmark dedicated for real-world surveillance face analysis and matching. (QMUL Computer Vision Group)
 82. [Re-labeled Faces in the Wild](#) - original images, but aligned using "deep funneling" method. (University of Massachusetts, Amherst)
 83. [RT-GENE: Real-Time Eye Gaze Estimation in Natural Environments](#) 122,531 images with the subjects' ground truth eye gaze and head pose labels under free-viewing conditions and large camera-subject distances (Fischer, Chang, Demiris, Imperial College London)
 84. [S3DFM](#) - Edinburgh Speech-driven 3D Facial Motion Database. 77 people with 10 repetitions of speaking a passphrase: 1 second of 500 frame per

- second 600x600 pixels of {IR intensity video, registered depth images} plus synchronized 44.1 Khz audio. There are an additional 26 people (10 repetitions) moving their heads while speaking (Zhang, Fisher)
85. [Salient features in gaze-aligned recordings of human visual input](#) - TB of human gaze-contingent data "in the wild" (Frank Schumann etc.)
 86. [SAMM Dataset of Micro-Facial Movements](#) - The dataset contains 159 spontaneous micro-facial movements obtained from 32 participants from 13 different ethnicities. (A.Davison, C.Lansley, N.Costen, K.Tan, M.H.Yap)
 87. [SCface](#) - Surveillance Cameras Face Database (Mislav Grgic, Kresimir Delac, Sonja Grgic, Bozidar Klmpak)
 88. [SiblingsDB](#) - The SiblingsDB contains two datasets depicting images of individuals related by sibling relationships. (Politecnico di Torino/Computer Graphics & Vision Group)
 89. [Solving the Robot-World Hand-Eye\(s\) Calibration Problem with Iterative Methods](#) - These datasets were generated for calibrating robot-camera systems. (Amy Tabb)
 90. [Spontaneous Emotion Multimodal Database \(SEM-db\)](#) - non-posed reactions to visual stimulus data recorded with HD RGB, depth and IR frames of the face, EEG signal and eye gaze data (Fernandez. Montenegro, Gkelias, Argyriou)
 91. [The UNBC-McMaster Shoulder Pain Expression Archive Database](#) - Painful data: The UNBC-McMaster Shoulder Pain Expression Archive Database (Lucy et al.)
 92. [The York 3D Ear Dataset](#) - The York 3D Ear Dataset is a set of 500 3D ear images, synthesized from detailed 2D landmarking, and available in both Matlab format (.mat) and PLY format (.ply). (Nick Pears, Hang Dai, Will Smith, University of York)
 93. [Trondheim Kinect RGB-D Person Re-identification Dataset](#) (Igor Barros Barbosa)
 94. [UB KinFace Database](#) - University of Buffalo kinship verification and recognition database
 95. [UBIRIS: Noisy Visible Wavelength Iris Image Databases](#) (University of Beira)
 96. [UMDFaces](#) - About 3.7 million annotated video frames from 22,000 videos and 370,000 annotated still images. (Ankan Bansal et al.)
 97. [UPNA Head Pose Database](#) - head pose database, with 120 webcam videos containing guided-movement sequences and free-movement sequences, including ground-truth head pose and automatically annotated 2D facial points. (Ariz, Bengoechea, Villanueva, Cabeza)
 98. [UPNA Synthetic Head Pose Database](#) - a synthetic replica of the UPNA Head Pose Database, with 120 videos with their 2D ground truth landmarks projections, their corresponding head pose ground truth, 3D head models and camera parameters. (Larumbe, Segura, Ariz, Bengoechea, Villanueva, Cabeza)

99. [UTIRIS cross-spectral iris image databank](#) (Mahdi Hosseini)
100. [VGGFace2](#) - VGGFace2 is a large-scale face recognition dataset covering large variations in pose, age, illumination, ethnicity and profession. (Oxford Visual Geometry Group)
101. [VIPSL Database](#) - VIPSL Database is for research on face sketch-photo synthesis and recognition, including 200 subjects (1 photo and 5 sketches per subject).(Nannan Wang)
102. [Visual Search Zero Shot Database](#) - Collection of human eyetracking data in three increasingly complex visual search tasks: object arrays, natural images and Waldo images. (Kreiman lab)
103. [VT-KFER](#): A Kinect-based RGBD+Time Dataset for Spontaneous and Non-Spontaneous Facial Expression Recognition - 32 subjects, 1,956 sequences of RGBD, six facial expressions in 3 poses (Aly, Trubanova, Abbott, White, and Youssef)
104. [Washington Facial Expression Database \(FERG-DB\)](#) - a database of 6 stylized (Maya) characters with 7 annotated facial expressions (Deepali Aneja, Alex Colburn, Gary Faigin, Linda Shapiro, and Barbara Mones)
105. [WebCaricature Dataset](#) - The WebCaricature dataset is a large photograph-caricature dataset consisting of 6042 caricatures and 5974 photographs from 252 persons collected from the web. (Jing Huo, Wenbin Li, Yinghuan Shi, Yang Gao and Hujun Yin)
106. [WIDER FACE: A Face Detection Benchmark](#) - 32,203 images with 393,703 labeled faces, 61 event classes (Shuo Yang, Ping Luo, Chen Change Loy, Xiaoou Tang)
107. [XM2VTS Face video sequences \(295\): The extended M2VTS Database \(XM2VTS\)](#) - (Surrey University)
108. [Yale Face Database - 11 expressions of 10 people](#) (A. Georghaides)
109. [Yale Face Database B - 576 viewing conditions of 10 people](#) (A. Georghaides)
110. [York Univ Eye Tracking Dataset \(120 images\)](#) (Neil Bruce)
111. [YouTube Faces DB](#) - 3,425 videos of 1,595 different people. (Wolf, Hassner, Maoz)
112. [Zurich Natural Image](#) - the image material used for creating natural stimuli in a series of eye-tracking studies (Frey et al.)

Fingerprints

1. [FVC fingerpring verification competition 2002 dataset](#) (University of Bologna)
2. [FVC fingerpring verification competition 2004 dataset](#) (University of Bologna)
3. [Fingerprint Manual Minutiae Marker \(FM3\) Databases](#): - Fingerprint Manual Minutiae Marker (FM3) Databases(Mehmet Kayaoglu, Berkay

Topcu and Umut Uludag)

4. [NIST fingerprint databases](#) (USA National Institute of Standards and Technology)
5. [SPD2010 Fingerprint Singular Points Detection Competition](#) (SPD 2010 committee)

General Images

1. [A Dataset for Real Low-Light Image Noise Reduction](#) - It contains pixel and intensity aligned pairs of images corrupted by low-light camera noise and their low-noise counterparts. (J. Anaya, A. Barbu)
2. [A database of paintings related to Vincent van Gogh](#) - This is the dataset VGDB-2016 built for the paper "From Impressionism to Expressionism: Automatically Identifying Van Gogh's Paintings" (Guilherme Folego and Otavio Gomes and Anderson Rocha)
3. [AMOS: Archive of Many Outdoor Scenes \(20+m\)](#) (Nathan Jacobs)
4. [Aerial images](#) Building detection from aerial images using invariant color features and shadow information. (Beril Sirmacek)
5. [Approximated overlap error dataset](#) Image pairs with sparse sets of ground-truth matches for evaluating local image descriptors (Fabio Bellavia)
6. [AutoDA \(Automatic Dataset Augmentation\)](#) - An automatically constructed image dataset including 12.5 million images with relevant textual information for the 1000 categories of ILSVRC2012 (Bai, Yang, Ma, Zhao)
7. [BGU Hyperspectral Image Database of Natural Scenes](#) (Ohad Ben-Shahar and Boaz Arad)
8. [Brown Univ Large Binary Image Database](#) (Ben Kimia)
9. [Butterfly-200](#) - Butterfly-20 is a image dataset for fine-grained image classification, which contains 25,279 images and covers four levels categories of 200 species, 116 genera, 23 subfamilies, and 5 families. (Tianshui Chen)
10. [CMP Facade Database](#) - Includes 606 rectified images of facades from various places with 12 architectural classes annotated.(Radim Tylecek)
11. [Caltech-UCSD Birds-200-2011](#) (Catherine Wah)
12. [Color correction dataset](#) - Homography-based registered images for evaluating color correction algorithms for image stitching. (Fabio Bellavia)
13. [Columbia Multispectral Image Database](#) (F. Yasuma, T. Mitsunaga, D. Iso, and S.K. Nayar)
14. [DAQUAR \(Visual Turing Challenge\)](#) - A dataset containing questions and answers about real-world indoor scenes.(Mateusz Malinowski, Mario Fritz)
15. [Darmstadt Noise Dataset](#) - 50 pairs of real noisy images and corresponding ground truth images (RAW and sRGB) (Tobias Plotz and Stefan Roth)
16. [Dataset of American Movie Trailers 2010-2014](#) - Contains links to 474

- hollywood movie trailers along with associated metadata (genre, budget, runtime, release, MPAA rating, screens released, sequel indicator) (USC Signal Analysis and Interpretation Lab)
17. [DIML Multimodal Benchmark](#) - To evaluate matching performance under photometric and geometric variations, 100 images of 1200 x 800 size. (Yonsei University)
 18. [DSLR Photo Enhancement Dataset \(DPED\)](#) - 22K photos taken synchronously in the wild by three smartphones and one DSLR camera, useful for comparing inferred high quality images from multiple low quality images (Ignatov, Kobyshev, Timofte, Vanhoey, and Van Gool).
 19. [Flickr-style](#) - 80K Flickr photographs annotated with 20 curated style labels, and 85K paintings annotated with 25 style/genre labels (Sergey Karayev)
 20. [Forth Multispectral Imaging Datasets](#) - images from 23 spectral bands each from 5 paintings. Images are annotated with ground truth data. (Karamaounas Polykarpos et al)
 21. [General 100 Dataset](#) - General-100 dataset contains 100 bmp-format images (with no compression), which are well-suited for super-resolution training (Dong, Chao and Loy, Chen Change and Tang, Xiaoou)
 22. [GOPRO dataset](#) - Blurred image dataset with sharp image ground truth (Nah, Kim, and Lee)
 23. [HIPR2 Image Catalogue of different types of images](#) (Bob Fisher et al)
 24. [HPatches](#) - A benchmark and evaluation of handcrafted and learned local descriptors (Balntas, Lenc, Vedaldi, Mikolajczyk)
 25. [Hyperspectral images for spatial distributions of local illumination in natural scenes](#) - Thirty calibrated hyperspectral radiance images of natural scenes with probe spheres embedded for local illumination estimation. (Nascimento, Amano & Foster)
 26. [Hyperspectral images of natural scenes - 2002](#) (David H. Foster)
 27. [Hyperspectral images of natural scenes - 2004](#) (David H. Foster)
 28. [ISPRS multi-platform photogrammetry dataset](#) - 1: Nadir and oblique aerial images plus 2: Combined UAV and terrestrial images (Francesco Nex and Markus Gerke)
 29. [Image & Video Quality Assessment at LIVE](#) - used to develop picture quality algorithms (the University of Texas at Austin)
 30. [ImageNet Large Scale Visual Recognition Challenges](#) - Currently 200 object classes and 500+K images (Alex Berg, Jia Deng, Fei-Fei Li and others)
 31. [ImageNet Linguistically organised \(WordNet\) Hierarchical Image Database - 10E7 images, 15K categories](#) (Li Fei-Fei, Jia Deng, Hao Su, Kai Li)
 32. [Improved 3D Sparse Maps for High-performance Structure from Motion with Low-cost Omnidirectional Robots - Evaluation Dataset](#) - Data set used in research paper doi:10.1109/ICIP.2015.7351744 (Breckon, Toby P., Cavestany, Pedro)
 33. [LabelMeFacade Database](#) - 945 labeled building images (Erik Rodner et al)

34. [Local illumination hyperspectral radiance images](#) - Thirty hyperspectral radiance images of natural scenes with embedded probe spheres for local illumination estimates(Sgio M. C. Nascimento, Kinjiro Amano, David H. Foster)
35. [McGill Calibrated Colour Image Database](#) (Adriana Olmos and Fred Kingdom)
36. [Multiply Distorted Image Database](#) -a database for evaluating the results of image quality assessment metrics on multiply distorted images.(Fei Zhou)
37. [NPRgeneral](#) - A standardized collection of images for evaluating image stylization algorithms. (David Mould, Paul Rosin)
38. [nuTonomy scenes dataset \(nuScenes\)](#) - The nuScenes dataset is a large-scale autonomous driving dataset. It features: Full sensor suite (1x LIDAR, 5x RADAR, 6x camera, IMU, GPS), 1000 scenes of 20s each, 1,440,000 camera images, 400,000 lidar sweeps, two diverse cities: Boston and Singapore, left versus right hand traffic, detailed map information, manual annotations for 25 object classes, 1.1M 3D bounding boxes annotated at 2Hz, attributes such as visibility, activity and pose. (Caesar et al)
39. [NYU Symmetry Database](#) - 176 single-symmetry and 63 multiple-symmetry images (Marcelo Cicconet and Davi Geiger)
40. [OceanDark dataset](#) - 100 low-lighting underwater images from underwater sites in the Northeast Pacific Ocean. 1400x1000 pixels, varying lighting and recording conditions (Ocean Networks Canada)
41. [OTCBVS Thermal Imagery Benchmark Dataset Collection](#) (Ohio State Team)
42. [PANorama Sparsely STructured Areas Datasets](#) - the PASSTA datasets used for evaluation of the image alignment (Andreas Robinson)
43. [QMUL-OpenLogo](#) - A logo detection benchmark for testing the model generalisation capability in detecting a variety of logo objects in natural scenes with the majority logo classes unlabelled. (QMUL Computer Vision Group)
44. [RESIDE \(Realistic Single Image DEhazing\)](#) - The current largest-scale benchmark consisting of both synthetic and real-world hazy images, for image dehazing research. RESIDE highlights diverse data sources and image contents, and serves various training or evaluation purposes. (Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, Zhangyang Wang)
45. [Rijksmuseum Challenge 2014](#) - It consist of 100K art objects from the rijksmuseum and comes with an extensive xml files describing each object. (Thomas Mensink and Jan van Gemert)
46. [See in the Dark](#) - 77 Gb of dark images (Chen, Chen, Xu, and Koltun)
47. [Smartphone Image Denoising Dataset \(SIDD\)](#) - The Smartphone Image Denoising Dataset (SIDD) consists of about 30,000 noisy images with corresponding high-quality ground truth in both raw-RGB and sRGB spaces obtained from 10 scenes with different lighting conditions using five

- representative smartphone cameras. (Abdelrahman Abdelhamed, Stephen Lin, Michael S. Brown)
48. [Stanford Street View Image, Pose, and 3D Cities Dataset](#) - a large scale dataset of street view images (25 million images and 118 matching image pairs) with their relative camera pose, 3D models of cities, and 3D metadata of images. (Zamir, Wekel, Agrawal, Malik, Savarese)
 49. [TESTIMAGES](#) - Huge and free collection of sample images designed for analysis and quality assessment of different kinds of displays (i.e. monitors, televisions and digital cinema projectors) and image processing techniques. (Nicola Asuni)
 50. [The Konstanz visual quality databases](#) - Large-scale image and video databases for the development and evaluation of visual quality assessment algorithms. (MMSP group, University of Konstanz)
 51. [Time-Lapse Hyperspectral Radiance Images of Natural Scenes](#) - Four time-lapse sequences of 7-9 calibrated hyperspectral radiance images of natural scenes taken over the day. (Foster, D.H., Amano, K., & Nascimento, S.M.C.)
 52. [Time-lapse hyperspectral radiance images](#) - Four time-lapse sequences of 7-9 calibrated hyperspectral images of natural scenes, spectra at 10-nm intervals(David H. Foster, Kinjiro Amano, Sgio M. C. Nascimento)
 53. [Tiny Images Dataset](#) 79 million 32x32 color images (Fergus, Torralba, Freeman)
 54. [TURBID Dataset](#) - five different subsets of degraded images with its respective ground-truth. Subsets Milk and DeepBlue have 20 images each and the subset Chlorophyll has 42 images (Amanda Duarte)
 55. [UT Snap Angle 360° Dataset](#) - A list of 360° videos of four activities (disney, parade, ski, concert) from youtube (Kristen Grauman, UT Austin)
 56. [UT Snap Point Dataset](#) - Human judgement on snap point quality of a subset of frames from UT Egocentric dataset and a newly collected mobile robot dataset (frames are also included) (Bo Xiong, Kristen Grauman, UT Austin)
 57. [Visual Dialog](#) - 120k human-human dialogs on COCO images, 10 rounds of QA per dialog (Das, Kottur, Gupta, Singh, Yadav, Moura, Parikh, Batra)
 58. [Visual Question Answering](#) - 254K imgs, 764K questions, ground truth (Agrawal, Lu, Antol, Mitchell, Zitnick, Batra, Parikh)
 59. [Visual Question Generation](#) - 15k images (including both object-centric and event-centric images), 75k natural questions asked about the images which can evoke further conversation(Nasrin Mostafazadeh, Ishan Misra, Jacob Devlin, Margaret Mitchell, Xiao dong He, Lucy Vanderwende)
 60. [VOA Human Attention](#) - 60k human attention maps for visual question answering i.e. where humans choose to look to answer questions about images (Das, Agrawal, Zitnick, Parikh, Batra)
 61. [Wild Web tampered image dataset](#) - A large collection of tampered images from Web and social media sources, including ground-truth annotation masks for tampering localization (Markos Zampoglou, Symeon

Papadopoulos)

62. [YFCC100M: The New Data in Multimedia Research](#) - This publicly available curated dataset of 100 million photos and videos is free and legal for all.(Bart Thomee, Yahoo Labs and Flickr in San Francisco,etc.)

General RGBD and Depth Datasets

Note: there are 3D datasets elsewhere as well, *e.g.* in [Objects](#), [Scenes](#), and [Actions](#).

See also: [List of RGBD datasets](#).

1. [360D](#) - A dataset of paired color and depth 360 spherical panoramas from 22096 unique viewpoints to be used for evaluating omnidirectional dense depth estimation methods. (Nikolaos Zioulis, Antonis Karakottas, Dimitrios Zarpalas, Petros Daras)
2. [3D-Printed RGB-D Object Dataset](#) - 5 objects with groundtruth CAD models and camera trajectories, recorded with various quality RGB-D sensors. (Siemens & TUM)
3. [3DCOMET](#) - 3DCOMET is a dataset for testing 3D data compression methods.(Miguel Cazorla, Javier Navarrete,Vicente Morell, Miguel Cazorla, Diego Viejo, Jose Garcia-Rodriguez, Sergio Orts.)
4. [3D articulated body](#) - 3D reconstruction of an articulated body with rotation and translation. Single camera, varying focal. Every scene may have an articulated body moving. There are four kinds of data sets included. A sample reconstruction result included which uses only four images of the scene. (Prof Jihun Park)
5. [A Dataset for Non-Rigid Reconstruction from RGB-D Data](#) - Eight scenes for reconstructing non-rigid geometry from RGB-D data, each containing several hundred frames along with our results. (Matthias Innmann, Michael Zollhoefer, Matthias Niessner, Christian Theobalt, Marc Stamminger)
6. [A Large Dataset of Object Scans](#) - 392 objects in 9 cases, hundreds of frames each (Choi, Zhou, Miller, Koltun)
7. [Articulated Object Challenge](#) - 4 articulated objects consisting of rigid parts connected by 1D revolute and prismatic joints, 7000+ RGBD images with annotations for 6D pose estimation(Frank Michel, Alexander Krull, Eric Brachmann, Michael. Y. Yang, Stefan Gumhold, Carsten Rother)
8. [BigBIRD](#) - 100 objects with for each object, 600 3D point clouds and 600 high-resolution color images spanning all views (Singh, Sha, Narayan, Achim, Abbeel)
9. [CAESAR](#) Civilian American and European Surface Anthropometry Resource Project - 4000 3D human body scans (SAE International)
10. [CIN 2D+3D object classification dataset](#) - segmented color and depth images of objects from 18 categories of common household and office

- objects (Bjorn Browatzki et al)
11. [CoRBS](#) - an RGB-D SLAM benchmark, providing the combination of real depth and color data together with a ground truth trajectory of the camera and a ground truth 3D model of the scene (Oliver Wasenmuller)
 12. [CSIRO synthetic deforming people](#) - synthetic RGBD dataset for evaluating non-rigid 3D reconstruction: 2 subjects and 4 camera trajectories (Elanattil and Moghadam)
 13. [CTU Garment Folding Photo Dataset](#) - Color and depth images from various stages of garment folding.(Sushkov R., Melkumov I., Smutn y V. (Czech Technical University in Prague))
 14. [CTU Garment Sorting Dataset](#) - Dataset of garment images, detailed stereo images, depth images and weights.(Petrik V., Wagner L. (Czech Technical University in Prague))
 15. [Clothing part dataset](#) - The clothing part dataset consists of image and depth scans, acquired with a Kinect, of garments laying on a table, with over a thousand part annotations (collar, cuffs, hood, etc) using polygonal masks.(Arnau Ramisa, Guillem Aleny, Francesc Moreno-Noguer and Carme Torras)
 16. [Cornell-RGBD-Dataset](#) - Office Scenes (Hema Koppula)
 17. [CVSSP Dynamic RGBD Modelling 2015](#) - This dataset contains eight RGBD sequences of general dynamic scenes captured using the Kinect V1/V2 as well as two synthetic sequences. (Charles Malleson, CVSSP, University of Surrey)
 18. [Deformable 3D Reconstruction Dataset](#) - two single-stream RGB-D sequences of dynamically moving mechanical toys together with ground-truth 3D models in the canonical rest pose. (Siemens, TUM)
 19. [Delft Windmill Interior and Exterior Laser Scanning Point Clouds](#) (Beril Sirmacek)
 20. [Diabetes60](#) - RGB-D images of 60 western dishes, home made. Data was recorded using a Microsoft Kinect V2. (Patrick Christ and Sebastian Schlecht)
 21. [ETH3D](#) - Benchmark for multi-view stereo and 3D reconstruction, covering a variety of indoor and outdoor scenes, with ground truth acquired by a high-precision laser scanner. (Thomas Sch??ps, Johannes L. Sch??nberger, Silvano Galliani, Torsten Sattler, Konrad Schindler, Marc Pollefeys, Andreas Geiger)
 22. [EURECOM Kinect Face Database](#) - 52 people, 2 sessions, 9 variations, 6 facial landmarks. (Jean-Luc DUGELAY et al)
 23. [G4S meta rooms](#) - RGB-D data 150 sweeps with 18 images per sweep. (John Folkesson et al.)
 24. [Georgiatech-Metz Symphony Lake Dataset](#) - 5 million RGBD outdoor images over 4 years from 121 surveys of a lakeshore. (Griffith and Pradalier)
 25. [Goldfinch: GOogLe image-search Dataset for FINE grained CHallenges](#) - a

- largescale dataset for finegrained bird (11K species),butterfly (14K species), aircraft (409 types), and dog (515 breeds) recognition.(Jonathan Krause, Benjamin Sapp, Andrew Howard, Howard Zhou, Alexander Toshev, Tom Duerig, James Philbin, Li Fei-Fei)
26. [House3D](#) - House3D is a virtual 3D environment which consists of thousands of indoor scenes equipped with a diverse set of scene types, layouts and objects sourced from the SUNCG dataset. It consists of over 45k indoor 3D scenes, ranging from studios to two-storied houses with swimming pools and fitness rooms. All 3D objects are fully annotated with category labels. Agents in the environment have access to observations of multiple modalities, including RGB images, depth, segmentation masks and top-down 2D map views. The renderer runs at thousands frames per second, making it suitable for large-scale RL training. (Yi Wu, Yuxin Wu, Georgia Gkioxari, Yuandong Tian, facebook research)
 27. [IMPART multi-view/multi-modal 2D+3D film production dataset](#) - LIDAR, video, 3D models, spherical camera, RGBD, stereo, action, facial expressions, etc. (Univ. of Surrey)
 28. [Industrial 3D Object Detection Dataset \(MVTec ITODD\)](#) - depth and gray value data of 28 objects in 3500 labeled scenes for 3D object detection and pose estimation with a strong focus on industrial settings and applications (MVTec Software GmbH, Munich)
 29. [Kinect v2 Dataset](#) - Efficient Multi-Frequency Phase Unwrapping using Kernel Density Estimation (Felix etc.)
 30. [KOMATSUNA dataset](#) - The datasets is designed for instance segmentation, tracking and reconstruction for leaves using both sequential multi-view RGB images and depth images. (Hideaki Uchiyama, Kyushu University)
 31. [Make3D Laser+Image data](#) - about 1000 RGB outdoor images with aligned laser depth images (Saxena, Chung, Ng, Sun)
 32. [McGill-Reparti Artificial Perception Database](#) - RGBD data from four cameras and unfiltered Vicon skeletal data of two human subjects performing simulated assembly tasks on a car door (Andrew Phan, Olivier St-Martin Cormier, Denis Ouellet, Frank P. Ferrie).
 33. [Meta rooms](#) - RGB-D data comprised of 28 aligned depth camera images collected by having robot go to specific place and do 360 degrees of pan with various tilts. (John Folkesson et al.)
 34. [METU Multi-Modal Stereo Datasets ???Benchmark Datasets for Multi-Modal Stereo-Vision???](#) - The METU Multi-Modal Stereo Datasets includes benchmark datasets for for Multi-Modal Stereo-Vision which is composed of two datasets: (1) The synthetically altered stereo image pairs from the Middlebury Stereo Evaluation Dataset and (2) the visible-infrared image pairs captured from a Kinect device. (Dr. Mustafa Yaman, Dr. Sinan Kalkan)
 35. [MHT RGB-D](#) - collected by a robot every 5 min over 16 days by the University of Lincoln. (John Folkesson et al.)
 36. [Moving Infants In RGB-D \(MINI-RGBD\)](#) - A synthetic, realistic RGB-D data

- set for infant pose estimation containing 12 sequences of moving infants with ground truth joint positions. (N. Hesse, C. Bodensteiner, M. Arens, U. G. Hofmann, R. Weinberger, A. S. Schroeder)
37. [Multi-sensor 3D Object Dataset for Object Recognition with Full Pose Estimation](#) - Multi-sensor 3D Object Dataset for Object Recognition and Pose Estimation(Alberto Garcia-Garcia, Sergio Orts-Escolano, Sergiu Oprea,etc.)
 38. [NTU RGB+D Action Recognition Dataset](#) - NTU RGB+D is a large scale dataset for human action recognition(Amir Shahroudy)
 39. [nuTonomy scenes dataset \(nuScenes\)](#) - The nuScenes dataset is a large-scale autonomous driving dataset. It features: Full sensor suite (1x LIDAR, 5x RADAR, 6x camera, IMU, GPS), 1000 scenes of 20s each, 1,440,000 camera images, 400,000 lidar sweeps, two diverse cities: Boston and Singapore, left versus right hand traffic, detailed map information, manual annotations for 25 object classes, 1.1M 3D bounding boxes annotated at 2Hz, attributes such as visibility, activity and pose. (Caesar et al)
 40. [NYU Depth Dataset V2](#) - Indoor Segmentation and Support Inference from RGBD Images
 41. [Oakland 3-D Point Cloud Dataset](#) (Nicolas Vandapel)
 42. [Pacman project](#) - Synthetic RGB-D images of 400 objects from 20 classes. Generated from 3D mesh models (Vladislav Kramarev, Umit Rusen Aktas, Jeremy L. Wyatt.)
 43. [Procedural Human Action Videos](#) - This dataset contains about 40,000 videos for human action recognition that had been generated using a 3D game engine. The dataset contains about 6 million frames which can be used to train and evaluate models not only action recognition but also models for depth map estimation, optical flow, instance segmentation, semantic segmentation, 3D and 2D pose estimation, and attribute learning. (Cesar Roberto de Souza)
 44. [RGB-D-based Action Recognition Datasets](#) - Paper that includes the list and links of different rgb-d action recognition datasets. (Jing Zhang, Wanqing Li, Philip O. Ogunbona, Pichao Wang, Chang Tang)
 45. [RGB-D Part Affordance Dataset](#) - RGB-D images and ground-truth affordance labels for 105 kitchen, workshop and garden tools, and 3 cluttered scenes (Myers, Teo, Fermuller, Aloimonos)
 46. [ScanNet: Richly-annotated 3D Reconstructions of Indoor Scenes](#) - ScanNet is a dataset of richly-annotated RGB-D scans of real-world environments containing 2.5M RGB-D images in more than 1500 scans, annotated with 3D camera poses, surface reconstructions, and instance-level semantic segmentations. (Angela Dai, Angel X. Chang, Manolis Savva, Maciej Halber, Thomas Funkhouser, Matthias Niessner)
 47. [SceneNN: A Scene Meshes Dataset with aNNotations](#) - RGB-D scene dataset with 100+ indoor scenes, labeled triangular mesh, voxel and pixel. (Hua, Pham, Nguyen, Tran, Yu, and Yeung)

48. [Semantic-8](#): 3D point cloud classification with 8 classes (ETH Zurich)
49. [Small office data sets](#) - Kinect depth images every 5 seconds beginning in April 2014 and on-going. (John Folkesson et al.)
50. [Stereo and ToF dataset with ground truth](#) - The dataset contains 5 different scenes acquired with a Time-of-flight sensor and a stereo setup. Ground truth information is also provided.(Carlo Dal Mutto, Pietro Zanuttigh, Guido M. Cortelazzo)
51. [SYNTHIA](#) - Large set (~half million) of virtual-world images for training autonomous cars to see. (ADAS Group at Computer Vision Center)
52. [Taskonomy](#) - Over 4.5 million real images each with ground truth for 25 semantic, 2D, and 3D tasks. (Zamir, Sax, Shen, Guibas, Malik, Savarese)
53. [The Headspace dataset](#) - The Headspace dataset is a set of 3D images of the full human head, consisting of 1519 subjects wearing tight fitting latex caps to reduce the effect of hairstyles. (Christian Duncan, Rachel Armstrong, Alder Hey Craniofacial Unit, Liverpool, UK)
54. [The York 3D Ear Dataset](#) - The York 3D Ear Dataset is a set of 500 3D ear images, synthesized from detailed 2D landmarking, and available in both Matlab format (.mat) and PLY format (.ply). (Nick Pears, Hang Dai, Will Smith, University of York)
55. [THU-READ\(Tsinghua University RGB-D Egocentric Action Dataset\)](#) - THU-READ is a large-scale dataset for action recognition in RGBD videos with pixel-level hand annotation. (Yansong Tang, Yi Tian, Jiwen Lu, Jianjiang Feng, Jie Zhou)
56. [TUM RGB-D Benchmark](#) - Dataset and benchmark for the evaluation of RGB-D visual odometry and SLAM algorithms (Jorgen Sturm, Nikolas Engelhard, Felix Endres, Wolfram Burgard and Daniel Cremers)
57. [UC-3D Motion Database](#) - Available data types encompass high resolution Motion Capture, acquired with MVN Suit from Xsens and Microsoft Kinect RGB and depth images.(Institute of Systems and Robotics, Coimbra, Portugal)
58. [Uni Bremen Open, Abdominal Surgery RGB Dataset](#) - Recording of a complete, open, abdominal surgery using a Kinect v2 that was mounted directly above the patient looking down at patient and staff. (Joern Teuber, Gabriel Zachmann, University of Bremen)
59. [USF Range Image Database](#) - 400+ laser range finder and structured light camera images, many with ground truth segmentations (Adam et al.)
60. [Washington RGB-D Object Dataset](#) - 300 common household objects and 14 scenes. (University of Washington and Intel Labs Seattle)
61. [Witham Wharf](#) - For RGB-D of eight locations collect by robot every 10 min over ~10 days by the University of Lincoln. (John Folkesson et al.)

General Videos

1. [AlignMNIST](#) - An artificially extended version of the MNIST handwritten dataset. (en Hauberg)
2. [Audio-Visual Event \(AVE\) dataset](#)- AVE dataset contains 4143 YouTube videos covering 28 event categories and videos in AVE dataset are temporally labeled with audio-visual event boundaries. (Yapeng Tian, Jing Shi, Bochen Li, Zhiyao Duan, and Chenliang Xu)
3. [Dataset of Multimodal Semantic Egocentric Video \(DoMSEV\)](#) - Labeled 80-hour Dataset of Multimodal Semantic Egocentric Videos (DoMSEV) covering a wide range of activities, scenarios, recorders, illumination and weather conditions. (UFMG, Michel Silva, Washington Ramos, Jo??o Ferreira, Felipe Chamone, Mario Campos, Erickson R. Nascimento)
4. [DAVIS: Video Object Segmentation dataset 2016](#) - A Benchmark Dataset and Evaluation Methodology for Video Object Segmentation (F. Perazzi, J. Pont-Tuset, B. McWilliams, L. Van Gool, M. Gross, and A. Sorkine-Hornung)
5. [DAVIS: Video Object Segmentation dataset 2017](#) - The 2017 DAVIS Challenge on Video Object Segmentation (J. Pont-Tuset, F. Perazzi, S. Caelles, P. Arbelaez, A. Sorkine-Hornung, and L. Van Gool)
6. [GoPro-Gyro Dataset](#) - ego centric videos (Linkoping Computer Vision Laboratory)
7. [Image & Video Quality Assessment at LIVE](#) - used to develop picture quality algorithms (the University of Texas at Austin)
8. [Large scale YouTube video dataset](#) - 156,823 videos (2,907,447 keyframes) crawled from YouTube videos (Yi Yang)
9. [Movie Memorability Dataset](#) - memorable movie clips and ground truth of detail memorability, 660 short movie excerpts extracted from 100 Hollywood-like movies (Cohendet, Yadati, Duong and Demarty)
10. [MovieQA](#) - each machines to understand stories by answering questions about them. 15000 multiple choice QAs, 400+ movies.(M. Tapaswi, Y. Zhu, R. Stiefelhagen, A. Torralba, R. Urtasun, and S. Fidler)
11. [Multispectral visible-NIR video sequences](#) - Annotated multispectral video, visible + NIR (LE2I, Universit de Bourgogne)
12. [Moments in Time Dataset](#) - Moments in Time Dataset 1M 3-second videos annotated with action type, the largest dataset of its kind for action recognition and understanding in video. (Monfort, Oliva, et al.)
13. [Near duplicate video retrieval dataset](#) - This database consists of 156,823 videos sequences (2,907,447 keyframes), which were crawled from YouTube during the period of July 2010 to September 2010. (Jingkuan Song, Yi Yang, Zi Huang, Heng Tao Shen, Richang Hong)
14. [PHD2: Personalized Highlight Detection Dataset](#) - PHD2 is a dataset with personalized highlight information, which allows to train highlight detection models that use information about the user, when making predictions. (Ana Garcia del Molino, Michael Gygli)
15. [Sports-1M](#) - Dataset for sports video classification containing 487 classes and 1.2M videos.(Andrej Karpathy and George Toderici and Sanketh Shetty)

- and Thomas Leung and Rahul Sukthankar and Li Fei-Fei.)
16. [nuTonomy scenes dataset \(nuScenes\)](#) - The nuScenes dataset is a large-scale autonomous driving dataset. It features: Full sensor suite (1x LIDAR, 5x RADAR, 6x camera, IMU, GPS), 1000 scenes of 20s each, 1,440,000 camera images, 400,000 lidar sweeps, two diverse cities: Boston and Singapore, left versus right hand traffic, detailed map information, manual annotations for 25 object classes, 1.1M 3D bounding boxes annotated at 2Hz, attributes such as visibility, activity and pose. (Caesar et al)
 17. [Video Sequences](#) used for research on Euclidean upgrades based on minimal assumptions about the camera (Kenton McHenry)
 18. [Video Stacking Dataset](#) - A Virtual Tripod for Hand-held Video Stacking on Smartphones (Erik Ringaby etc.)
 19. [YFCC100M videos](#) - A benchmark on the video subset of YFCC100M which includes the videos, the video content features and the API to a state-of-the-art video content engine. (Lu Jiang)
 20. [YFCC100M: The New Data in Multimedia Research](#) - This publicly available curated dataset of 100 million photos and videos is free and legal for all. (Bart Thomee, Yahoo Labs and Flickr in San Francisco, etc.)
 21. [YouTube-BoundingBoxes](#) - 5.6 million accurate human-annotated BB from 23 object classes tracked across frames, from 240,000 YouTube videos, with a strong focus on the person class (1.3 million boxes) (Real, Shlens, Pan, Mazzocchi, Vanhoucke, Khan, Kakarla et al)
 22. [YouTube-8M](#) - Dataset for video classification in the wild, containing pre-extracted frame level features from 8M videos, and 4800 classes. (Sami Abu-El-Haija, Nisarg Kothari, Joonseok Lee, Paul Natsev, George Toderici, Balakrishnan Varadarajan, Sudheendra Vijayanarasimhan)
 23. [YUP++ / Dynamic Scenes dataset](#) - 20 outdoor scene classes, each with 60 colour videos (each 5 seconds, 480 pixels wide, 24-30 fps) from 60 different scenes. Half of the videos are with a static camera and half with a moving camera (Feichtenhofer, Pinz, Wildes)

Hand, Hand Grasp, Hand Action and Gesture Databases

1. [11k Hands](#) - 11,076 hand images (1600 x 1200 pixels) of 190 subjects, of varying ages between 18 - 75, with metadata (id, gender, age, skin color, handedness, which hand, accessories, etc). (Mahmoud Afifi)
2. [20bn-Jester](#) - densely-labeled video clips that show humans performing predefined hand gestures in front of a laptop camera or webcam (Twenty Billion Neurons GmbH)
3. [3D Articulated Hand Pose Estimation with Single Depth Images](#) (Tang, Chang, Tejani, Kim, Yu)
4. [A Dataset of Human Manipulation Actions](#) - RGB-D of 25 objects and 6

- actions (Alessandro Pieropan)
5. [A Hand Gesture Detection Dataset](#) (Javier Molina et al)
 6. [A-STAR Annotated Hand-Depth Image Dataset and its Performance Evaluation](#) - depth data and data glove data, 29 images of 30 volunteers, Chinese number counting and American Sign Language (Xu and Cheng)
 7. [Bosphorus Hand Geometry Database and Hand-Vein Database](#) (Bogazici University)
 8. [DemCare dataset](#) - DemCare dataset consists of a set of diverse data collection from different sensors and is useful for human activity recognition from wearable/depth and static IP camera, speech recognition for Alzheimer's disease detection and physiological data for gait analysis and abnormality detection. (K. Avgerinakis, A.Karakostas, S.Vrochidis, I. Kompatsiaris)
 9. [EgoGesture Dataset](#) - First-person view gestures with 83 classes, 50 subjects, 6 scenes, 24161 RGB-D video samples (Zhang, Cao, Cheng, Lu)
 10. [EgoHands](#) - A large dataset with over 15,000 pixel-level-segmented hands recorded from egocentric cameras of people interacting with each other. (Sven Bambach)
 11. [EgoYouTubeHands dataset](#) - An egocentric hand segmentation dataset consists of 1290 annotated frames from YouTube videos recorded in unconstrained real-world settings. The videos have variation in environment, number of participants, and actions. This dataset is useful to study hand segmentation problem in unconstrained settings. (Aisha Urooj, A. Borji)
 12. [FORTH Hand tracking library](#) (FORTH)
 13. [General HANDS: general hand detection and pose challenge](#) - 22 sequences with different gestures, activities and viewpoints (UC Irvine)
 14. [Grasp UNderstanding \(GUN-71\) dataset](#) - 12,000 first-person RGB-D images of object manipulation scenes annotated using a taxonomy of 71 fine-grained grasps.(Rogez, Supancic and Ramanan)
 15. [Hand gesture and marine silhouettes](#) (Euripides G.M. Petrakis)
 16. [HandNet: annotated depth images of articulated hands](#) 214971 annotated depth images of hands captured by a RealSense RGBD sensor of hand poses. Annotations: per pixel classes, 6D fingertip pose, heatmap. Train: 202198, Test: 10000, Validation: 2773. Recorded at GIP Lab, Technion.
 17. [HandOverFace dataset](#) - A hand segmentation dataset consists of 300 annotated frames from the web to study the hand-occluding-face problem. (Aisha Urooj, A. Borji)
 18. [IDIAP Hand pose/gesture datasets](#) (Sebastien Marcel)
 19. [Kinect and Leap motion gesture recognition dataset](#) - The dataset contains 1400 different gestures acquired with both the Leap Motion and the Kinect devices(Giulio Marin, Fabio Dominio, Pietro Zanuttigh)
 20. [Kinect and Leap motion gesture recognition dataset](#) - The dataset contains several different static gestures acquired with the Creative Senz3D

- camera.(A. Memo, L. Minto, P. Zanuttigh)
21. [LISA CVRR-HANDS 3D](#) - 19 gestures performed by 8 subjects as car driver and passengers (Ohn-Bar and Trivedi)
 22. [MPI Dexter 1 Dataset for Evaluation of 3D Articulated Hand Motion Tracking](#) - Dexter 1: 7 sequences of challenging, slow and fast hand motions, RGB + depth (Sridhar, Oulasvirta, Theobalt)
 23. [MSR Realtime and Robust Hand Tracking from Depth](#) - (Qian, Sun, Wei, Tang, Sun)
 24. [Mobile and Webcam Hand images database](#) - MOHI and WEHI - 200 people, 30 images each (Ahmad Hassanat)
 25. [NTU-Microsoft Kinect HandGesture Dataset](#) - This is a RGB-D dataset of hand gestures, 10 subjects x 10 hand gestures x 10 variations. (Zhou Ren, Junsong Yuan, Jingjing Meng, and Zhengyou Zhang)
 26. [NUIG_Palm1](#) - Database of palmprint images acquired in unconstrained conditions using consumer devices for palmprint recognition experiments. (Adrian-Stefan Ungureanu)
 27. [NYU Hand Pose Dataset](#) - 8252 test-set and 72757 training-set frames of captured RGBD data with ground-truth hand-pose, 3 views (Tompson, Stein, Lecun, Perlin)
 28. [PRAXIS gesture dataset](#) - RGB-D upper-body data from 29 gestures, 64 volunteers, several repetitions, many volunteers have some cognitive impairment (Farhood Negin, INRIA)
 29. [Rendered Handpose Dataset](#) - Synthetic dataset for 2D/ 3D Handpose Estimation with RGB, depth, segmentation masks and 21 keypoints per hand (Christian Zimmermann and Thomas Brox)
 30. [Sahand Dynamic Hand Gesture Database](#) - This database contains 11 Dynamic gestures designed to convey the functions of mouse and touch screens to computers.(Behnam Maleki, Hossein Ebrahimnezhad)
 31. [Sheffield gesture database](#) - 2160 RGBD hand gesture sequences, 6 subjects, 10 gestures, 3 postures, 3 backgrounds, 2 illuminations (Ling Shao)
 32. [UT Grasp Data Set](#) - 4 subjects grasping a variety of objects with a variety of grasps (Cai, Kitani, Sato)
 33. [Yale human grasping data set](#) - 27 hours of video with tagged grasp, object, and task data from two housekeepers and two machinists (Bullock, Feix, Dollar)

Image, Video and Shape Database Retrieval

1. [2D-to-3D Deformable Sketches](#) - A collection of deformable 2D contours in pointwise correspondence with deformable 3D meshes of the same class; around 10 object classes are provided, including humans and animals. (Lahner, Rodola)

2. [3D Deformable Objects in Clutter](#) - A dataset for 3D deformable object-in-clutter, with point-wise ground truth correspondence across hundreds of scenes and spanning multiple classes (humans, animals). (Cosmo, Rodola, Masci, Torsello, Bronstein)
3. [ANN_SIFT1M](#) - 1M Flickr images encoded by 128D SIFT descriptors (Jegou et al)
4. [Brown Univ 25/99/216 Shape Databases](#) (Ben Kimia)
5. [CIFAR-10](#) - 60K 32x32 images from 10 classes, with a 512D GIST descriptor (Alex Krizhevsky)
6. [CLEF-IP 2011 evaluation on patent images](#)
7. [DeepFashion](#) - Large-scale Fashion Database(Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, Xiaoou Tang)
8. [EMODB](#) - Thumbnails of images in the picsearch image search engine together with the picsearch emotion keywords (Reiner Lenz etc.)
9. [ETU10 Silhouette Dataset](#) - The dataset consists of 720 silhouettes of 10 objects, with 72 views per object.(M. Akimaliev and M.F. Demirci)
10. [European Flood 2013](#) - 3,710 images of a flood event in central Europe, annotated with relevance regarding 3 image retrieval tasks (multi-label) and important image regions. (Friedrich Schiller University Jena, Deutsches GeoForschungsZentrum Potsdam)
11. [Fashion-MNIST](#) - A MNIST-like fashion product database. (Han Xiao, Zalando Research)
12. [Fish Shape Database](#) - It's a Fish Shape Database with 100, 2D point set shapes. (Adrian M. Peter)
13. [Flickr 30K](#) - images, actions and captions (Peter Young et al)
14. [Flickr15k - Sketch based Image Retrieval \(SBIR\) Benchmark](#) - Dataset of 330 sketches and 15,024 photos comprising 33 object categories,benchmark dataset commonly used to evaluate Sketch based Image Retrieval (SBIR) algorithms.(Hu and Collomosse, CVIU 2013)
15. [Hands in action \(HIC\) IJCV dataset](#) - Data (images, models, motion) for tracking 1 hand or 2 hands with/o 1 object. Includes both *single-view RGB-D sequences (1 subject, >18 annotated sequences, 4 objects, complete RGB image), and *multi-view RGB sequences (1 subject, HD, 8 views, 8 sequences - 1 annotated, 2 objects). (Dimitrios Tzionas, Luca Ballan, Abhilash Srikantha, Pablo Aponte, Marc Pollefeys, Juergen Gall)
16. [IAPR TC-12 Image Benchmark](#) (Michael Grubinger)
17. [IAPR-TC12 Segmented and annotated image benchmark \(SAIAPR TC-12\):](#) (Hugo Jair Escalante)
18. [ImageCLEF 2010 Concept Detection and Annotation Task](#) (Stefanie Nowak)
19. [ImageCLEF 2011 Concept Detection and Annotation Task](#) - multi-label classification challenge in Flickr photos
20. [METU Trademark dataset](#)The METU Dataset is composed of more than 900K real logos belonging to companies worldwide. (Usta Bilgi Sistemleri A.S. and Grup Ofis Marka Patent A.S)

21. [McGill 3D Shape Benchmark](#) (Siddiqi, Zhang, Macrini, Shokoufandeh, Bouix, Dickinson)
22. [MPI MANO & SMPL+H dataset](#) - Models, 4D scans and registrations for the statistical models MANO (hand-only) and SMPL+H (body+hands). For MANO there are ~2k static 3D scans of 31 subjects performing up to 51 poses. For SMPL+H we include 39 4D sequences of 11 subjects. (Javier Romero, Dimitrios Tzionas and Michael J Black)
23. [Multiview Stereo Evaluation](#) - Each dataset is registered with a "ground-truth" 3D model acquired via a laser scanning process(Steve Seitz et al)
24. [NIST SHREC - 2014 NIST retrieval contest databases and links](#) (USA National Institute of Standards and Technology)
25. [NIST SHREC - 2013 NIST retrieval contest databases and links](#) (USA National Institute of Standards and Technology)
26. [NIST SHREC 2010 - Shape Retrieval Contest of Non-rigid 3D Models](#) (USA National Institute of Standards and Technology)
27. [NIST TREC Video Retrieval Evaluation Database](#) (USA National Institute of Standards and Technology)
28. [NUS-WIDE](#) - 269K Flickr images annotated with 81 concept tags, enclded as a 500D BoVW descriptor (Chau et al)
29. [Princeton Shape Benchmark](#) (Princeton Shape Retrieval and Analysis Group)
30. [PairedFrames - evaluation of 3D pose tracking error](#) - Synthetic and Real dataset to test 3D pose tracking/refinement with pose initialization close/far to/from minima. Establishes testing frame pairs of increasing difficulty, to measure the pose estimation error separately, without employing a full tracking pipeline. (Dimitrios Tzionas, Juergen Gall)
31. [Queensland cross media dataset](#) - millions of images and text documents for "cross-media" retrieval (Yi Yang)
32. [Reconstructing Articulated Riggged Models from RGB-D Videos \(RecArt-D\)](#) - Dataset of objects deforming during manipulation. Includes 4 RGB-D sequences (RGB image complete), result of deformable tracking for each object, as well as 3D mesh and Ground-Truth 3D skeleton for each object. (Dimitrios Tzionas, Juergen Gall)
33. [Reconstruction from Hand-Object Interactions \(R-HOI\)](#) - Dataset of one hand interacting with an unknown object. Includes 4 RGB-D sequences, in total 4 objects, the RGB image is complete. Includes tracked 3D motion and Ground-Truth meshes for the objects. (Dimitrios Tzionas, Juergen Gall)
34. [Revisiting Oxford and Paris \(RevisitOP\)](#) - Improved and more challenging version (fixed errors, new annotation and evaluation protocols, new query images) of the well known landmark/building retrieval datasets accompanied with 1M distractor images. (F. Radenovic, A. Iscen, G. Tolas, Y. Avrithis, O. Chum)
35. [SHREC'16 Deformable Partial Shape Matching](#) - A collection of around 400 3D deformable shapes undergoing strong partiality transformations, with

- point-to-point ground truth correspondence included. (Cosmo, Rodola, Bronstein, Torsello)
36. [SHREC 2016 - 3D Sketch-Based 3D Shape Retrieval](#) - data to evaluate the performance of different 3D sketch-based 3D model retrieval algorithms using a hand-drawn 3D sketch query dataset on a generic 3D model dataset (Bo Li)
 37. [SHREC'17 Deformable Partial Shape Retrieval](#) - A collection of around 4000 deformable 3D shapes undergoing severe partiality transformations, in the form of irregular missing parts and range data; ground truth class information is provided. (Lahner, Rodola)
 38. [SHREC Watertight Models Track \(of SHREC 2007\)](#) - 400 watertight 3D models (Daniela Giorgi)
 39. [SHREC Partial Models Track \(of SHREC 2007\)](#) - 400 watertight 3D DB models and 30 reduced watertight query models (Daniela Giorgi)
 40. [SBU Captions Dataset](#) - image captions collected for 1 million images from Flickr (Ordonez, Kulkarni and Berg)
 41. [Sketch me That Shoe](#) - Sketch-based object retrieval in a fine-grained setting. Match sketches to specific shoes and chairs. (Qian Yu, QMUL, T. Hospedales Edinburgh/QMUL).
 42. [TOSCA 3D shape database](#) (Bronstein, Bronstein, Kimmel)
 43. [Totally Looks Like](#) - A benchmark for assessment of predicting human-based image similarity (Amir Rosenfeld, Markus D. Solbach, John Tsotsos)
 44. [UCF-CrossView Dataset: Cross-View Image Matching for Geo-localization in Urban Environments](#) - A new dataset of street view and bird's eye view images for cross-view image geo-localization. (Center for Research in Computer Vision, University of Central Florida)
 45. [YouTube-8M Dataset](#) - A Large and Diverse Labeled Video Dataset for Video Understanding Research. (Google Inc.)

Object Databases

1. [2.5D/3D Datasets of various objects and scenes](#) (Ajmal Mian)
2. [3D Object Recognition Stereo Dataset](#) This dataset consists of 9 objects and 80 test images. (Akash Kushal and Jean Ponce)
3. [3D Photography Dataset](#) a collection of ten multiview data sets captured in our lab (Yasutaka Furukawa and Jean Ponce)
4. [3D-Printed RGB-D Object Dataset](#) - 5 objects with groundtruth CAD models and camera trajectories, recorded with various quality RGB-D sensors (Siemens & TUM)
5. [3DNet Dataset](#) - The 3DNet dataset is a free resource for object class recognition and 6DOF pose estimation from point cloud data. (John Folkesson et al.)
6. [Aligned 2.5D/3D datasets of various objects](#) - Synthesized and real-world

- datasets for object reconstruction from a single depth view. (Bo Yang, Stefano Rosa, Andrew Markham, Niki Trigoni, Hongkai Wen)
7. [Amsterdam Library of Object Images \(ALOI\): 100K views of 1K objects](#) (University of Amsterdam/Intelligent Sensory Information Systems)
 8. [Animals with Attributes 2](#) - 37322 (freely licensed) images of 50 animal classes with 85 per-class binary attributes. (Christoph H. Lampert, IST Austria)
 9. [ASU Office-Home Dataset](#) - Object recognition dataset of everyday objects for domain adaptation (Venkateswara, Eusebio, Chakraborty, Panchanathan)
 10. [B3DO: Berkeley 3-D Object Dataset](#) - household object detection (Janoch et al)
 11. [Bristol Egocentric Object Interactions Dataset](#) - egocentric object interactions with synchronised gaze (Dima Damen)
 12. [CORE image dataset](#) - to help learn more detailed models and for exploring cross-category generalization in object recognition. (Ali Farhadi, Ian Endres, Derek Hoiem, and David A. Forsyth)
 13. [CTU Color and Depth Image Dataset of Spread Garments](#) - Images of spread garments with annotated corners.(Wagner, L., Krejov D., and Smutn V. (Czech Technical University in Prague))
 14. [Caltech 101 \(now 256\) category object recognition database](#) (Li Fei-Fei, Marco Andreeto, Marc'Aurelio Ranzato)
 15. [Catania Fish Species Recognition](#) - 15 fish species, with about 20,000 sample training images and additional test images (Concetto Spampinato)
 16. [COCO-Stuff dataset](#) - 164K images labeled with 'things' and 'stuff' (Caesar, Uijlings, Ferrari)
 17. [Columbia COIL-100 3D object multiple views](#) (Columbia University)
 18. [Deeper, Broader and Artier Domain Generalization](#) - Domain generalisation task dataset. (Da Li, QMUL)
 19. [Densely sampled object views: 2500 views of 2 objects, eg for view-based recognition and modeling](#) (Gabriele Peters, Universiteit Dortmund)
 20. [Edinburgh Kitchen Utensil Database](#) - 897 raw and binary images of 20 categories of kitchen utensil, a resource for training future domestic assistance robots (D. Fullerton, A. Goel, R. B. Fisher)
 21. [EDUB-Obj](#) - Egocentric dataset for object localization and segmentation.(Marc Bolaños and Petia Radeva.)
 22. [Ellipse finding dataset](#) (Dilip K. Prasad et al)
 23. [FIN-Benthic](#) - This is a dataset for automatic fine-grained classification of benthic macroinvertebrates. There are 15074 images from 64 categories. The number of images per category varies from 577 to 7. (Jenni Raitoharju, Ekaterina Riabchenko, Iftikhar Ahmad, Alexandros Iosifidis, Moncef Gabbouj, Serkan Kiranyaz, Ville Tirronen, Johanna Arje)
 24. [GERMS](#) - The object set we use for GERMS data collection consists of 136 stuffed toys of different microorganisms. The toys are divided into 7 smaller

categories, formed by semantic division of the toy microbes. The motivation for dividing the objects into smaller categories is to provide benchmarks with different degrees of difficulty. (Malmir M, Sikka K, Forster D, Movellan JR, Cottrell G.)

25. [GDxray: X-ray images for X-ray testing and Computer Vision](#) - GDxray includes five groups of images: Castings, Welds*, Baggages, Nature and Settings. (Domingo Mery, Catholic University of Chile)
26. [GMU Kitchens Dataset](#) - instance level annotation of 11 common household products from BigBird dataset across 9 different kitchens (George Mason University)
27. [Grasping In The Wild](#) - Egocentric video dataset of natural everyday life objects. 16 objects in 7 kitchens. (Benois-Pineau, Larrousse, de Ruyg)
28. [GRAZ-02 Database \(Bikes, cars, people\)](#) (A. Pinz)
29. [GREYC 3D](#) - The GREYC 3D Colored mesh database is a set of 15 real objects with different colors, geometries and textures that were acquired using a 3D color laser scanner. (Anass Nouri, Christophe Charrier, Olivier Lezoray)
30. [GTSDb: German Traffic Sign Detection Benchmark](#) (Ruhr-Universität Bochum)
31. [iCubWorld](#) - iCubWorld datasets are collections of images acquired by recording from the cameras of the iCub humanoid robot while it observes daily objects. (Giulia Pasquale, Carlo Ciliberto, Giorgio Metta, Lorenzo Natale, Francesca Odone and Lorenzo Rosasco.)
32. [Industrial 3D Object Detection Dataset \(MVTec ITODD\)](#) - depth and gray value data of 28 objects in 3500 labeled scenes for 3D object detection and pose estimation with a strong focus on industrial settings and applications (MVTec Software GmbH, Munich)
33. [Instagram Food Dataset](#) - A database of 800,000 food images and associated metadata posted to Instagram over 6 week period. Supports food type recognition and social network analysis. (T. Hospedales. Edinburgh/QMUL)
34. [Keypoint-5 dataset](#) - a dataset of five kinds of furniture with their 2D keypoint labels (Jiajun Wu, Tianfan Xue, Joseph Lim, Yuandong Tian, Josh Tenenbaum, Antonio Torralba, Bill Freeman)
35. [KTH-3D-TOTAL](#) - RGB-D Data with objects on desktops annotated. 20 Desks, 3 times per day, over 19 days. (John Folkesson et al.)
36. [LISA Traffic Light Dataset](#) - 6 light classes in various lighting conditions (Jensen, Philipsen, Mogelmoose, Moeslund, and Trivedi)
37. [LISA Traffic Sign Dataset](#) - video of 47 US sign types with 7855 annotations on 6610 frames (Mogelmoose, Trivedi, and Moeslund)
38. [Linköping 3D Object Pose Estimation Database](#) (Fredrik Viksten and Per-Erik Forssen)
39. [Linköping Traffic Signs Dataset](#) - 3488 traffic signs in 20K images (Larsson and Felsberg)

40. [Longterm Labeled](#) - This dataset contains a subset of the observations from the longterm dataset (longterm dataset above). (John Folkesson et al.)
41. [Main Product Detection Dataset](#) - Contains textual metadata of fashion products and their images with bounding boxes of the main product (the one referred by the text). (A. Rubio, L. Yu, E. Simo-Serra and F. Moreno-Noguer)
42. [MCIndoor20000](#) - 20,000 digital images from three different indoor object categories: doors, stairs, and hospital signs. (Bashiri, LaRose, Peissig, and Tafti)
43. [Mexculture142](#) - Mexican Cultural heritage objects and eye-tracker gaze fixations (Montoya Obeso, Benois-Pineau, Garcia-Vazquez, Ramirez Acosta)
44. [MIT CBCL Car Data](#) (Center for Biological and Computational Learning)
45. [MIT CBCL StreetScenes Challenge Framework](#): (Stan Bileschi)
46. [Microsoft COCO](#) - Common Objects in Context (Tsung-Yi Lin et al)
47. [Microsoft Object Class Recognition image databases](#) (Antonio Criminisi, Pushmeet Kohli, Tom Minka, Carsten Rother, Toby Sharp, Jamie Shotton, John Winn)
48. [Microsoft salient object databases \(labeled by bounding boxes\)](#) (Liu, Sun Zheng, Tang, Shum)
49. [Moving Labeled](#) - This dataset extends the longterm dataset with more locations within the same office environment at KTH. (John Folkesson et al.)
50. [NABirds Dataset](#) - 70,000 annotated photographs of the 400 species of birds commonly observed in North America (Grant Van Horn)
51. [NEC Toy animal object recognition or categorization database](#) (Hossein Mobahi)
52. [NORB 50 toy image database](#) (NYU)
53. [NTU-VOI: NTU Video Object Instance Dataset](#) - video clips with frame-level bounding box annotations of object instances for evaluating object instance search and localization in large scale videos.(Jingjing Meng, et. al.)
54. [Object Pose Estimation Database](#) - This database contains 16 objects, each sampled at 5 degrees angle increments along two rotational axes (F. Viksten etc.)
55. [Object Recognition Database](#) This database features modeling shots of eight objects and 51 cluttered test shots containing multiple objects.(Fred Rothganger, Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce.)
56. [Open Images Dataset V4](#) 15,440,132 boxes on 600 categories, 30,113,078 image-level labels on 19,794 categories. (Ferrari, Duerig, Gomes)
57. [Open Museum Identification Challenge \(Open MIC\)](#) Open MIC contains photos of exhibits captured in 10 distinct exhibition spaces (painting, sculptures, jewellery, etc.) of several museums and the protocols for the domain adaptation and few-shot learning problems. (P. Koniusz, Y. Tas, H. Zhang, M. Harandi, F. Porikli, R. Zhang)
58. [Osnabrück Synthetic Scalable Cube Dataset](#) - 830000 different cubes captured from 12 different viewpoints for ANN training (Schöning,

- Behrens, Faion, Kheiri, Heidemann & Krumnack)
59. [Princeton ModelNet](#) - 127,915 CAD Models, 662 Object Categories, 10 Categories with Annotated Orientation (Wu, Song, Khosla, Yu, Zhang, Tang, Xiao)
 60. [PacMan datasets](#) - RGB and 3D synthetic and real data for graspable cookware and crockery (Jeremy Wyatt)
 61. [PACS \(Photo Art Cartoon Sketch\)](#) - An object category recognition dataset for testing domain generalisation: How well can a classifier trained on object images in one domain recognise objects in another domain? (Da Li QMUL, T. Hospedales. Edinburgh/QMUL)
 62. [PASCAL 2007 Challenge Image Database \(motorbikes, cars, cows\)](#) (PASCAL Consortium)
 63. [PASCAL 2008 Challenge Image Database](#) (PASCAL Consortium)
 64. [PASCAL 2009 Challenge Image Database](#) (PASCAL Consortium)
 65. [PASCAL 2010 Challenge Image Database](#) (PASCAL Consortium)
 66. [PASCAL 2011 Challenge Image Database](#) (PASCAL Consortium)
 67. [PASCAL 2012 Challenge Image Database](#) Category classification, detection, and segmentation, and still-image action classification (PASCAL Consortium)
 68. [PASCAL Image Database \(motorbikes, cars, cows\)](#) (PASCAL Consortium)
 69. [PASCAL Parts dataset](#) - PASCAL VOC with segmentation annotation for semantic parts of objects (Alan Yuille)
 70. [PASCAL-Context dataset](#) - annotations for 400+ additional categories (Alan Yuille)
 71. [PASCAL 3D/Beyond PASCAL: A Benchmark for 3D Object Detection in the Wild](#) - 12 class, 3000+ images each with 3D annotations (Yu Xiang, Roozbeh Mottaghi, Silvio Savarese)
 72. [Physics 101 dataset](#) - a video dataset of 101 objects in five different scenarios (Jiajun Wu, Joseph Lim, Hongyi Zhang, Josh Tenenbaum, Bill Freeman)
 73. [Plant seedlings dataset](#) - High-resolution images of 12 weed species. (Aarhus University)
 74. [Raindrop Detection](#) - Improved Raindrop Detection using Combined Shape and Saliency Descriptors with Scene Context Isolation - Evaluation Dataset (Breckon, Toby P., Webster, Dereck D.)
 75. [ReferIt Dataset \(IAPRTC-12 and MS-COCO\)](#) - referring expressions for objects in images from the IAPRTC-12 and MS-COCO datasets (Kazemzadeh, Matten, Ordonez, and Berg)
 76. [ShapeNet](#) - 3D models of 55 common object categories with about 51K unique 3D models. Also 12K models over 270 categories. (Princeton, Stanford and TTIC)
 77. [SHORT-100 dataset](#) - 100 categories of products found on a typical shopping list. It aims to benchmark the performance of algorithms for recognising hand-held objects from either snapshots or videos acquired

- using hand-held or wearable cameras. (Jose Rivera-Rubio, Saad Idrees, Anil A. Bharath)
78. [SOR3D](#) - The SOR3D dataset consists of over 20k instances of human-object interactions, 14 object types, and 13 object affordances. (pyridon Thermos)
 79. [Stanford Dogs Dataset](#) - The Stanford Dogs dataset contains images of 120 breeds of dogs from around the world. This dataset has been built using images and annotation from ImageNet for the task of fine-grained image categorization. (Aditya Khosla, Nityananda Jayadevaprakash, Bangpeng Yao, Li Fei-Fei, Stanford University)
 80. [SVHN: Street View House Numbers Dataset](#) - like MNIST, but an order of magnitude more labeled data (over 600,000 digit images) and comes from a significantly harder, unsolved, real world problem (recognizing digits and numbers in natural scene images). (Netzer, Wang, Coates, Bissacco, Wu, Ng)
 81. [Swedish Leaf Dataset](#) - These images contains leaves from 15 treeclasses (Oskar J. O. Söderkvist)
 82. [T-LESS](#) - An RGB-D dataset for 6D pose estimation of texture-less objects. (Tomas Hodan, Pavel Haluza, Stepan Obdrzalek, Jiri Matas, Manolis Lourakis, Xenophon Zabulis)
 83. [Taobao Commodity Dataset](#) - TCD contains 800 commodity images (dresses, jeans, T-shirts, shoes and hats) for image salient object detection from the shops on the Taobao website. (Keze Wang, Keyang Shi, Liang Lin, Chenglong Li)
 84. [The Laval 6 DOF Object Tracking Dataset](#) - A Dataset of 297 RGB-D sequences with 11 objects for 6 DOF object Tracking. (Mathieu Garon, Denis Laurendeau, Jean-Francois Lalonde)
 85. [ToolArtec point clouds](#) - 50 kitchen tool 3D scans (ply) from an Artec EVA scanner. See also [ToolKinect](#) - 13 scans using a Kinect 2 and [ToolWeb](#) - 116 point clouds of synthetic household tools with mass and affordance groundtruth for 5 tasks. (Paulo Abelha)
 86. [TUW Object Instance Recognition Dataset](#) - Annotations of object instances and their 6DoF pose for cluttered indoor scenes observed from various viewpoints and represented as Kinect RGB-D point clouds (Thomas, A. Aldoma, M. Zillich, M. Vincze)
 87. [TUW dat sets](#) - Several RGB-D Ground truth and annotated data sets from TUW. (John Folkesson et al.)
 88. [UAH Traffic Signs Dataset](#) (Arroyo etc.)
 89. [UIUC Car Image Database](#) (UIUC)
 90. [UIUC Dataset of 3D object categories](#) (S. Savarese and L. Fei-Fei)
 91. [VAIS](#) - VAIS contains simultaneously acquired unregistered thermal and visible images of ships acquired from piers, and it was created to facilitate autonomous ship development. (Mabel Zhang, Jean Choi, Michael Wolf, Kostas Daniilidis, Christopher Kanan)
 92. [Venezia 3D object-in-clutter recognition and segmentation](#) (Emanuele

Rodola)

93. [Visual Attributes Dataset](#) visual attribute annotations for over 500 object classes (animate and inanimate) which are all represented in ImageNet. Each object class is annotated with visual attributes based on a taxonomy of 636 attributes (e.g., has fur, made of metal, is round).
94. [Visual Hull Data Sets](#) a collection of visual hull datasets (Svetlana Lazebnik, Yasutaka Furukawa, and Jean Ponce)
95. [YouTube-BoundingBoxes](#) - 5.6 million accurate human-annotated BB from 23 object classes tracked across frames, from 240,000 YouTube videos, with a strong focus on the person class (1.3 million boxes) (Real, Shlens, Pan, Mazzocchi, Vanhoucke, Khan, Kakarla et al)

People (static and dynamic), human body pose

1. [3D articulated body](#) - 3D reconstruction of an articulated body with rotation and translation. Single camera, varying focal. Every scene may have an articulated body moving. There are four kinds of data sets included. A sample reconstruction result included which uses only four images of the scene. (Prof Jihun Park)
2. [BUFF dataset](#) - About 10K scans of people in clothing and the estimated body shape of people underneath. Scans contain texture so synthetic videos/images are easy to generate.(Zhang, Pujades, Black and Pons-Moll)
3. [Dynamic Dyna](#) - More than 40K 4D 60fps high resolution scans and models of people very accurately registered. Scans contain texture so synthetic videos/images are easy to generate. (Pons-Moll, Romero, Mahmood and Black)
4. [Dynamic Faust](#) - More than 40K 4D 60fps high resolution scans of people very accurately registered. Scans contain texture so synthetic videos/images are easy to generate. (Bogo, Romero, Pons-Moll and Black)
5. [Extended Chictopia dataset](#) - 14K image Chictopia dataset with additional processed annotations (face) and SMPL body model fits to the images. (Lassner, Pons-Moll and Gehler)
6. [Frames Labeled In Cinema \(FLIC\)](#) - 20928 frames labeled with human pose (Sapp, Taskar)
7. [KIDS dataset](#) - A collection of 30 high-resolution 3D shapes undergoing nearly-isometric and non-isometric deformations, with point-to-point ground truth as well as ground truth for left-to-right bilateral symmetry. (Rodola, Rota Bulo, Windheuser, Vestner, Cremers)
8. [Kinect2 Human Pose Dataset \(K2HPD\)](#) - Kinect2 Human Pose Dataset (K2HPD) includes about 100K depth images with various human poses under challenging scenarios. (Keze Wang, Liang Lin, Shengfu Zhai, Dengke Dong)
9. [Leeds Sports Pose Dataset](#) - 2000 pose annotated images of mostly sports

- people (Johnson, Everingham)
10. [Look into Person Dataset](#) - 50,000 images with elaborated pixel-wise annotations with 19 semantic human part labels and 2D hposes with 16 key points. (Gong, Liang, Zhang, Shen, Lin)
 11. [Mannequin in-bed pose datasets via RGB webcam](#) - This in-bed pose dataset is collected via regular webcam in a simulated hospital room at Northeastern University. (Shuangjun Liu and Sarah Ostadabbas, ACLab)
 12. [Mannequin IRS in-bed dataset](#) - This in-bed pose dataset is collected via our infrared selective (IRS) system in a simulated hospital room at Northeastern University. (Shuangjun Liu and Sarah Ostadabbas, ACLab)
 13. [MoPoTS-3D](#) - Multi-person 3D body pose benchmark for monocular RGB based methods, with 20 sequences in indoor and outdoor settings (MPI For Informatics)
 14. [MPI-INF-3DHP](#) - Single-person 3D body pose dataset and evaluation benchmark, with extensive pose coverage across a broad set of activities, and extensive scope of appearance augmentation. Multi-view RGB frames are available for the training set, and monocular view frames for the test set. (MPI For Informatics)
 15. [MPI MANO & SMPL+H dataset](#) - Models, 4D scans and registrations for the statistical models MANO (hand-only) and SMPL+H (body+hands). For MANO there are ~2k static 3D scans of 31 subjects performing up to 51 poses. For SMPL+H we include 39 4D sequences of 11 subjects. (Javier Romero, Dimitrios Tzionas and Michael J Black)
 16. [MPII Human Pose Dataset](#) - 25K images containing over 40K people with annotated body joints, 410 human activities {Andriluka, Pishchulin, Gehler, Schiele)
 17. [MPII Human Pose Dataset](#) - MPII Human Pose dataset is a de-facto standard benchmark for evaluation of articulated human pose estimation. (Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, Bernt Schiele)
 18. [MuCo-3DHP](#) - Large scale dataset of composited multi-person RGB images with 3D pose annotations, generated from MPI-INF-3DHP dataset (MPI For Informatics)
 19. [People In Photo Albums](#) - Social media photo dataset with images from Flickr, and manual annotations on person heads and their identities.(Ning Zhang and Manohar Paluri and Yaniv Taigman and Rob Fergus and Lubomir Bourdev)
 20. [People Snapshot Dataset](#) - Monocular video of 24 subjects rotating in front of a fixed camera. Annotation in form of segmentation and 2D joint positions is provided. (Alldieck, Magnor, Xu, Theobalt, Pons-Moll)
 21. [Person Recognition in Personal Photo Collections](#) - we introduced three harder splits for evaluation and long-term attribute annotations and per-photo timestamp metadata.(Oh, Seong Joon and Benenson, Rodrigo and Fritz, Mario and Schiele, Bernt)
 22. [Pointing'04 ICPR Workshop Head Pose Image Database](#)

23. [Pose estimation](#) - This dataset has a total of 155,530 images. These images were obtained through the recording of members of CIDIS, in 4 sessions. In total, 10 videos with a duration of 4 minutes each were obtained. The participants were asked to bring different clothes, in order to give variety to the images. After this, the frames of the videos were separated at a rate of 5 frames per second. All these images were captured from a top view perspective. The original images have a resolution of 1280x720 pixels. (CIDIS)
24. [SHREC'16 Topological KIDS](#) - A collection of 40 high-resolution and low-resolution 3D shapes undergoing nearly-isometric deformations in addition to strong topological artifacts, self-contacts and mesh gluing, with point-to-point ground truth. (Lahner, Rodola)
25. [SURREAL](#) - 60,000 synthetic videos of people under large variations in shape, texture, view-point and pose. (Varol, Romero, Martin, Mahmood, Black, Laptev, Schmid)
26. [TNT 15 dataset](#) - Several sequences of video synchronised by 10 Inertial Sensors (IMU) worn at the extremities. (von Marcard, Pons-Moll and Rosenhahn)
27. [UC-3D Motion Database](#) - Available data types encompass high resolution Motion Capture, acquired with MVN Suit from Xsens and Microsoft Kinect RGB and depth images. (Institute of Systems and Robotics, Coimbra, Portugal)
28. [United People \(UP\) Dataset](#) - ~8,000 images with keypoint and foreground segmentation annotations as well as 3D body model fits. (Lassner, Romero, Kiefel, Bogo, Black, Gehler)
29. [VGG Human Pose Estimation datasets](#) including the BBC Pose (20 videos with an overlaid sign language interpreter), Extended BBC Pose (72 additional training videos), Short BBC Pose (5 one hour videos with sign language signers), and ChaLearn Pose (23 hours of Kinect data of 27 persons performing 20 Italian gestures). (Charles, Everingham, Pfister, Magee, Hogg, Simonyan, Zisserman)
30. [VRLF: Visual Lip Reading Feasibility](#) - audio-visual corpus of 24 speakers recorded in Spanish (Fernandez-Lopez, Martinez and Sukno)

People Detection and Tracking Databases

1. [3D KINECT Gender Walking data base](#) (L. Igual, A. Lapedriza, R. Borrajo; from UB, CVC and UOC, Spain)
2. [AGORASET: a dataset for crowd video analysis](#) (Nicolas Courty et al)
3. [CASIA gait database](#) (Chinese Academy of Sciences)
4. [CAVIAR project video sequences with tracking and behavior ground truth](#) (CAVIAR team/Edinburgh University - EC project IST-2001-37540)
5. [CMU Panoptic Studio Dataset](#) - Multiple people social interaction dataset

- captured by 500+ synchronized video cameras, with 3D full body skeletons and calibration data. (H. Joo, T. Simon, Y. Sheikh)
6. [CUHK Crowd Dataset](#) - 474 video clips from 215 crowded scenes (Shao, Loy, and Wang)
 7. [CUHK01 Dataset](#) : Person re-id dataset with 3, 884 images of 972 pedestrians (Rui Zhao et al)
 8. [CUHK02 Dataset](#) : Person re-id dataset with five camera view settings. (Rui Zhao et al)
 9. [CUHK03 Dataset](#) : Person re-id dataset with 13,164 images of 1,360 pedestrians (Rui Zhao et al)
 10. [Caltech Pedestrian Dataset](#) (P. Dollar, C. Wojek, B. Schiele and P. Perona)
 11. [Daimler Pedestrian Detection Benchmark](#) 21790 images with 56492 pedestrians plus empty scenes. (D. M. Gavrila et al)
 12. [Datasets \(Color & Infrared\) for Fusion](#) A series of images in color and infrared captured from a parallel two-camera setup under different environmental conditions. (Juan Serrano-Cuerda, Antonio Fernandez-Caballero, Maria T. Lopez)
 13. [Driver Monitoring Video Dataset](#) (RobeSafe + Jesus Nuevo-Chiquero)
 14. [DukeMTMC: Duke Multi-Target Multi-Camera tracking dataset](#) - 8 cameras, 85 min, 2m frames, 2000 people of video (Ergys Ristani, Francesco Solera, Roger S. Zou, Rita Cucchiara, Carlo Tomasi)
 15. [Edinburgh overhead camera person tracking dataset](#) (Bob Fisher, Bashia Majecka, Gurkirt Singh, Rowland Sillito)
 16. [GVVPerfcapEva](#) - Repository of human shape and performance capture data, including full body skeletal, hand tracking, body shape, face performance, interactions (Christian Theobalt)
 17. [HAT](#) Database of 27 human attributes (Gaurav Sharma, Frederic Jurie)
 18. [Immediacy Dataset](#) - This dataset is designed for estimation personal relationships. (Xiao Chu et al.)
 19. [Inria Dressed human bodies in motion benchmark](#) - Benchmark containing 3D motion sequences of different subjects, motions, and clothing styles that allows to quantitatively measure the accuracy of body shape estimates. (Jinlong Yang, Jean-Sbastien Franco, Franck H=E9troy-Wheeler, and Stefanie Wuhrer)
 20. [INRIA Person Dataset](#) (Navneet Dalal)
 21. [IU ShareView](#) - IU ShareView dataset consists of nine sets of synchronized (two first-person) videos with a total of 1,227 pixel-level ground truth segmentation maps of 2,654 annotated person instances. (Mingze Xu, Chenyou Fan, Yuchen Wang, Michael S. Ryoo, David J. Crandall)
 22. [Izmir](#) - omnidirectional and panoramic image dataset (with annotations) to be used for human and car detection (Yalin Bastanlar)
 23. [Joint Attention in Autonomous Driving \(JAAD\)](#) - The dataset includes instances of pedestrians and cars intended primarily for the purpose of behavioural studies and detection in the context of autonomous driving.

- (Iuliia Kotseruba, Amir Rasouli and John K. Tsotsos)
24. [JTL Stereo Tacking Dataset for Person Following Robots](#) - 11 different indoor and outdoor places for the task of robots following people under challenging situations (Chen, Sahdev, Tsotsos)
 25. [MAHNOB: MHI-Mimicry database](#) - A 2 person, multiple camera and microphone database for studying mimicry in human-human interaction scenarios. (Sun, Lichtenauer, Valstar, Nijholt, and Pantic)
 26. [MIT CBCL Pedestrian Data](#) (Center for Biological and Computational Learning)
 27. [MPI DYNA](#) - A Model of Dynamic Human Shape in Motion (Max Planck Tubingen)
 28. [MPI FAUST Dataset](#) A data set containing 300 real, high-resolution human scans, with automatically computed ground-truth correspondences (Max Planck Tubingen)
 29. [MPI JHMDB dataset](#) - Joint-annotated Human Motion Data Base - 21 actions, 928 clips, 33183 frames (Jhuang, Gall, Zuffi, Schmid and Black)
 30. [MPI MOSH](#) Motion and Shape Capture from Markers. MOCAP data, 3D shape meshes, 3D high resolution scans. (Max Planck Tubingen)
 31. [MVHAUS-PI](#) - a multi-view human interaction recognition dataset (Saeid et al.)
 32. [Market-1501 Dataset](#) - 32,668 annotated bounding boxes of 1,501 identities from up to 6 cameras (Liang Zheng et al)
 33. [Modena and Reggio Emilia first person head motion videos](#) (Univ of Modena and Reggio Emilia)
 34. [Multimodal Activities of Daily Living](#) - including video, audio, physiological, sleep, motion and plug sensors. (Alexia Briasouli)
 35. [Multiple Object Tracking Benchmark](#) - A collection of datasets with ground truth, plus a performance league table (ETHZ, U. Adelaide, TU Darmstadt)
 36. [Multispectral visible-NIR video sequences](#) - Annotated multispectral video, visible + NIR (LE2I, Universit de Bourgne)
 37. [NYU Multiple Object Tracking Benchmark](#) (Konrad Schindler et al)
 38. [Occluded Articulated Human Body Dataset](#) - Body pose extraction and tracking under occlusions, 6 RGB-D sequences in total (3500 frames) with one, two and three users, marker-based ground truth data. (Markos Sigalas, Maria Pateraki, Panos Trahanias)
 39. [OxUva](#) - A large-scale long-term tracking dataset composed of 366 long videos of about 14 hours in total, with separate dev (public annotations) and test sets (hidden annotations), featuring target object disappearance and continuous attributes. (Jack Valmadre, Luca Bertinetto, Joao F. Henriques, Ran Tao, Andrea Vedaldi, Arnold Smeulders, Philip Torr, Efstratios Gavves)
 40. [OU-ISIR Gait Database](#) - six video-based gait data sets, two inertial sensor-based gait datasets, and a gait-relevant biometric score data set. (Yasushi Makihara)

41. [PARSE Dataset Additional Data](#) - facial expression, gaze direction, and gender (Antol, Zitnick, Parikh)
42. [PARSE Dataset of Articulated Bodies](#) - 300 images of humans and horses (Ramanan)
43. [PathTrack dataset: a large-scale MOT dataset](#) - PathTrack is a large scale multi-object tracking dataset of more than 15,000 person trajectories in 720 sequences. (Santiago Manen, Michael Gygli, Dengxin Dai, Luc Van Gool)
44. [PDbm: People Detection benchmark repository](#) - realistic sequences, manually annotated people detection ground truth and a complete evaluation framework (Garc??a-Mart??n, Mart??nez, Besc??s)
45. [PDds: A Person Detection dataset](#) - several annotated surveillance sequences of different levels of complexity (Garc??a-Mart??n, Mart??nez, Besc??s)
46. [PETS 2009 Crowd Challenge dataset](#) (Reading University & James Ferryman)
47. [PETS Winter 2009 workshop data](#) (Reading University & James Ferryman)
48. [PETS: 2015 Performance Evaluation of Tracking and Surveillance](#) (Reading University & James Ferryman)
49. [PETS: 2015 Performance Evaluation of Tracking and Surveillance](#) (Reading University & Luis Patino)
50. [PETS 2016 datasets](#) - multi-camera (including thermal cameras) video recordings of human behavior around a stationary vehicle and around a boat (Thomas Cane)
51. [PIROPO](#) - People in Indoor ROoms with Perspective and Omnidirectional cameras, with more than 100,000 annotated frames (GTI-UPM, Spain)
52. [People-Art](#) - a databased containing people labelled in photos and artwork (Qi Wu and Hongping Cai)
53. [Photo-Art-50](#) - a databased containing 50 object classes annotated in photos and artwork (Qi Wu and Hongping Cai)
54. [Pixel-based change detection benchmark dataset](#) (Goyette et al)
55. [Precarious Dataset](#) - unusual people detection dataset (Huang)
56. [RAiD](#) - Re-Identification Across Indoor-Outdoor Dataset: 43 people, 4 cameras, 6920 images (Abir Das et al)
57. [RPIfield](#) - Person re-identification dataset containing 4108 person images with timestamps. (Meng Zheng, Srikrishna Karanam, Richard J. Radke)
58. [Singapore Maritime Dataset](#) - Visible range videos and Infrared videos. (Dilip K. Prasad)
59. [SYNTHIA](#) - Large set (~half million) of virtual-world images for training autonomous cars to see. (ADAS Group at Computer Vision Center)
60. [Shinpuhkan 2014](#) - A Person Re-identification dataset containing 22,000 images of 24 people captured by 16 cameras. (Yasutomo Kawanishi et al.)
61. [Stanford Structured Group Discovery dataset](#) - Discovering Groups of People in Images (W. Choi et al)

62. [TrackingNet](#) - Large-scale dataset for tracking in the wild: more than 30k annotated sequences for training, more than 500 sequestered sequences for testing, evaluation server and leaderboard for fair ranking. (Matthias Muller, Adel Bibi, Silvio Giancola, Salman Al-Subaihi and Bernard Ghanem)
63. [Transient Biometrics Nails Dataset V01](#) (Igor Barros Barbosa)
64. [Temple Color 128 - Color Tracking Benchmark](#) - Encoding Color Information for Visual Tracking (P. Liang, E. Blasch, H. Ling)
65. [TUM Gait from Audio, Image and Depth \(GAID\) database](#) - containing tracked RGB video, tracked depth video, and audio for 305 subjects (Babae, Hofmann, Geiger, Bachmann, Schuller, Rigoll)
66. [TVPR \(Top View Person Re-identification\) dataset](#) - person re-identification using an RGB-D camera in a Top-View configuration: indoor 23 sessions, 100 people, 8 days (Liciotti, Paolanti, Frontoni, Mancini and Zingaretti)
67. [UCLA Aerial Event Dataset](#) - Human activities in aerial videos with annotations of people, objects, social groups, activities and roles (Shu, Xie, Rothrock, Todorovic, and Zhu)
68. [Univ of Central Florida - Crowd Dataset](#) (Saad Ali)
69. [Univ of Central Florida - Crowd Flow Segmentation datasets](#) (Saad Ali)
70. [VIPeR: Viewpoint Invariant Pedestrian Recognition](#) - 632 pedestrian image pairs taken from arbitrary viewpoints under varying illumination conditions. (Gray, Brennan, and Tao)
71. [Visual object tracking challenge datasets](#) - The VOT datasets is a collection of fully annotated visual object tracking datasets used in the single-target short-term visual object tracking challenges.(The VOT committee)
72. [WIDER Attribute Dataset](#) - WIDER Attribute is a large-scale human attribute dataset, with 13789 images belonging to 30 scene categories, and 57524 human bounding boxes each annotated with 14 binary attributes.(Li, Yining and Huang, Chen and Loy, Chen Change and Tang, Xiaoou)
73. [WUds: Wheelchair Users Dataset](#) - wheelchair users detection data, to extend people detection, providing a more general solution to detect people in environments such as independent and assisted living, hospitals, healthcare centers and senior residences (Martínez-Nieto, García-Martínez, Martínez)
74. [YouTube-BoundingBoxes](#) - 5.6 million accurate human-annotated BB from 23 object classes tracked across frames, from 240,000 YouTube videos, with a strong focus on the person class (1.3 million boxes) (Real, Shlens, Pan, Mazzocchi, Vanhoucke, Khan, Kakarla et al)

Remote Sensing

1. [Aerial Imagery for Roof Segmentation \(AIRS\)](#) - 457 km² coverage of orthorectified aerial images with over 220,000 buildings for roof segmentation. (Lei Wang, Qi Chen)

2. [Brazilian Cerrado-Savanna Scenes Dataset](#) - Composition of IR-R-G scenes taken by RapidEye sensor for vegetation classification in Brazilian Cerrado-Savanna. (K. Nogueira, J. A. dos Santos, T. Fornazari, T. S. Freire, L. P. Morellato, R. da S. Torres)
3. [Brazilian Coffee Scenes Dataset](#) - Composition of IR-R-G scenes taken by SPOT sensor for identification of coffee crops in Brazilian mountains.(O. A. B. Penatti, K. Nogueira, J. A. dos Santos.)
4. [Building Detection Benchmark](#) -14 images acquired from IKONOS (1 m) and QuickBird (60 cm)(Ali Ozgun Ok and Caglar Senaras)
5. [CBERS-2B, Landsat 5 TM, Geoeye, Ikonos-2 MS and ALOS-PALSAR](#) - land-cover classification using optical images(D. Osaku et al.)
6. [Data Fusion Contest 2015 \(Zeebruges\)](#) - This dataset provides a RGB aerial dataset (5cm) and a Lidar point cloud (65pts/m²) over the harbor of the city of Zeebruges (Belgium). It also provided a DSM derived from the point cloud and a semantic segmentation ground truth of five of the seven 10000 x 10000 pixels tiles. An evaluation server is used to evaluate the results on the two other tiles. (Image analysis and Data Fusion Technical Committee, IEEE Geoscience, Remote Sensing Society)
7. [Data Fusion Contest 2017](#) - This dataset provides satellite (Landsat, Sentinel 2) and vector GIS layers (e.g. buildings and road footprint) for nine cities worldwide. The task is to predict land use classes useful for climate models at a 100m prediction grid, given data of different resolution and types of features. 5 cities come with labels, 4 others are kept hidden for scoring on an evaluation server. (Image analysis and Data Fusion Technical Committee, IEEE Geoscience, Remote Sensing Society)
8. [deepGlobe challenge](#) - This datasets comprises three challenges, road extraction, buildings detection and semantic segmentation of land cover. A series of satellite images from Digital Globe (RGB, 50 cm resolution) and labels over several countries worldwide are provided. The results were presented at the DeepGlobe workshop at CVPR 2018. (Facebook, Digital Globe)
9. [DeepGlobe Satellite Image Understanding Challenge](#) - Datasets and evaluation platforms for three deep learning tasks on satellite images: road extraction, building detection, and land type classification. (Demir, Ilke and Koperski, Krzysztof and Lindenbaum, David and Pang, Guan and Huang, Jing and Basu, Saikat and Hughes, Forest and Tuia, Devis and Raskar, Ramesh)
10. [FORTH Multispectral Imaging \(MSI\) datasets](#) - 5 datasets for Multispectral Imaging (MSI), annotated with ground truth data (Polykarpos Karamaoynas)
11. [Furnas and Tiete](#) - sediment yield classification(Pisani et al.)
12. [ISPRS 2D semantic labeling](#) - Height models and true ortho-images with a ground sampling distance of 5cm have been prepared over the city of Potsdam/Germany (Franz Rottensteiner, Gunho Sohn, Markus Gerke, Jan D.

Wegner)

13. [ISPRS 3D semantic labeling](#) - nine class airborne laser scanning data (Franz Rottensteiner, Gunho Sohn, Markus Gerke, Jan D. Wegner)
14. [Inria Aerial Image Labeling Dataset](#) - 9000 square kilometers of color aerial imagery over U.S. and Austrian cities. (Emmanuel Maggiori, Yuliya Tarabalka, Guillaume Charpiat, Pierre Alliez.)
15. [Lampert's Spectrogram Analysis](#) - Passive sonar spectrogram images derived from time-series data, these spectrograms are generated from recordings of acoustic energy radiated from propeller and engine machinery in underwater sea recordings. (Thomas Lampert)
16. [Linköping Thermal InfraRed dataset](#) - The LTIR dataset is a thermal infrared dataset for evaluation of Short-Term Single-Object (STSO) tracking (Linköping University)
17. [MASATI: MARitime SATellite Imagery dataset](#) - MASATI is a dataset composed of optical aerial imagery with 6212 samples which were obtained from Microsoft Bing Maps. They were labeled and classified into 7 classes of maritime scenes: land, coast, sea, coast-ship, sea-ship, sea with multi-ship, sea-ship in detail. (University of Alicante)
18. [MUUFL Gulfport Hyperspectral and LiDAR data set](#) - Co-registered aerial hyperspectral and lidar data over the University of Southern Mississippi Gulfport campus containing several sub-pixel targets. (Gader, Zare, Close, Aitken, Tuell)
19. [NWPU-RESISC45](#) - A large-scale benchmark dataset used for remote sensing image scene classification containing 31500 images covered by 45 scene classes. (Gong Cheng, Junwei Han, and Xiaoqiang Lu)
20. [RIT-18](#) - a high-resolution multispectral dataset for semantic segmentation. (Ronald Kemker, Carl Salvaggio, Christopher Kanan)
21. [Semantic Drone Dataset](#) - 20 houses from nadir (bird's eye) view acquired at 5 to 30 meters above ground. 400 public and 200 private high resolution images of 6000x4000px (24Mpx).
22. [UC Merced Land Use Dataset](#) 21 class land use image dataset with 100 images per class, largely urban, 256x256 resolution, 1 foot pixels (Yang and Newsam)
23. [UCF-CrossView Dataset: Cross-View Image Matching for Geo-localization in Urban Environments](#) - A new dataset of street view and bird's eye view images for cross-view image geo-localization. (Center for Research in Computer Vision, University of Central Florida)
24. [Zurich Summer dataset](#) - it is intended for semantic segmentation of very high resolution satellite images of urban scenes, with incomplete ground truth (Michele Volpi and Vitto Ferrari.)
25. [Zurich Urban Micro Aerial Vehicle Dataset](#) - time synchronized aerial high-resolution images of 2 km of Zurich, with associated other data (Majdik, Till, Scaramuzza)

Robotics

1. [Edinburgh Kitchen Utensil Database](#) - 897 raw and binary images of 20 categories of kitchen utensil, a resource for training future domestic assistance robots (D. Fullerton, A. Goel, R. B. Fisher)
2. [Improved 3D Sparse Maps for High-performance Structure from Motion with Low-cost Omnidirectional Robots - Evaluation Dataset](#) - Data set used in research paper doi:10.1109/ICIP.2015.7351744 (Breckon, Toby P., Cavestany, Pedro)
3. [Indoor Place Recognition Dataset for localization of Mobile Robots](#) - The dataset contains 17 different places built from 2 different robots (virtualMe and pioneer) (Raghavender Sahdev, John K. Tsotsos.)
4. [JTL Stereo Tacking Dataset for Person Following Robots](#) - 11 different indoor and outdoor places for the task of robots following people under challenging situations (Chen, Sahdev, Tsotsos)
5. [Meta rooms](#) - RGB-D data comprised of 28 aligned depth camera images collected by having robot go to specific place and do 360 degrees of pan with various tilts. (John Folkesson et al.)
6. [PanoNavi dataset](#) - A panoramic dataset for robot navigation, consisted of 5 videos lasting about 1 hour. (Lingyan Ran)
7. [Robotic 3D Scan Repository](#) - 3D point clouds from robotic experiments of scenes (Osnabruck and Jacobs Universities)
8. [Solving the Robot-World Hand-Eye\(s\) Calibration Problem with Iterative Methods](#) - These datasets were generated for calibrating robot-camera systems. (Amy Tabb)
9. [The Event-Camera Dataset](#) - This presents the world's first collection of datasets with an event-based camera for high-speed robotics (E. Mueggler, H. Rebecq, G. Gallego, T. Delbruck, D. Scaramuzza)
10. [ViDRILO](#) - ViDRILO is a dataset containing 5 sequences of annotated RGB-D images acquired with a mobile robot in two office buildings under challenging lighting conditions.(Miguel Cazorla, J. Martinez-Gomez, M. Cazorla, I. Garcia-Varea and V. Morell.)
11. [Witham Wharf](#) - For RGB-D of eight locations collect by robot every 10 min over ~10 days by the University of Lincoln. (John Folkesson et al.)

Scenes or Places, Scene Segmentation or Classification

1. [Barcelona](#) - 15,150 images, urban views of Barcelona (Tighe and Lazebnik)
2. [Cross-modal Landmark Identification Benchmark](#) - Dandmark-identification benchmark taken under varying weather conditions, which consists of 17 landmark images taken under several weather conditions, e.g., sunny,

- cloudy, snowy, and sunset. (Yonsei University)
3. [CMU Visual Localization Data Set](#) - Dataset collected over the period of a year using the Navlab 11 equipped with IMU, GPS, INS, Lidars and cameras.(Hernan Badino, Daniel Huber and Takeo Kanade)
 4. [COLD \(COsy Localization Database\) - place localization](#) (Ullah, Pronobis, Caputo, Luo, and Jensfelt)
 5. [DAVIS: Video Object Segmentation dataset 2016](#) - A Benchmark Dataset and Evaluation Methodology for Video Object Segmentation (F. Perazzi, J. Pont-Tuset, B. McWilliams, L. Van Gool, M. Gross, and A. Sorkine-Hornung)
 6. [DAVIS: Video Object Segmentation dataset 2017](#) - The 2017 DAVIS Challenge on Video Object Segmentation (J. Pont-Tuset, F. Perazzi, S. Caelles, P. Arbelaez, A. Sorkine-Hornung, and L. Van Gool)
 7. [EDUB-Seg](#)- Egocentric dataset for event segmentation.(Mariella Dimiccoli, Marc Bolaños, Estefania Talavera, Maedeh Aghaei, Stavri G. Nikolov, and Petia Radeva.)
 8. [European Flood 2013](#) - 3,710 images of a flood event in central Europe, annotated with relevance regarding 3 image retrieval tasks (multi-label) and important image regions. (Friedrich Schiller University Jena, Deutsches GeoForschungsZentrum Potsdam)
 9. [Fieldsafe](#) - A multi-modal dataset for obstacle detection in agriculture. (Aarhus University)
 10. [Fifteen Scene Categories](#) - A dataset of fifteen natural scene categories. (Fei-Fei Li and Aude Oliva)
 11. [FIGRIM \(Fine Grained Image Memorability Dataset\)](#) - A subset of images from the SUN database used for human memory experiments, and provided along with memorability scores. (Bylinskii, Isola, Bainbridge, Torralba, Oliva)
 12. [Geometric Context - scene interpretation images](#) (Derek Hoiem)
 13. [HyKo: A Spectral Dataset for Scene Understanding](#) - The HyKo dataset was captured with compact, low-cost, snapshot mosaic (SSM) imaging cameras, which are able to capture a whole spectral cube in one shot recorded from a moving vehicle enabling hyperspectral scene analysis for road scene understanding. (Active Vision Group, University of Koblenz-Landau)
 14. [Indoor Place Recognition Dataset for localization of Mobile Robots](#) - The dataset contains 17 different places built from 2 different robots (virtualMe and pioneer) (Raghavender Sahdev, John K. Tsotsos.)
 15. [Indoor Scene Recognition](#) - 67 Indoor categories, 15620 images (Quattoni and Torralba)
 16. [Intrinsic Images in the Wild \(IIW\)](#) - Intrinsic Images in the Wild, is a large-scale, public dataset for evaluating intrinsic image decompositions of indoor scenes (Sean Bell, Kavita Bala, Noah Snavely)
 17. [LM+SUN](#) - 45,676 images, mainly urban or human related scenes (Tighe and Lazebnik)
 18. [Maritime Imagery in the Visible and Infrared Spectrums](#) - VAIS contains

- simultaneously acquired unregistered thermal and visible images of ships acquired from piers (Zhang, Choi, Daniilidis, Wolf, & Kanan)
19. [MASATI: MARitime SATellite Imagery dataset](#) - MASATI is a dataset composed of optical aerial imagery with 6212 samples which were obtained from Microsoft Bing Maps. They were labeled and classified into 7 classes of maritime scenes: land, coast, sea, coast-ship, sea-ship, sea with multi-ship, sea-ship in detail. (University of Alicante)
 20. [Materials in Context \(MINC\)](#) - The Materials in Context Database (MINC) builds on OpenSurfaces, but includes millions of point annotations of material labels. (Sean Bell, Paul Upchurch, Noah Snavely, Kavita Bala)
 21. [MIT Intrinsic Images](#) - 20 objects (Roger Grosse, Micah K. Johnson, Edward H. Adelson, and William T. Freeman)
 22. [NYU V2 Mixture of Manhattan Frames Dataset](#) - We provide the Mixture of Manhattan Frames (MMF) segmentation and MF rotations on the full NYU depth dataset V2 by Silberman et al. (Straub, Julian and Rosman, Guy and Freifeld, Oren and Leonard, John J. and Fisher III, John W.)
 23. [OpenSurfaces](#) - OpenSurfaces consists of tens of thousands of examples of surfaces segmented from consumer photographs of interiors, and annotated with material parameters, texture information, and contextual information . (Kavita Bala et al.)
 24. [Oxford Audiovisual Segmentation Dataset](#) - Oxford Audiovisual Segmentation Dataset with Oxford Audiovisual Segmentation Dataset including audio recordings of objects being struck (Arnab, Sapienza, Golodetz, Miksik and Torr)
 25. [Thermal Road Dataset](#) - Our thermal-road dataset provides around 6000 thermal-infrared images captured in the road scene with manually annotated ground-truth. (3500: general road, 1500: complicated road, 1000: off-road). (Jae Shin Yoon)
 26. [Places 2 Scene Recognition database](#) -365 scene categories and 8 millions of images (Zhou, Khosla, Lapedriza, Torralba and Oliva)
 27. [Places Scene Recognition database](#) - 205 scene categories and 2.5 millions of images (Zhou, Lapedriza, Xiao, Torralba, and Oliva)
 28. [RGB-NIR Scene Dataset](#) - 477 images in 9 categories captured in RGB and Near-infrared (NIR) (Brown and Susstrunk)
 29. [RMS2017 - Reconstruction Meets Semantics outdoor dataset](#) - 500 semantically annotated images with poses and point cloud from a real garden (Tylecek, Sattler)
 30. [RMS2018 - Reconstruction Meets Semantics virtual dataset](#) - 30k semantically annotated images with poses and point cloud from 6 virtual gardens (An, Tylecek)
 31. [Southampton-York Natural Scenes Dataset](#) 90 scenes, 25 indoor and outdoor scene categories, with spherical LiDAR, HDR intensity, stereo intensity panorama. (Adams, Elder, Graf, Leyland, Lugtigheid, Murry)
 32. [SUN 2012](#) - 16,873 fully annotated scene images for scene categorization

- (Xiao et al)
33. [SUN 397](#) - 397 scene categories for scene classification (Xiao et al)
 34. [SUN RGB-D: A RGB-D Scene Understanding Benchmark Suite](#) - 10,000 RGB-D images, 146,617 2D polygons and 58,657 3D bounding boxes (Song, Lichtenberg, and Xiao)
 35. [SYNTIA](#) - Large set (~half million) of virtual-world images for training autonomous cars to see. (ADAS Group at Computer Vision Center)
 36. [Sift Flow \(also known as LabelMe Outdoor, LMO\)](#) - 2688 images, mainly outdoor natural and urban (Tighe and Lazebnik)
 37. [Stanford Background Dataset](#) - 715 images of outdoor scenes containing at least one foreground object (Gould et al)
 38. [Surface detection](#) - Real-time traversable surface detection by colour space fusion and temporal analysis - Evaluation Dataset (Breckon, Toby P., Katramados, Ioannis)
 39. [Taskonomy](#) - Over 4.5 million real images each with ground truth for 25 semantic, 2D, and 3D tasks. (Zamir, Sax, Shen, Guibas, Malik, Savarese)
 40. [The iNaturalist Species Classification and Detection Dataset](#) - The iNaturalist 2017 species classification and detection dataset has been collected and annotated by citizen scientists and contains 859,000 images from over 5,000 different species of plants and animals. (Caltech)
 41. [ViDRILO](#) - ViDRILO is a dataset containing 5 sequences of annotated RGB-D images acquired with a mobile robot in two office buildings under challenging lighting conditions.(Miguel Cazorla, J. Martinez-Gomez, M. Cazorla, I. Garcia-Varea and V. Morell.)
 42. [Wireframe dataset](#) - A set of RGB images of man-made scenes are annotated with junctions and lines, which describes the large-scale geometry of the scenes.(Huang et al.)

Segmentation (General)

1. [A Dataset for Sky Segmentation](#) - sentence describing it: This Sky dataset was used to evaluate the method IFT-SLIC and other superpixel algorithms, using the superpixel-based sky segmentation method proposed by Juraj Kostolansky. It contains a collection of 60 images based on the Caltech Airplanes Side dataset by R. Fergus with ground truth for sky segmentation. (Eduardo B. Alexandre, Paulo A. V. Miranda, R. Fergus)
2. [Aberystwyth Leaf Evaluation Dataset](#) - Timelapse plant images with hand marked up leaf-level segmentations for some time steps, and biological data from plant sacrifice. (Bell, Jonathan; Dee, Hannah M.)
3. [ADE20K](#) - 22+K hierarchically segmented and labeled scene images (900 scene categories, 3+K classes and subpart classes) (Zhou, Zhao, Puig, Fidler, Barriuso, Torralba)
4. [Alpert et al. Segmentation evaluation database](#) (Sharon Alpert, Meirav

- Galun, Ronen Basri, Achi Brandt)
5. [BMC \(Background Model Challenge\)](#) - A dataset for comparing background subtraction algorithms, composed of real and synthetic videos(Antoine)
 6. [Berkeley Segmentation Dataset and Benchmark](#) (David Martin and Charless Fowlkes)
 7. [CAD 120 affordance dataset](#) - Pixelwise affordance annotation in human context (Sawatzky, Srikantha, Gall)
 8. [COLT](#) - The dataset contains 40 imagenet categories with manually annotated per-pixel object masks. (Jia Li)
 9. [CO-SKEL dataset](#) - This dataset consists of categorized skeleton and segmentation masks for evaluating co-skeletonization methods. (Koteswar Rao Jerripothula, Jianfei Cai, Jiangbo Lu, Junsong Yuan)
 10. [Crack detection on 2D pavement images](#) - five sets of pavement images that contain cracks with the manual ground truth associated and 5 automatic segmentations obtained with existing approaches (Sylvie Chambon)
 11. [CTU Color and Depth Image Dataset of Spread Garments](#) - Images of spread garments with annotated corners.(Wagner, L., Krejov D., and Smutn V. (Czech Technical University in Prague))
 12. [CTU Garment Folding Photo Dataset](#) - Color and depth images from various stages of garment folding.(Sushkov R., Melkumov I., Smutn y V. (Czech Technical University in Prague))
 13. [DeformIt 2.0](#) - Image Data Augmentation Tool: Simulate novel images with ground truth segmentations from a single image-segmentation pair (Brian Booth and Ghassan Hamarneh)
 14. [GrabCut Image database](#) (C. Rother, V. Kolmogorov, A. Blake, M. Brown)
 15. [Histology Image Collection Library \(HICL\)](#) - The HICL is a compilation of 3870 histopathological images (so far) from various diseases, such as brain cancer, breast cancer and HPV (Human Papilloma Virus)-Cervical cancer. (Medical Image and Signal Processing (MEDISP) Lab., Department of Biomedical Engineering, School of Engineering, University of West Attica)
 16. [ICDAR'15 Smartphone document capture and OCR competition - challenge 1](#) - videos of documents filmed by a user with a smartphone to simulate mobile document capture, and ground truth coordinates of the document corners to detect. (Burie, Chazalon, Coustaty, Eskenazi, Luqman, Mehri, Nayef, Ogier, Prum and Rusinol)
 17. [Intrinsic Images in the Wild \(IIW\)](#) - Intrinsic Images in the Wild, is a large-scale, public dataset for evaluating intrinsic image decompositions of indoor scenes (Sean Bell, Kavita Bala, Noah Snavely)
 18. [LabelMe images database and online annotation tool](#) (Bryan Russell, Antonio Torralba, Kevin Murphy, William Freeman)
 19. [LITS Liver Tumor Segmentation](#) - 130 3D CT scans with segmentations of the liver and liver tumor. Public benchmark with leaderboard at Codalab.org (Patrick Christ)
 20. [Materials in Context \(MINC\)](#) - The Materials in Context Database (MINC)

- builds on OpenSurfaces, but includes millions of point annotations of material labels. (Sean Bell, Paul Upchurch, Noah Snavely, Kavita Bala)
21. [Multi-species fruit flower detection](#) - This dataset consists of four sets of flower images, from three different tree species: apple, peach, and pear, and accompanying ground truth images. (Philipe A. Dias, Amy Tabb, Henry Medeiros)
 22. [Objects with thin and elongated parts](#) - The three datasets used to evaluate our method Oriented Image Foresting Transform with Connectivity Constraints, which contain objects with thin and elongated parts. These databases are composed of 280 public images of birds and insects with ground truths. (Lucy A. C. Mansilla (IME-USP), Paulo A. V. Miranda)
 23. [OpenSurfaces](#) - OpenSurfaces consists of tens of thousands of examples of surfaces segmented from consumer photographs of interiors, and annotated with material parameters, texture information, and contextual information . (Kavita Bala et al.)
 24. [Osnabrück gaze tracking data](#) - 318 video sequences from several different gaze tracking data sets with polygon based object annotation. (Schöning, Faion, Heidemann, Krumnack, Gert, Açık, Kietzmann, Heidemann & König)
 25. [PASCAL-Scribble Dataset](#) - Our PASCAL-Scribble Dataset provides scribble-annotations on 59 object/stuff categories. (Di Lin)
 26. [PetroSurf3D](#) - 26 high resolution (sub-millimeter accuracy) 3D scans of rock art with pixelwise labeling of petroglyphs for segmentation. (Poier, Seidl, Zeppelzauer, Reinbacher, Schaich, Bellandi, Marretta, Bischof)
 27. [Shadow Detection/Texture Segmentation Computer Vision Dataset](#) - Video based sequences for shadow detection/suppression, with ground truth (Newey, C., Jones, O., & Dee, H. M.)
 28. [SYNTHIA](#) - Large set (~half million) of virtual-world images for training autonomous cars to see. (ADAS Group at Computer Vision Center)
 29. [Stony Brook University Shadow Dataset \(SBU-Shadow5k\)](#) - Large scale shadow detection dataset from a wide variety of scenes and photo types, with human annotations (Tomas F.Y. Vicente, Le Hou, Chen-Ping Yu, Minh Hoai, Dimitris Samaras)
 30. [TRoM: Tsinghua Road Markings](#) - This is a dataset which contributes to the area of road marking segmentation for Automated Driving and ADAS. (Xiaolong Liu, Zhidong Deng, Lele Cao, Hongchao Lu)
 31. [VOS](#) - A dataset with 200 Internet videos for video-based salient object detection and segmentation. (Jia Li, Changqun Xia)
 32. [XPIE](#) - An image dataset with 10000 images containing manually annotated salient objects and 8596 containing no salient objects. (Jia Li, Changqun Xia)

Simultaneous Localization and Mapping

1. [Collaborative SLAM Dataset \(CSD\)](#) - The dataset consists of four different subsets - Flat, House, Priory and Lab - each containing several RGB-D sequences that can be reconstructed and successfully relocalised against each other to form a combined 3D model. Each sequence was captured using an Asus ZenFone AR, and we provide an accurate local 6D pose for each RGB-D frame in the dataset. We also provide the calibration parameters for the depth and colour sensors, optimised global poses for the sequences in each subset, and a pre-built mesh of each sequence. (Golodetz, Cavallari, Lord, Prisacariu, Murray, Torr)
2. [Event-Camera Data for Pose Estimation, Visual Odometry, and SLAM](#) - The data also include intensity images, inertial measurements, and ground truth from a motion-capture system. (ETH)
3. [House3D](#) - House3D is a virtual 3D environment which consists of thousands of indoor scenes equipped with a diverse set of scene types, layouts and objects sourced from the SUNCG dataset. It consists of over 45k indoor 3D scenes, ranging from studios to two-storied houses with swimming pools and fitness rooms. All 3D objects are fully annotated with category labels. Agents in the environment have access to observations of multiple modalities, including RGB images, depth, segmentation masks and top-down 2D map views. The renderer runs at thousands frames per second, making it suitable for large-scale RL training. (Yi Wu, Yuxin Wu, Georgia Gkioxari, Yuandong Tian, facebook research)
4. [Indoor Dataset of Quadrotor with Down-Looking Camera](#) - This dataset contains the recording of the raw images, IMU measurements as well as the ground truth poses of a quadrotor flying a circular trajectory in an office size environment. (Scaramuzza, ETH Zurich, University of Zurich)
5. [InLoc](#) - Benchmark for evaluating the accuracy of 6DoF visual localization algorithms in challenging indoor scenarios. (Hajime Taira, Masatoshi Okutomi, Torsten Sattler, Mircea Cimpoi, Marc Pollefeys, Josef Sivic, Tomas Pajdla, Akihiko Torii)
6. [Long-term visual localization](#) - TBenchmark for evaluating visual localization and mapping algorithms under various illumination and seasonal condition. (Torsten Sattler, Will Maddern, Carl Toft, Akihiko Torii, Lars Hammarstrand, Erik Stenborg, Daniel Safari, Masatoshi Okutomi, Marc Pollefeys, Josef Sivic, Fredrik Kahl, Tomas Pajdla)
7. [PanoNavi dataset](#) - A panoramic dataset for robot navigation, consisted of 5 videos lasting about 1 hour. (Lingyan Ran)
8. [RAWSEEDS SLAM benchmark datasets](#) (Rawseeds Project)
9. [Rijksmuseum Challenge 2014](#) - It consist of 100K art objects from the rijksmuseum and comes with an extensive xml files describing each object. (Thomas Mensink and Jan van Gemert)
10. [RSM dataset of Visual Paths](#) - Visual dataset of indoor spaces to benchmark localisation/navigation methods. It consists of 1.5 km of corridors and indoor spaces with ground truth for every frame, measured as distance in

- centimetres from starting point. Includes a synthetically generated corridor for benchmark. (Jose Rivera-Rubio, Ioannis Alexiou, Anil A. Bharath)
11. [The Multi Vehicle Stereo Event Camera Dataset](#) - Multiple sequences containing a stereo pair of DAVIS 346b event cameras with ground truth poses, depth maps and optical flow. (lex Zihao Zhu, Dinesh Thakur, Tolga Ozaslan, Bernd Pfrommer, Vijay Kumar, Kostas Daniilidis)
 12. [TUM RGB-D Benchmark](#) - Dataset and benchmark for the evaluation of RGB-D visual odometry and SLAM algorithms (BCRgen Sturm, Nikolas Engelhard, Felix Endres, Wolfram Burgard and Daniel Cremers)
 13. [TUM VI Benchmark](#) - 28 sequences, indoor and outdoor, sensor data from stereo camera and IMU, accurate ground truth at beginning and end segments. (David Schubert, Thore Goll, Nikolaus Demmel, Vladyslav Usenko, Joerg Stueckler, Daniel Cremers)
 14. [Visual Odometry / SLAM Evaluation](#) - The odometry benchmark consists of 22 stereo sequences (Andreas Geiger and Philip Lenz and Raquel Urtasun)
 15. [Visual Odometry Dataset with Plenoptic and Stereo Data](#) - The dataset contains 11 sequences recorded by a hand-held platform consisting of a plenoptic camera and a pair of stereo cameras. The sequences comprising different indoor and outdoor sequences with trajectory length ranging from 25 meters up to several hundred meters. The recorded sequences show moving objects as well as changing lighting conditions. (Niclas Zeller and Franz Quint, Hochschule Karlsruhe, Karlsruhe University of Applied Sciences)

Surveillance and Tracking

1. [A collection of challenging motion segmentation benchmark datasets](#) - These datasets enclose real-life long and short sequences, with increased number of motions and frames per sequence, and also real distortions with missing data. The ground truth is provided on all the frames of all the sequences. (Muhammad Habib Mahmood, Yago Diez, Joaquim Salvi, Xavier Llado)
2. [ATOMIC GROUP ACTIONS dataset](#) - (Ricky J. Sethi et al.)
3. [AVSS07: Advanced Video and Signal based Surveillance 2007 datasets](#) (Andrea Cavallaro)
4. [Activity modeling and abnormality detection dataset](#) - The dataset contains a 45 minutes video with annotated anomalies. (Jagan Varadarajan and Jean-Marc Odobez)
5. [Background subtraction](#) - a list of datasets about background subtraction (Thierry BOUWMANS)
6. [CAMO-UOW Dataset](#) - 10 high resolution videos captured in real scenes for camouflaged background subtraction (Shuai Li and Wanqing Li)
7. [CCTV-Fights](#) - 1,000 videos picturing real-world fights, recorded from

- CCTVs or mobile cameras, and temporally annotated at the frame level.
(Mauricio Perez, ROSE Lab, NTU)
8. [CMUSRD: Surveillance Research Dataset](#) - multi-camera video for indoor surveillance scenario (K. Hattori, H. Hattori, et al)
 9. [DukeMTMC: Duke Multi-Target Multi-Camera tracking dataset](#) - 8 cameras, 85 min, 2m frames, 2000 people of video (Ergys Ristani, Francesco Solera, Roger S. Zou, Rita Cucchiara, Carlo Tomasi)
 10. [DukeMTMC-reID](#) - A subset of the DukeMTMC for image-based person re-identification (8 cameras, 16,522 training images of 702 identities, 2,228 query images of the other 702 identities and 17,661 gallery images.) (Zheng, Zheng, and Yang)
 11. [ETISEO Video Surveillance Download Datasets](#) (INRIA Orion Team and others)
 12. [FMO dataset](#) - FMO dataset contains annotated video sequences with Fast Moving Objects - objects which move over a projected distance larger than their size in one frame. (Denys Rozumnyi, Jan Kotera, Lukas Novotny, Ales Hrabalik, Filip Sroubek, Jiri Matas)
 13. [HDA+ Multi-camera Surveillance Dataset](#) - video from a network of 18 heterogeneous cameras (different resolutions and frame rates) distributed over 3 floors of a research institute with 13 fully labeled sequences, 85 persons, and 64028 bounding boxes of persons. (D. Figueira, M. Taiana, A. Nambiar, J. Nascimento and A. Bernardino)
 14. [Human click data](#) - 20K human clicks on a tracking target (including click errors) (Zhu and Porikli)
 15. [Immediacy Dataset](#) - This dataset is designed for estimation personal relationships. (Xiao Chu et al.)
 16. [MAHNOB Databases](#) -including Laughter Database, HCI-tagging Database, MHI-Mimicry Database(M. Pantic. etc.)
 17. [Moving Infants In RGB-D \(MINI-RGBD\)](#) - A synthetic, realistic RGB-D data set for infant pose estimation containing 12 sequences of moving infants with ground truth joint positions. (N. Hesse, C. Bodensteiner, M. Arens, U. G. Hofmann, R. Weinberger, A. S. Schroeder)
 18. [MSMT17](#) - Person re-identification dataset. 180 hours of videos, 12 outdoor cameras, 3 indoor cameras, and 12 time slots. (Wei Longhui, Zhang Shiliang, Gao Wen, Tian Qi)
 19. [MVHAUS-PI](#) - a multi-view human interaction recognition dataset (Saeid et al.)
 20. [Multispectral visible-NIR video sequences](#) - Annotated multispectral video, visible + NIR (LE2I, Universit de Bourgogne)
 21. [Openvisor - Video surveillance Online Repository](#) (Univ of Modena and Reggio Emilia)
 22. [Parking-Lot dataset](#) - Parking-Lot dataset is a car dataset which focus on moderate and heavily occlusions on cars in the parking lot scenario.(B. Li, T.F. Wu and S.C. Zhu)

23. [Pornography Database](#) - The Pornography database is a pornography detection dataset containing nearly 80 hours of 400 pornographic and 400 non-pornographic videos extracted from pornography websites and Youtube. (Avila, Thome, Cord, Valle, de Araujo)
24. [Princeton Tracking Benchmark](#) - 100 RGBD tracking datasets (Song and Xiao)
25. [QMUL Junction Dataset 1 and 2](#) - Videos of busy road junctions. Supports anomaly detection tasks. (T. Hospedales Edinburgh/QMUL)
26. [Queen Mary Multi-Camera Distributed Traffic Scenes Dataset \(QMDTS\)](#) - The QMDTS is collected from urban surveillance environment for the study of surveillance behaviours in distributed scenes.(Dr. Xun Xu. Prof. Shaogang Gong and Dr. Timothy Hospedales)
27. [Road Anomaly Detection](#) - 22km, 11 vehicles, normal + 4 defect categories (Hameed, Mazhar, Hassan)
28. [SALSA: Synergetic sociAL Scene Analysis](#) - A Novel Dataset for Multimodal Group Behavior Analysis(Xavier Alameda-Pineda etc.)
29. [SBMnet \(Scene Background Modeling.NET\)](#) - A dataset for testing background estimation algorithms(Jodoin, Maddalena, and Petrosino)
30. [SBM-RGBD dataset](#) - 35 Kinect indoor RGBD videos to evaluate and compare scene background modelling methods for moving object detection (Camplani, Maddalena, Moy?? Alcover, Petrosino, Salgado)
31. [SCOUTER](#) - video surveillance ground truthing (shifting perspectives, different setups/lighting conditions, large variations of subject). 30 videos and approximately 36,000 manually labeled frames. (Catalin Mitrea)
32. [SJTU-BEST](#)One surveillance-specified datasets platform with realistic, on-using camera-captured, diverse set of surveillance images and videos (Shanghai Jiao Tong University)
33. [SPEVI: Surveillance Performance Evaluation Initiative](#) (Queen Mary University London)
34. [Shinpuhkan 2014](#) - A Person Re-identification dataset containing 22,000 images of 24 people captured by 16 cameras. (Yasutomo Kawanishi et al.)
35. [Stanford Drone Dataset](#) - 60 images and videos of various types of agents (not just pedestrians, but also bicyclists, skateboarders, cars, buses, and golf carts) that navigate in a real world outdoor environment such as a university campus (Robicquet, Sadeghian, Alahi, Savarese)
36. [The S-Hock dataset](#) - A new Benchmark for Spectator Crowd Analysis. (Francesco Setti, Davide Conigliaro, Paolo Rota, Chiara Bassetti, Nicola Conci, Nicu Sebe, Marco Cristani)
37. [Tracking in extremely cluttered scenes](#) - this single object tracking dataset has 28 highly cluttered sequences with per frame annotation(Jingjing Xiao,Linbo Qiao,Rustam Stolkin,Ale Leonardis)
38. [TrackingNet](#) - Large-scale dataset for tracking in the wild: more than 30k annotated sequences for training, more than 500 sequestered sequences for testing, evaluation server and leaderboard for fair ranking. (Matthias

- Muller, Adel Bibi, Silvio Giancola, Salman Al-Subaihi and Bernard Ghanem)
39. [UCF-Crime Dataset: Real-world Anomaly Detection in Surveillance Videos](#) - A large-scale dataset for real-world anomaly detection in surveillance videos. It consists of 1900 long and untrimmed real-world surveillance videos (of 128 hours), with 13 realistic anomalies such as fighting, road accident, burglary, robbery, etc. as well as normal activities. (Center for Research in Computer Vision, University of Central Florida)
 40. [UCLA Aerial Event Dataset](#) - Human activities in aerial videos with annotations of people, objects, social groups, activities and roles (Shu, Xie, Rothrock, Todorovic, and Zhu)
 41. [UCSD Anomaly Detection Dataset](#) - a stationary camera mounted at an elevation, overlooking pedestrian walkways, with unusual pedestrian or non-pedestrian motion.
 42. [UCSD trajectory clustering and analysis datasets](#) - (Morris and Trivedi)
 43. [USC Information Sciences Institute's ATOMIC PAIR ACTIONS dataset](#) - (Ricky J. Sethi et al.)
 44. [Udine Trajectory-based anomalous event detection dataset](#) - synthetic trajectory datasets with outliers (Univ of Udine Artificial Vision and Real Time Systems Laboratory)
 45. [Visual Tracker Benchmark](#) - 100 object tracking sequences with ground truth with [Visual Tracker Benchmark evaluation](#), including tracking results from a number of trackers (Wu, Lim, Yang)
 46. [WIDER Attribute Dataset](#) - WIDER Attribute is a large-scale human attribute dataset, with 13789 images belonging to 30 scene categories, and 57524 human bounding boxes each annotated with 14 binary attributes. (Li, Yining and Huang, Chen and Loy, Chen Change and Tang, Xiaoou)

Textures

1. [Brodatz Texture, Normalized Brodatz Texture, Colored Brodatz Texture, Multiband Brodatz Texture](#) 154 new images plus 112 original images with various transformations (A. Safia, D. He)
2. [Color texture images by category](#) (textures.forrest.cz)
3. [Columbia-Utrecht Reflectance and Texture Database](#) (Columbia & Utrecht Universities)
4. [DynTex: Dynamic texture database](#) (Renaud Piteri, Mark Huiskes and Sandor Fazekas)
5. [Houses dataset](#) - Benchmark dataset for houses prices that contains both visual and textual information about 535 houses. (Ahmed, Eman and Moustafa, Mohamed)
6. [Intrinsic Images in the Wild \(IIW\)](#) - Intrinsic Images in the Wild, is a large-scale, public dataset for evaluating intrinsic image decompositions of indoor scenes (Sean Bell, Kavita Bala, Noah Snavely)

7. [KTH TIPS & TIPS2 textures](#) - pose/lighting/scale variations (Eric Hayman)
8. [Materials in Context \(MINC\)](#) - The Materials in Context Database (MINC) builds on OpenSurfaces, but includes millions of point annotations of material labels. (Sean Bell, Paul Upchurch, Noah Snavely, Kavita Bala)
9. [OpenSurfaces](#) - OpenSurfaces consists of tens of thousands of examples of surfaces segmented from consumer photographs of interiors, and annotated with material parameters, texture information, and contextual information . (Kavita Bala et al.)
10. [Oulu Texture Database](#) (Oulu University)
11. [Oxford Describable Textures Dataset](#) - 5640 images in 47 categories (M.Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, A. Vedaldi)
12. [Prague Texture Segmentation Data Generator and Benchmark](#) (Mikes, Haindl)
13. [Salzburg Texture Image Database \(STex\)](#) - a large collection of 476 color texture image that have been captured around Salzburg, Austria. (Roland Kwitt and Peter Meerwald)
14. [Synthetic SVBRDFs and renderings](#) - The dataset contains 200000 renderings of 20000 different materials associated with their ground truth representation in the Cook-Torrance model. Distributed under research only, non commercial use license. ("GraphDeco" team, Inria)
15. [Texture Database](#)The texture database features 25 texture classes, 40 samples each(Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce)
16. [Uppsala texture dataset of surfaces and materials](#) - fabrics, grains, etc.
17. [Vision Texture](#) (MIT Media Lab)

Urban Datasets

1. [Barcelona](#) - 15,150 images, urban views of Barcelona (Tighe and Lazebnik)
2. [Cityscapes](#) - a large-scale dataset that contains a diverse set of stereo video sequences recorded in street scenes from 50 different cities, with high quality pixel-level annotations of 5.000 frames in addition to a larger set of 20.000 weakly annotated frames. (Cityscapes Team)
3. [CMP Facade Database](#) - Includes 606 rectified images of facades from various places with 12 architectural classes annotated.(Radim Tylecek)
4. [DeepGlobe Satellite Image Understanding Challenge](#) - Datasets and evaluation platforms for three deep learning tasks on satellite images: road extraction, building detection, and land type classification. (Demir, Ilke and Koperski, Krzysztof and Lindenbaum, David and Pang, Guan and Huang, Jing and Basu, Saikat and Hughes, Forest and Tuia, Devis and Raskar, Ramesh)
5. [DroNet: Learning to Fly by Driving](#) - Videos from a bicycle with labeled collision data used for learning to predict potentially dangerous situations for vehicles. (Loquercio, Maqueda, Del Blanco, Scaramuzza)

6. [European Flood 2013](#) - 3,710 images of a flood event in central Europe, annotated with relevance regarding 3 image retrieval tasks (multi-label) and important image regions. (Friedrich Schiller University Jena, Deutsches GeoForschungsZentrum Potsdam)
7. [Houses dataset](#) - Benchmark dataset for houses prices that contains both visual and textual information about 535 houses. (Ahmed, Eman and Moustafa, Mohamed)
8. [LM+SUN](#) - 45,676 images, mainly urban or human related scenes (Tighe and Lazebnik)
9. [MIT CBCL StreetScenes Challenge Framework](#): (Stan Bileschi)
10. [Queen Mary Multi-Camera Distributed Traffic Scenes Dataset \(QMDTS\)](#) - The QMDTS is collected from urban surveillance environment for the study of surveillance behaviours in distributed scenes.(Dr. Xun Xu. Prof. Shaogang Gong and Dr. Timothy Hospedales)
11. [Robust Global Translations with 1DSfM](#)the numerical data describing global structure from motion problems for each dataset (Kyle Wilson and Noah Snavely)
12. [Sift Flow \(also known as LabelMe Outdoor, LMO\)](#) - 2688 images, mainly outdoor natural and urban (Tighe and Lazebnik)
13. [Street-View Change Detection with Deconvolutional Networks](#) - Database with aligned image pairs from street-view imagery with structural, lighting, weather and seasonal changes.(Pablo F. Alcantarilla, Simon Stent, German Ros, Roberto Arroyo and Riccardo Gherardi)
14. [SydneyHouse](#) - Streetview house images with accurate 3D house shape, facade object label, dense point correspondence, and annotation toolbox.(Hang Chu, Shenlong Wang, Raquel Urtasun, Sanja Fidler)
15. [Traffic Signs Dataset](#) - recording sequences from over 350 km of Swedish highways and city roads (Fredrik Larsson)
16. [nuTonomy scenes dataset \(nuScenes\)](#) - The nuScenes dataset is a large-scale autonomous driving dataset. It features: Full sensor suite (1x LIDAR, 5x RADAR, 6x camera, IMU, GPS), 1000 scenes of 20s each, 1,440,000 camera images, 400,000 lidar sweeps, two diverse cities: Boston and Singapore, left versus right hand traffic, detailed map information, manual annotations for 25 object classes, 1.1M 3D bounding boxes annotated at 2Hz, attributes such as visibility, activity and pose. (Caesar et al)

Vision and Natural Language

1. [INRIA BL-database](#) - an audio-visual speech corpus multimodal automatic speech recognition, audio/visual synchronization or speech-driven lip animation systems (Benezeth, Bachman, Lejan, Souviraa-Labastie, Bimbot)
2. [CrisisMMD: Multimodal Twitter Datasets from Natural Disasters](#) - The CrisisMMD multimodal Twitter dataset consists of several thousands of

manually annotated tweets and images collected during seven major natural disasters including earthquakes, hurricanes, wildfires, and floods that happened in the year 2017 across different parts of the World. (Firoj Alam, Ferda Ofli, Muhammad Imran)

3. [DAQUAR](#) - A dataset of human question answer pairs about images, which manifests our vision on a Visual Turing Test. (Mateusz Malinowski, Mario Fritz)
4. [Dataset of Structured Queries and Spatial Relations](#) - Dataset of structured queries about images with the emphasise on spatial relations.(Mateusz Malinowski, Mario Fritz)
5. [DVOA: Understanding Data Visualization through Question Answering](#) - a dataset for VQA about bar charts: 3 types of questions, 300,000 Images, 3,487,194 question-answer pairs, detailed metadata (Kafle, Cohen, Price, Kanan)
6. [FigureQA](#) - a dataset for VQA about bar and pie charts, and numerical graphs: 100,000 images, 1,327,368 question-answer pairs, 100 colors and figure plot element names, 15 question types (Kahou, Michalski, Atkinson, Kadar, Trischler, Bengio)
7. [Hannah and her sisters database](#) - a dense audio-visual person-oriented ground-truth annotation of faces, speech segments, shot boundaries (Patrick Perez, Technicolor)
8. [Large scale Movie Description Challenge \(LSMDC\)](#) - A large scale dataset and challenge for movie description, including over 128K video-sentence pairs, mainly sourced from Audio Description (also known as DVS). (Rohrbach, Torabi, Rohrbach, Tandon, Pal, Larochelle, Courville and Schiele)
9. [MPII dataset](#) - A dataset about correcting inaccurate sentences based on the videos. (Amir Mazaheri)
10. [MPI Movie Description dataset - text and video](#) - A dataset of movie clips associated with natural language descriptions sourced from movie scripts and Audio Description. (Rohrbach, Rohrbach, Tandon and Schiele)
11. [Recipe1M - A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images](#) - Recipe1M is a new large-scale, structured corpus of over one million cooking recipes and 13 million food images. As the largest publicly available collection of recipe data, Recipe1M affords the ability to train high-capacity models on aligned, multi-modal data. (Javier Marin, Aritro Biswas, Ferda Ofli, Nicholas Hynes, Amaia Salvador, Yusuf Aytar, Ingmar Weber, Antonio Torralba)
12. [SemArt dataset](#) - A dataset for semantic art understanding, including 21,384 fine-art painting images with attributes and artistic comments. (Noa Garcia, George Vogiatzis)
13. [TACoS Multi-Level Corpus](#) - Dataset of cooking videos associated with natural language descriptions at three levels of detail (long, short and single sentence). (Rohrbach, Rohrbach, Qiu, Friedrich, Pinkal and Schiele)

14. [TallyQA](#) - The largest dataset for open-ended counting as of 2018, and it includes test sets that evaluate both simple and more advanced capabilities. (Manoj Acharya, Kushal Kafle, Christopher Kanan)
15. [TDIUC \(Task-driven image understanding\)](#) - As of 2018, this is the largest VQA dataset and it facilitates analysis for 12 kinds of questions. (Kushal Kafle, Christopher Kanan)
16. [TGIF](#) - 100K animated GIFs from Tumblr and 120K natural language descriptions. (Li, Song, Cao, Tetreault, Goldberg, Jaimes, Luo)
17. [Toronto COCO-QA Dataset](#) - Automatically generated from image captions. 123287 images 78736 train questions 38948 test questions 4 types of questions: object, number, color, location Answers are all one-word. (Mengye Ren, Ryan Kiros, Richard Zemel)
18. [Totally Looks Like](#) - A benchmark for assessment of predicting human-based image similarity (Amir Rosenfeld, Markus D. Solbach, John Tsotsos)
19. [Twitter for Sentiment Analysis \(T4SA\)](#) - About 1 million tweets (text and associated images) labelled according to the sentiment polarity of the text; the data can be used for sentiment analysis as well as other analysis in the wild since the tweets were randomly sampled tweets from the stream of all globally produced tweets. (Lucia Vadicamo, Fabio Carrara, Andrea Cimino, Stefano Cresci, Felice Dell'Orletta, Fabrizio Falchi, Maurizio Tesconi)
20. [UCF-CrossView Dataset: Cross-View Image Matching for Geo-localization in Urban Environments](#) - A new dataset of street view and bird's eye view images for cross-view image geo-localization. (Center for Research in Computer Vision, University of Central Florida)
21. [Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations](#) - Visual Genome is a dataset, a knowledge base, an ongoing effort to connect structured image concepts to language. (Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li Jia-Li, David Ayman Shamma, Michael Bernstein, Li Fei-Fei)
22. [Visual Relationship Detection with Language Priors](#) - 5000 images, 37,993 thousand relationships, 100 object categories, 70 predicate categories (Lu, Krishna, Bernstein, Fei-Fei)
23. [VQA: Visual Question Answering](#) - a new dataset containing open-ended questions about images. These questions require an understanding of vision, language and commonsense knowledge to answer. (Yash Goyal, Tejas Khot, Georgia Institute of Technology, Army Research Laboratory, Virginia Tech)
24. [VQA v1 - VQA: Visual Question Answering](#) - For every image, we collected 3 free-form natural-language questions with 10 concise open-ended answers each. We provide two formats of the VQA task. (Stanislaw Antol, Aishwarya Agrawal, Jiasen Lu)
25. [YouCook2](#) - 2000 long YouTube cooking videos, where each recipe step is temporally localized and described by an imperative English sentence.

Bounding box annotations are available for the validation & test splits.
(Luwei Zhou, Chenliang Xu, and Jason Corso)

26. [YouTube Movie Summaries](#) - movie summary videos from YouTube, annotated with the correspondence between the video segments and the movie synopsis text at the sentence level and the phrase level. (Pelin Dogan, Boyang Li, Leonid Sigal, Markus Gross)

Other Collections

1. [4D Light Field Dataset](#) - 24 synthetic scenes with 9x9x512x512x3 input images, depth and disparity ground truth, camera parameters, and evaluation masks. (Katrin Honauer, Ole Johannsen, Daniel Kondermann, Bastian Goldluecke)
2. [AMADI LontarSet](#) - Balinese Palm Leaf Manuscript Images Dataset for Binarization, Query-by-Example Word Spotting, and Isolated Character Recognition of Balinese Script. (The AMADI Project et al.)
3. [Annotated Web Ears Dataset \(AWE Dataset\)](#) - All images were acquired by cropping ears from images from the internet of known persons. (Ziga Emersic, Vitomir Struc and Peter Peer)
4. [Biometrics Evaluation and Testing](#) - Evaluation of identification technologies, including Biometrics (European computing e-infrastructure)
5. [CALVIN research group datasets](#) - object detection with eye tracking, imagenet bounding boxes, synchronised activities, stickman and body poses, youtube objects, faces, horses, toys, visual attributes, shape classes (CALVIN group)
6. [CANTATA Video and Image Database Index site](#) (Multitel)
7. [Chinese University of Hong Kong datasets](#) - Face sketch, face alignment, image search, public square observation, occlusion, central station, MIT single and multiple camera trajectories, person re-identification (Multimedia lab)
8. [Computer Vision Homepage list of test image databases](#) (Carnegie Mellon Univ)
9. [Computer Vision Lab OCR DataBase \(CVL OCR DB\)](#) - CVL OCR DB is a public annotated image dataset of 120 binary annotated images of text in natural scenes. (Andrej Ikica and Peter Peer.)
10. [ETHZ various datasets](#) - including ETH 3D head pose, BIWI audiovisual data, ETHZ shape classes, BIWI walking pedestrians, pedestrians, buildings, 4D MRI, personal events, liver ultrasound, Food 101. (ETH Zurich, Computer Vision Lab)
11. [Finger Vein USM \(FV-USM\) Database](#) - An infrared finger image database consists of finger vein and also finger geometry information. (Bakhtiar Affendi Rosdi, Universiti Sains Malaysia)
12. [General 100 Dataset](#) - General-100 dataset contains 100 bmp-format

- images (with no compression), which are well-suited for super-resolution training(Dong, Chao and Loy, Chen Change and Tang, Xiaoou)
13. [GPDS Bengali and Devanagari Synthetic Signature Databases](#) - Dual Off line and On line signature databases of Bengali and Devanagari signatures. (Miguel A. Ferrer, GPDS, ULPGC)
 14. [GPDS Synthetic OnLine and OffLine Signature database](#) - Dual Off line and On line Latin signature database. (Miguel A. Ferrer, GPDS, ULPGC)
 15. [HKU-IS](#) - 4447 images with pixel labeling groundtruth for salient object detection.(Guanbin Li, Yizhou Yu)
 16. [High-res 3D-Models](#) - it includes high-res renderings of these data-sets. (Hubert etc.)
 17. [I3 - Yahoo Flickr Creative Commons 100M](#) - This dataset contains a list of photos and videos. (B. Thomee, D.A. Shamma, G. Friedland et al.)
 18. [The Int. Assoc. for Pattern Recognition's Technical Committee TC11 on Reading Systems index of datasets concerning document text reading](#)
 19. [IDIAP dataset collection](#) - 26 different datasets - multimodal, attack, biometric, cursive characters, discourse, eye gaze, posters, maya codex, MOBIO, face spoofing, game playing, finger vein, youtube-personality traits (IDIAP team)
 20. [Kinect v2 dataset](#) - Dataset for evaluating unwrapping in kinect2 depth decoding (Felix etc.)
 21. [Laval HDR Sky Database](#) - The database contains 800 hemispherical, full HDR photos of the sky that can be used for outdoor lighting analysis. (Jean-Francois Lalonde et al.)
 22. [Leibe's Collection of people/vehicle/object databases](#) (Bastian Leibe)
 23. [Lotus Hill Image Database Collection with Ground Truth](#) (Sealeen Ren, Benjamin Yao, Michael Yang)
 24. [MIT Saliency Benchmark dataset](#) - collection (pointers to 23 datasets) (Bylinskii, Judd, Borji, Itti, Durand, Oliva, Torralba}
 25. [Michael Firman's List of RGBD datasets](#)
 26. [Msspoof:2D multi-spectral face spoofing](#) - Presentation attack (spoofing) dataset with samples from both real data subjects and spoofed data subjects performed with paper to a NIR and VIS camera(Idiap research institute)
 27. [Multiview Stereo Evaluation](#) - Each dataset is registered with a "ground-truth" 3D model acquired via a laser scanning process(Steve Seitz et al)
 28. [Oxford Misc, including Buffy, Flowers, TV characters, Buildings, etc](#) (Oxford Visual geometry Group)
 29. [PEIPA Image Database Summary](#) (Pilot European Image Processing Archive)
 30. [PalmVein spoofing](#) - Presentation attack (spoofing) dataset with samples from spoofed data subjects (corresponding to VERA Palmvein) performed with paper(Idiap research institute)
 31. [RSBA dataset](#) - Sequences for evaluating rolling shutter bundle adjustment

- (Per-Erik etc.)
32. [Replay Attack:2D face spoofing](#) - Presentation attack (spoofing) dataset with samples from both real data subjects and spoofed data subjects performed with paper, photos and videos from a mobile device to a laptop. (Idiap research institute)
 33. [Replay Mobile:2D face spoofing](#) - Presentation attack (spoofing) dataset with samples from both real data subjects and spoofed data subjects performed with paper, photos and videos to/from a mobile device.(Idiap research institute)
 34. [Synthetic Sequence Generator](#) - Synthetic Sequence Generator (G. Hamarneh)
 35. [The Event-Camera Dataset](#) - This presents the world's first collection of datasets with an event-based camera for high-speed robotics (E. Mueggler, H. Rebecq, G. Gallego, T. Delbruck, D. Scaramuzza)
 36. [The world from a cat perspective](#) - videos recorded from the head of a freely behaving cat (Belinda Y. Betsch, Wolfgang Einh?user)
 37. [USC Annotated Computer Vision Bibliography database publication summary](#) (Keith Price)
 38. [USC-SIPI image databases: texture, aerial, favorites \(eg. Lena\)](#) (USC Signal and Image Processing Institute)
 39. [Univ of Bern databases on handwriting, online documents, string edit and graph matching](#) (Univ of Bern, Computer Vision and Artificial Intelligence)
 40. [VERA Fingervein spoofing](#) - Presentation attack (spoofing) dataset with samples from spoofed data subjects (corresponding to VERA Fingervein) performed with paper(Idiap research institute)
 41. [VERA Fingervein](#) - Fingervein dataset with data subjects recorded with a open fingervein sensor(Idiap research institute)
 42. [VERA PalmVein:PalmVein](#) - Palmvein dataset with data subjects recorded with a open palmvein sensor(Idiap research institute)
 43. [Vehicle Detection in Aerial Imagery](#) - VEDAI is a dataset for Vehicle Detection in Aerial Imagery, provided as a tool to benchmark automatic target recognition algorithms in unconstrained environments. (Sebastien Razakarivony and Frederic Jurie)
 44. [Video Stacking Dataset](#) - Dataset for evaluating video stacking on cell-phones (Erik Ringaby etc.)
 45. [Wrist-mounted camera video dataset](#) - Activities of Daily Living videos captured from a wrist- mounted camera and a head-mounted camera(Katsunori Ohnishi, Atsushi Kanehira,Asako Kanezaki, Tatsuya Harada)
 46. [Yummlly-10k dataset](#) - The goal was to understand human perception, in this case of food taste similarity.(SE(3) Computer Vision Group at Cornell Tech)

Miscellaneous

1. [3D mesh watermarking benchmark dataset](#) (Guillaume Lavoue)
2. [4D Light Field Dataset](#) - 24 synthetic scenes with 9x9x512x512x3 input images, depth and disparity ground truth, camera parameters, and evaluation masks. (Katrin Honauer, Ole Johannsen, Daniel Kondermann, Bastian Goldluecke)
3. [A Dataset for Real Low-Light Image Noise Reduction](#) - It contains pixel and intensity aligned pairs of images corrupted by low-light camera noise and their low-noise counterparts. (J. Anaya, A. Barbu)
4. [AF 4D dataset](#) - Based on our observations, we settled on 10 representative scenes that are categorized into three types: (1) scenes containing no face (NF), (2) scenes with a face in the foreground (FF), and (3) scenes with faces in the background (FB). For each of these scenes, we allowed different arrangements in terms of textured backgrounds, whether the camera moves, and how many types of objects in the scene change their directions (referred to as motion switches). (Abdullah Abuolaim, York University)
5. [AMADI LontarSet](#) - Balinese Palm Leaf Manuscript Images Dataset for Binarization, Query-by-Example Word Spotting, and Isolated Character Recognition of Balinese Script. (The AMADI Project et al.)
6. [Active Appearance Models datasets](#) (Mikkel B. Stegmann)
7. [Aircraft tracking](#) (Ajmal Mian)
8. [Annotated Web Ears Dataset \(AWE Dataset\)](#) - All images were acquired by cropping ears from images from the internet of known persons. (Ziga Emersic, Vitomir Struc and Peter Peer)
9. [CITIUS Video Database](#) - A database of 72 videos with eye-tracking data for evaluate dynamic saliency visual models. (Xose)
10. [CrowdFlow](#) - Optical flow dataset and benchmark for crowd analytics (Gregory Schroeder, Tobias Senst, Erik Bochinski, Thomas Sikora)
11. [CVSSP 3D data repository](#) - The datasets are designed to evaluate general multi-view reconstruction algorithms. (Armin Mustafa, Hansung Kim, Jean-Yves Guillemaut and Adrian Hilton)
12. [California-ND](#) - 701 photos from a personal photo collection, including many challenging real-life non-identical near-duplicates (Vassilios Vonikakis)
13. [Cambridge Motion-based Segmentation and Recognition Dataset](#) (Brostow, Shotton, Fauqueur, Cipolla)
14. [Catadioptric camera calibration images](#) (Yalin Bastanlar)
15. [Chars74K dataset - 74 English and Kannada characters](#) (Teo de Campos - t.decampos@surrey.ac.uk)
16. [Coin Image Dataset](#) - The coin image dataset is a dataset of 60 classes of Roman Republican coins (Sebastian Zambanini, Klaus Vondrovec)

17. [Columbia Camera Response Functions: Database \(DoRF\) and Model \(EMOR\)](#) (M.D. Grossberg and S.K. Nayar)
18. [Columbia Database of Contaminants' Patterns and Scattering Parameters](#) (Jinwei Gu, Ravi Ramamoorthi, Peter Belhumeur, Shree Nayar)
19. [COVERAGE](#) - copy-move forged (CMFD) images and their originals with similar but genuine objects (SGOs), which highlight and address tamper detection ambiguity of popular methods, caused by self-similarity within natural images (Wen, Zhu, Subramanian, Ng, Shen, and Winkler)
20. [Crime Scene Footwear Impression Database](#) - crime scene and reference footwear impression images (Adam Kortylewski)
21. [Curve tracing database for an automatic grading system.](#) - The ground truth database of 70 public images used to evaluate our method Bandeirantes and other curve tracing methods in an automatic grading system. (Marcos A. Tejada Condori, Paulo A. V. Miranda)
22. [D-HAZY](#) - : A DATASET TO EVALUATE QUANTITATIVELY DEHAZING ALGORITHMS (Cosmin Ancuti et al.)
23. [DR\(eye\)VE](#) - A driver's attention dataset (University of Modena and Reggio Emilia)
24. [DTU controlled motion and lighting image dataset \(135K images\)](#) (Henrik Aanaes)
25. [Database for Visual Eye Movements \(DOVES\)](#) - A set of eye movements collected from 29 human observers as they viewed 101 natural calibrated images. (van der Linde, I., Rajashekar, U., Bovik, A. C. etc.)
26. [DeformIt 2.0](#) - Image Data Augmentation Tool: Simulate novel images with ground truth segmentations from a single image-segmentation pair (Brian Booth and Ghassan Hamarneh)
27. [Dense outdoor correspondence ground truth datasets, for optical flow and local keypoint evaluation](#) (Christoph Strecha)
28. [EISATS](#): .enpeda.. Image Sequence Analysis Test Site (Auckland University Multimedia Imaging Group)
29. [Featureless object tracking](#) - This dataset contains several videosequences with limited texture, intended for visual tracking, including manually annotated per-frame pose.(Lebeda, Hadfield, Matas, Bowden)
30. [FlickrLogos-32](#) - 8240 images of 32 product logos (Stefan Romberg)
31. [General 100 Dataset](#) - General-100 dataset contains 100 bmp-format images (with no compression), which are well-suited for super-resolution training(Dong, Chao and Loy, Chen Change and Tang, Xiaoou)
32. [Geometry2view](#) - This dataset contains image pairs for 2-view geometry computation, including manually annotated point coordinates.(Lebeda, Matas, Chum)
33. [Hannover Region Detector Evaluation Data Set](#) - Feature detector evaluation sequences in multiple image resolutions from 1.5 up to 8 megapixels (Kai Cordes)
34. [Hillclimb and CubicGlobe datasets](#) - a video of a rally car, separated into

- several independent shots (for visual tracking and modelling). (Lebeda, Hadfield, Bowden)
35. [Houston Multimodal Distracted Driving Dataset](#) - 68 volunteers that drove the same simulated highway under four different conditions (Dcosta, Buddharaju, Khatri, and Pavlidis)
 36. [HyperSpectral Salient Object Detection Dataset \(HS-SOD Dataset\)](#) - Hyperspectral (visible spectrum) image data for benchmarking on salient object detection with a collection of 60 hyperspectral images with their respective ground-truth binary images and representative rendered colour images (rendered in sRGB). (Nevrez Imamoglu, Yu Oishi, Xiaoqiang Zhang, Guanqun Ding, Yuming Fang, Toru Kouyama, Ryosuke Nakamura)
 37. [I3 - Yahoo Flickr Creative Commons 100M](#) - This dataset contains a list of photos and videos. (B. Thomee, D.A. Shamma, G. Friedland et al.)
 38. [ICDAR'15 Smartphone document capture and OCR competition - challenge 2](#) - pictures of documents captured with smartphones under various conditions of perspective, lighting, etc. The ground truth is the textual content which should be extracted. (Burie, Chazalon, Coustaty, Eskenazi, Luqman, Mehri, Nayef, Ogier, Prum and Rusinol)
 39. [L-HAZE](#) - A dehazing benchmark with real hazy and haze-free indoor images. (ethz)
 40. [Intrinsic Images in the Wild \(IIW\)](#) - Intrinsic Images in the Wild, is a large-scale, public dataset for evaluating intrinsic image decompositions of indoor scenes (Sean Bell, Kavita Bala, Noah Snavely)
 41. [IISc - Dissimilarity between Isolated Objects \(IISc-DIO\)](#) - The dataset has a total of 26,675 perceived dissimilarity measurements made on 269 human subjects using a Visual Search task with a diverse set of objects.(RT Pramod & SP Arun, IISc)
 42. [INRIA feature detector evaluation sequences](#) (Krystian Mikolajczyk)
 43. [Image/video quality assessment database summary](#) (Stefan Winkler)
 44. [INRIA's PERCEPTION's database of images and videos gathered with several synchronized and calibrated cameras](#) (INRIA Rhone-Alpes)
 45. [KITTI dataset for stereo, optical flow and visual odometry](#) (Geiger, Lenz, Urtasun)
 46. [LabelMe images database and online annotation tool](#) (Bryan Russell, Antonio Torralba, Kevin Murphy, William Freeman)
 47. [Large scale 3D point cloud data from terrestrial LiDAR scanning](#) (Andreas Nuechter)
 48. [LFW-10 dataset for learning relative attributes](#) - A dataset of 10,000 pairs of face images with instance-level annotations for 10 attributes.(CVIT, IIIT Hyderabad.)
 49. [Light-field Material Dataset](#) - 1.2k annotated images of 12 material classes taken with the Lytro ILLUM camera(Ting-Chun Wang, Jun-Yan Zhu, Ebi Hiroaki,Manmohan Chandraker, Alexei Efros, Ravi Ramamoorthi)
 50. [Linkoping Rolling Shutter Rectification Dataset](#) (Per-Erik Forssen and Erik

Ringaby)

51. [LIRIS-ACCEDE Dataset](#) - a collection of video excerpts with a large content diversity annotated along affective dimensions (Technicolor)
52. [MARIS Portofino dataset](#) - A dataset of underwater stereo images depicting cylindrical pipe objects and collected to test object detection and pose estimation algorithms. (RIMLab (Robotics and Intelligent Machines Laboratory), University of Parma.)
53. [Materials in Context \(MINC\)](#) - The Materials in Context Database (MINC) builds on OpenSurfaces, but includes millions of point annotations of material labels. (Sean Bell, Paul Upchurch, Noah Snavely, Kavita Bala)
54. [MASSVIS \(Massive Visualization Dataset\)](#) - Over 5K different information visualizations from a variety of sources, a subset of which have been categorized, segmented, and come with memorability and eye tracking recordings. (Borkin, Bylinskii, Kim, Oliva, Pfister)
55. [MPI Sintel Flow Dataset](#) A data set for the evaluation of optical flow derived from the open source 3D animated short film, Sintel. It has been extended for Stereo and disparity, Depth and camera motion, and Segmentation. (Max Planck Tubingen)
56. [MPI-Sintel optical flow evaluation dataset](#) (Michael Black)
57. [MSR-VTT](#) - video to text database of 200K+ video clip/sentence pairs
58. [Middlebury College stereo vision research datasets](#) (Daniel Scharstein and Richard Szeliski)
59. [Modelling of 2D Shapes with Ellipses](#) - he dataset contains 4,526 2D shapes included in standard as well as in home-build datasets.(Costas Panagiotakis and Antonis Argyros)
60. [Multi-FoV](#) - Photo-realistic video sequences that allow benchmarking of the impact of the Field-of-View (FoV) of the camera on various vision tasks. (Zhang, Rebecq, Forster, Scaramuzza)
61. [Multiview Stereo Evaluation](#) - Each dataset is registered with a "ground-truth" 3D model acquired via a laser scanning process(Steve Seitz et al)
62. [Multiview stereo images with laser based groundtruth](#) (ESAT-PSI/VISICS,FGAN-FOM,EPFL/IC/ISIM/CVLab)
63. [NCI Cancer Image Archive](#) - prostate images (National Cancer Institute)
64. [NIST 3D Interest Point Detection](#) (Helin Dutagaci, Afzal Godil)
65. [NRCS natural resource/agricultural image database](#) (USDA Natural Resources Conservation Service)
66. [O-HAZE](#) - A dehazing benchmark with real hazy and haze-free outdoor images. (ethz)
67. [Object recognition dataset for domain adaptation](#) - Consists of images from 4 different domains: Artistic images, Clip Art, Product images and Real-World images. For each domain, the dataset contains images of 65 object categories found typically in Office and Home settings. (Venkateswara Hemanth, Eusebio Jose, Chakraborty Shayok, Panchanathan Sethuraman)
68. [Object Removal](#) - Generalized Dynamic Object Removal for Dense Stereo

- Vision Based Scene Mapping using Synthesised Optical Flow - Evaluation Dataset (Hamilton, O.K., Breckon, Toby P.)
69. [Occlusion detection test data](#) (Andrew Stein)
 70. [OpenSurfaces](#) - OpenSurfaces consists of tens of thousands of examples of surfaces segmented from consumer photographs of interiors, and annotated with material parameters, texture information, and contextual information . (Kavita Bala et al.)
 71. [OSIE - Object and Semantic Images and Eye-tracking](#) - 700 images, 5551 segmented objects, eye tracking data (Xu, Jiang, Wang, Kankanhalli, Zhao)
 72. [Osnabrück gaze tracking data](#) - 318 video sequences from several different gaze tracking data sets with polygon based object annotation (Schöning, Faion, Heidemann, Krumnack, Gert, Açık, Kietzmann, Heidemann & König)
 73. [OTIS: Open Turbulent Image Set](#) - several sequences (either static or dynamic) of long distance imaging through a turbulent atmosphere (Jerome Gilles, Nicholas B. Ferrante)
 74. [PanoNavi dataset](#) - A panoramic dataset for robot navigation, consisted of 5 videos lasting about 1 hour. (Lingyan Ran)
 75. [PetroSurf3D](#) - 26 high resolution (sub-millimeter accuracy) 3D scans of rock art with pixelwise labeling of petroglyphs for segmentation(Poier, Seidl, Zeppelzauer, Reinbacher, Schaich, Bellandi, Marretta, Bischof)
 76. [PHOS \(illumination invariance dataset\)](#) - 15 scenes captured under different illumination conditions * 15 images (Vassilios Vonikakis)
 77. [PIRM](#) - perceptual quality of super-resolution benchmark (Blau, Y., Mechrez, R., Timofte, R., Michaeli, T., Zelnik-Manor, L)
 78. [PittsStereo-RGBNIR](#) - A Large RGB-NIR Stereo Dataset Collected in Pittsburgh with challenging Materials. (Tiancheng Zhi, Bernardo R. Pires, Martial Hebert and Srinivasa G. Narasimha)
 79. [PRINTART](#): Artistic images of prints of well known paintings, including detail annotations. A benchmark for automatic annotation and retrieval tasks with this database was published at ECCV. (Nuno Miguel Pinho da Silva)
 80. [Pics 'n' Trails - Dataset of Continuously archived GPS and digital photos](#) (Gamhewage Chaminda de Silva)
 81. [Pitt Image and Video Advertisement Understanding](#) - rich annotations encompassing the topic and sentiment of the ads, questions and answers describing what actions the viewer is prompted to take and the reasoning that the ad presents to persuade the viewer (Hussain, Zhang, Zhang, Ye, Thomas, Agha, Ong, Kovashka (University of Pittsburgh)>
 82. [RAWSEEDS SLAM benchmark datasets](#) (Rawseeds Project)
 83. [ROMA \(ROad MARKings\) : Image database for the evaluation of road markings extraction algorithms](#) (Jean-Philippe Tarel, et al)
 84. [Robotic 3D Scan Repository](#) - 3D point clouds from robotic experiments of scenes (Osnabruck and Jacobs Universities)
 85. [Rolling Shutter Rectification Dataset](#) - Rectifying rolling shutter video from

- hand-held devices (Per-Erik etc.)
86. [SALICON](#) - Saliency in Context eye tracking dataset c. 1000 images with eye-tracking data in 80 image classes. (Jiang, Huang, Duan, Zhao)
 87. [Scripps Plankton Camera System](#) - thousands of images of c. 50 classes of plankton and other small marine objects (Jaffe et al)
 88. [ScriptNet: ICDAR2017 Competition on Historical Document Writer Identification \(Historical-WI\)](#) - The dataset consists of 4782 handwritten pages written by more than 1100 writers and dating from the 13th to 20th century. (Fiel Stefan, Kleber Florian, Diem Markus, Christlein Vincent, Louloudis Georgios, Stamatopoulos Nikos, Gatos Basili)
 89. [Seam Carving JPEG Image Database](#) - Our seam-carving-based forgery database contains 500 untouched JPEG images and 500 JPEG images that were manipulated by seam-carving, both at the quality of 75 (Qingzhong Liu)
 90. [SIDIRE: Synthetic Image Dataset for Illumination Robustness Evaluation](#) - SIDIRE is a freely available image dataset which provides synthetically generated images allowing to investigate the influence of illumination changes on object appearance (Sebastian Zambanini)
 91. [Smartphone document capture and OCR 2015 - Quality Assessment](#) - pictures of documents captured with smartphones under various conditions perspective, lighting, etc. It also features text ground truth and OCR accuracies to train and test document image quality assessment systems. (Nayef, Luqman, Prum, Eskenazi, Chazalon, and Ogier)
 92. [Smartphone document capture and OCR 2017 - mobile video capture](#) - video recording of documents, along with the reference ground truth image to reconstruct using the video stream. (Chazalon, Gomez-Kr??mer, Burie, Coustaty, Eskenazi, Luqman, Nayef, Rusi??ol, Sid??re, and Ogier)
 93. [Stony Brook University Real-World Clutter Dataset \(SBU-RwC90\)](#) - Images of different level of clutteriness, ranked by humans (Chen-Ping Yu, Dimitris Samaras, Gregory Zelinsky)
 94. [Street-View Change Detection with Deconvolutional Networks](#) - Database with aligned image pairs from street-view imagery with structural, lighting, weather and seasonal changes. (Pablo F. Alcantarilla, Simon Stent, German Ros, Roberto Arroyo and Riccardo Gherardi)
 95. [SydneyHouse](#) - Streetview house images with accurate 3D house shape, facade object label, dense point correspondence, and annotation toolbox. (Hang Chu, Shenlong Wang, Raquel Urtasun, Sanja Fidler)
 96. [SYNTHIA](#) - Large set (~half million) of virtual-world images for training autonomous cars to see. (ADAS Group at Computer Vision Center)
 97. [Stony Brook University Shadow Dataset \(SBU-Shadow5k\)](#) - Large scale shadow detection dataset from a wide variety of scenes and photo types, with human annotations (Tomas F.Y. Vicente, Le Hou, Chen-Ping Yu, Minh Hoai, Dimitris Samaras)
 98. [Technicolor Interestingness Dataset](#) - a collection of movie excerpts and

- key-frames and their corresponding ground-truth files based on the classification into interesting and non-interesting samples (Technicolor)
99. [Technicolor Hannah Dataset](#) - 153,825 frames from the movie "Hannah and her sisters" annotated for several types of audio and visual information (Technicolor)
 100. [Technicolor HR-EEG4EMO Dataset](#) - EEG and other physiological recordings of 40 subjects collected during the viewing of neutral and emotional videos (Technicolor)
 101. [Technicolor VSD Violent Scenes Dataset](#) - a collection of ground-truth files based on the extraction of violent events in movies (Technicolor)
 102. [The Conflict Escalation Resolution \(CONFER\) Database](#) - 120 audio-visual episodes (~142 mins) of naturalistic interactions from televised political debates, annotated frame-by-frame in terms of real-valued conflict intensity. (Christos Georgakis, Yannis Panagakis, Stefanos Zafeiriou, Maja Pantic)
 103. [The Open Video Project](#) (Gary Marchionini, Barbara M. Wildemuth, Gary Geisler, Yaxiao Song)
 104. [The Toulouse Vanishing Points Dataset](#) - a dataset of Manhattan scenes for vanishing point estimation which also provide, for each image, the IMU data of the camera orientation. (Vincent Angladon and Simone Gasparini)
 105. [TMAGIC dataset](#) - Several videosequences for visual tracking, containing strong out-of-plane rotation (Lebeda, Hadfield, Bowden)
 106. [Totally Looks Like](#) - A benchmark for assessment of predicting human-based image similarity (Amir Rosenfeld, Markus D. Solbach, John Tsotsos)
 107. [TUM RGB-D Benchmark](#) - Dataset and benchmark for the evaluation of RGB-D visual odometry and SLAM algorithms (BCrgen Sturm, Nikolas Engelhard, Felix Endres, Wolfram Burgard and Daniel Cremers)
 108. [UCL Ground Truth Optical Flow Dataset](#) (Oisin Mac Aodha)
 109. [Underwater Single Image Color Restoration](#) - A dataset of forward-looking underwater images, enabling a quantitative evaluation of color restoration using color charts at different distances and ground truth distances using stereo imaging. (Berman, Levy, Avidan, Treibitz)
 110. [Univ of Genoa Datasets for disparity and optic flow evaluation](#) (Manuela Chessa)
 111. [Validation and Verification of Neural Network Systems](#) (Francesco Vivarelli)
 112. [Very Long Baseline Interferometry Image Reconstruction Dataset](#) (MIT CSAIL)
 113. [Virtual KITTI](#) - 40 high-resolution videos (17,008 frames) generated from five different virtual worlds, for : object detection and multi-object tracking, scene-level and instance-level semantic segmentation, optical flow, and depth estimation (Gaidon, Wang, Cabon, Vig)
 114. [Visual Object Tracking challenge](#) - This challenge is held annually as an ICCV/ECCV workshop, with a new dataset and an updated evaluation kit every year. (Kristan et al.)
 115. [WHOI-Plankton](#) - 3.5 million images of microscopic marine plankton on 103

categories (Olson, Sosik)

116. [WILD: Weather and Illumination Database](#) (S. Narasimhan, C. Wang. S. Nayar, D. Stolyarov, K. Garg, Y. Schechner, H. Peri)
117. [YACCLAB dataset](#) - YACCLAB dataset includes both synthetic and real binary images(Grana, Costantino; Bolelli, Federico; Baraldi, Lorenzo; Vezzani, Roberto)
118. [YtLongTrack](#) - This dataset contains two video sequences with challenges such as low quality, extreme length and full occlusions, including manually annotated per-frame pose.(Lebeda, Hadfield, Matas, Bowden)

Acknowledgements: Many thanks to all of the contributors for their suggestions of databases. Can Pu and Hanz Cuevas Velasquez were very helpful with the updating of this web page.

[Return to CVentry top level](#)

Date of last change to this page: 05/31/2019 16:47:27



© 2018 Robert Fisher