Chapter 11
3D Visual Content Datasets

Karel Fliegel, Federica Battisti, Marco Carli, Margrit Gelautz, Lukáš Krasula, Patrick Le Callet and Vladimir Zlokolica

Abstract Development and performance evaluation of efficient methods for coding, transmission, and quality assessment of 3D visual content require rich datasets of a suitable test material. The use of these databases allows a fair comparison of systems under test. Moreover, publicly available and widely used datasets are crucial for experimentation leading to reproducible research. This chapter presents an overview of 3D visual content datasets relevant to research in the field of coding, transmission, and quality assessment. Description of regular stereoscopic or multiview image and video datasets is presented. Databases created using emerging technologies, including light-field imaging, are also addressed. Moreover, there are databases of multimedia content annotated with ratings from the subjective experiment, which are a necessary resource for understanding the complex problem of quality of experience while consuming the 3D visual content.

K. Fliegel (✉)
Department of Radioelectronics, Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic
e-mail: fliegek@fel.cvut.cz

F. Battisti · M. Carli
University of Roma TRE, Rome, Italy
e-mail: federica.battisti@uniroma3.it

M. Carli
e-mail: marco.carli@uniroma3.it

M. Gelautz
Vienna University of Technology, Vienna, Austria
e-mail: margrit.gelautz@tuwien.ac.at

L. Krasula · P. Le Callet
University of Nantes, Nantes, France
e-mail: l.krasula@gmail.com

P. Le Callet
e-mail: patrick.lecallet@univ-nantes.fr

V. Zlokolica
Faculty of Technical Sciences, University of Novi Sad, Novi Sad, Serbia
e-mail: vzlokolica@uns.ac.rs

P. A. Assunção and A. Gotchev (eds.), *3D Visual Content Creation, Coding and Delivery*, Signals and Communication Technology,
https://doi.org/10.1007/978-3-319-77842-6_11

11.1 Introduction

The suitable test material, in this context 3D visual content, plays a crucial role in the development and performance evaluation of related coding, transmission, and quality assessment methods. Publicly available and widely used datasets are necessary for fair performance comparison and validation of systems under test and thus crucial for experimentation leading to reproducible research. Numerous research laboratories produced the relevant databases of 3D visual content. The content description is usually published in technical reports, research papers, and online resources, thus it is very scattered, and it is not easy to identify the most suitable dataset for the particular needs.

There were numerous efforts to provide overview and comparison of multimedia content datasets. Among the first published descriptions belong image and video quality resources website[1] by Stefan Winkler and related publications [1–3] providing in-depth analysis of multimedia content databases. Another notable achievement with the goal to provide rich and internationally recognized database of content of different sorts is "QUALINET Multimedia Databases Online" platform[2] created in the frame of ICT COST Action IC1003 "European Network on Quality of Experience in Multimedia Systems and Services" (QUALINET).[3] The platform, abbreviated "Qualinet Databases", is used to share the databases efficiently with other researchers and handles information on the multimedia content. The database was substantially extended to 3D visual content within the frame of ICT COST Action IC1105 "3D Content Creation, Coding and Transmission over Future Media Networks" (3D-ConTourNet).[4] As of September 2017, Qualinet Databases contains 241 registered datasets, from which about 30 datasets cover relevant 3D visual content, and there are more than 400 registered users.

3D visual content datasets relevant to research in the field of coding, transmission, and quality assessment are overviewed in this chapter. The chapter is focused mainly on selected databases available in the public domain. The databases are categorized, and then a detailed comparison of available datasets in various application domains is presented to help the users with the decision about which database is more suitable for the particular problem. For each discussed database an overview of the material is presented along with the details on how the content was created. Where available, also, experimental image acquisition setup and subjective experiment design are discussed.

The chapter has the following structure related to the fundamental categorization of 3D visual content datasets. At first, stereoscopic and multiview image and video content datasets are introduced in Sect. 11.2, with the basic description of related stereo dataset generation and multiview camera content for 3D reconstruction,

[1]Image and Video Quality Resources (http://stefan.winklerbros.net/resources.html).

[2]Qualinet Databases (http://dbq.multimediatech.cz/).

[3]COST Action IC1003 QUALINET (http://www.qualinet.eu/).

[4]COST Action IC1105 3D-ConTourNet (http://www.3d-contournet.eu/).

modeling, and visualization. Then, light-field content characterization and selection is addressed in Sect. 11.3 with focus on perceptual assessment. Special point-cloud and holographic content datasets are also reviewed in Sect. 11.4 with respect to image compression standardization activities. The most popular datasets annotated with ratings from subjective experiments are discussed in Sect. 11.5 and the chapter is concluded in Sect. 11.6.

11.2 Stereoscopic and Multiview Visual Content Datasets

In the following paragraphs, several stereo and multiview image and video datasets with reference depth are reviewed. These datasets have been made publicly available by the computer vision community. The creation of these datasets was primarily motivated by the need of depth ground truth to support the design and quantitative evaluation of computer vision algorithms, especially in the field of stereo matching. A particular merit of such repositories is the detailed information on how the data were created and the accuracy of their associated ground truth. Furthermore, ancillary information such as occlusion maps is often provided. Beyond their initial purpose of benchmarking computer vision algorithms, the stereo and multiview image plus depth data contained in these datasets can also give valuable support in the context of coding, transmission, and quality assessment of 3D visual content.

11.2.1 Stereo Dataset Generation for Different Scene Cases

Different approaches for generating stereo or multiview imagery with reference depth can be used to assess the quality of stereo matching or 3D reconstruction results. Similar to [4], datasets are distinguished between real scenes, laboratory scenes, and synthetic data, which are discussed in the following paragraphs.

Real Scenes

Real scenes have the advantage of rich natural texture and can faithfully represent diverse application scenarios. However, the simultaneous acquisition of depth ground truth for real outdoor videos typically requires the usage of a relatively expensive laser scanning device and techniques for coregistration between the depth measurements and RGB video. It may also result in missing depth information in areas that could not be mapped successfully by the employed 3D sensor. A well-established database that includes real stereo images and videos of traffic scenes taken by cameras mounted on a moving car is the KITTI dataset [5]. A stereo benchmark that also contains multiview data and stereo videos taken with

mobile devices of real indoor and outdoor environments has been published recently [4].

Laboratory Scenes

The acquisition of depth ground truth is alleviated for indoor laboratory scenes, where structured light techniques with multiple exposure patterns can achieve high reconstruction accuracy for stationary settings. The controlled laboratory environment supports the acquisition of multiple images of the same scene taken from different viewpoints or under varying illumination conditions. The chosen spatial arrangement and surface characteristics such as material properties or texture allow to specifically address challenges such as occlusions or specular reflections, which have an impact on the quality of the stereo reconstruction and derived new views. A notable example of this group is the widely known Middlebury benchmark for stereo matching [6], which has been a driving factor for the development of stereo matching algorithms over the past 15 years.

Simulated Data

The main advantage of simulated data, as opposed to real and laboratory scenes, is that dense coregistered depth maps are generated as a byproduct of the rendering process, and video acquisitions with complex camera motions can be synthesized using freely chosen viewpoints. Related to this, certain sets of 3D models have been standardized (such as "Stanford Bunny"), which are used in virtual environments under controllable conditions to render stereo images, from which the 3D can be reconstructed and compared to the starting ground-truth 3D object. One example of such case is the Sintel dataset [7], which comprises stereo videos that were acquired with different rendering options accounting for a variety of shading and illumination effects. The obvious downside of synthetic data is their limited degree of reality, especially for natural outdoor scenes, which has traditionally limited their applicability in the design and evaluation of vision algorithms. However, there are notable recent developments which exploit high-quality renderings delivered by commercial computer games to generate synthetic data whose high degree of realism has already proven effective in the training of machine learning algorithms [8]. This shows the potential of simulation approaches to substitute real data in applications where additional information such as semantic labeling or depth maps are necessary. It is an open question whether such highly photorealistic computer images, which can be produced with a large variety of rendering options, can also support, for example, the development of improved objective quality metrics.

Description of the three selected example datasets are given in the following paragraphs.

MPI-Sintel

The dataset, which is available at the dedicated website,[5] is derived from an open source 3D animated movie that was created using the 3D creation software Blender.[6] The main purpose of the data collection is to provide ground truth for evaluation of optical flow algorithms. However, the dataset as described in the original publication [7] has been augmented over time and offers now also stereo videos with ground-truth disparity[7] and ground-truth depth maps with corresponding camera parameters[8] for download. The stereo dataset was created by simulating two parallel-viewing cameras that are placed 10 cm apart. It comprises 23 scenes consisting of up to 50 frames, captured at a resolution of 1024×436 pixels. The stereo videos are offered in two rendering options denoted as "clean" and "final", with the latter one accommodating not only illumination effects such as shading and specular reflections but also additional blurring due to camera motion or depth of field. The reference disparity maps are accompanied by masks that indicate missing stereo correspondences due to occlusions or border effects.

ETH3D

A very recent dataset containing natural multiview stereo imagery with associated depth ground truth has been released [4] at the dedicated website.[9] The data is provided along with online benchmarks that evaluate multiview reconstructions based on accuracy and completeness, and two-view results by measuring the disparity errors. The ETH3D repository includes 13 multiview stereo scenes captured with high resolution (6048×4032 pixels) by a Nikon D3X camera and five lower resolution videos (752×480 pixels) recorded by a synchronized multi-camera rig in a mobile setup with automated exposure settings. A specific design goal was to provide a broad range of indoor and outdoor scenes with both natural and man-made content, including some fine scene details such as trees or wires, which are challenging to reconstruct. The (non-dense) depth ground truth is delivered by a Faro Focus X 330 laser scanner.

KITTI

The KITTI Vision Benchmark Suite[10] was developed in the specific context of autonomous driving. It includes ground-truth datasets and evaluation tables for a variety of computer vision tasks including stereo and optical flow. The stereo cameras and a Velodyne laser scanner for capturing the color/gray-value images and ground-truth depth, respectively, were mounted on the roof of a driving vehicle. Additional multiview data are provided by consecutive frames of recorded video

[5]MPI-Sintel dataset (http://sintel.is.tue.mpg.de/).

[6]Blender (http://www.blender.org).

[7]MPI-Sintel stereo videos with ground truth disparity (http://sintel.is.tue.mpg.de/stereo).

[8]MPI-Sintel ground truth depth maps (http://sintel.is.tue.mpg.de/depth).

[9]ETH3D dataset (www.eth3d.net).

[10]KITTI Vision Benchmark Suite (http://www.cvlibs.net/datasets/kitti/).

sequences. The repository includes stereo images (at a resolution of approximately 0.5 Megapixels) of static environments along with semi-dense ground truth, which was released in 2012, and a dataset from 2015 comprising also dynamic scene objects. More information on the ground-truth acquisition, evaluation method, and ancillary data can be found in the literature [5, 9].

11.2.2 Multiview Camera Content for 3D Reconstruction, Modeling, and Visualization

In the following paragraphs, the existing multiview stereo (and photometric stereo) datasets that have been made available by different research groups are described. Details are provided for each dataset, including where they can be found, and their aimed application is explained. These datasets have been standardized to certain extent and can be used for benchmarking of the algorithms related to their target application, which is indicated by the reference publication (also attached to the explanation of the corresponding dataset).

Reconstructing 3D content from multiview camera images [10] has shown to be an efficient approach, which in the most general case does not require special setup and special kind of sensors. Consequently, such an approach can be performed flexibly and inexpensively in different environmental conditions, in comparison to other more advanced 3D acquisition methodologies. An additional advantage of such an approach is that the subsequent texture mapping and 3D modeling can be subsequently performed and optimized more easily for the target 3D visualization application.

However, currently, there is still lack of the standardized datasets for validation of the 3D multiview stereo reconstruction approaches as well as the availability of objective metrics for evaluating reconstructed 3D content [11, 12], in full reference or non-reference sense. One common existing method for 3D data reconstruction validation [11, 12] is to have a priori known camera configuration setup and then make images from different views, provide calibration data and also additionally provide the ground truth (obtained by some other more precise and expensive sensor technology) so that reconstruction quality could be assessed objectively, in full reference sense, in terms of accuracy and completeness.[11] The accuracy is usually computed as mean square error to the known 3D point cloud, while the completeness is determined as the number of 3D points being reconstructed.

The multiview stereo dataset from Middlebury (see footnote 11) is one of the most standardized sites in 3D community for multiview 3D reconstruction [11]. It provides two high-quality datasets: (i) Dino and (ii) Temple, which can be used for benchmarking and performance evaluation of the multiview stereo reconstruction algorithms. Each dataset is registered with a ground-truth 3D model acquired via a

[11]Middlebury dataset (http://vision.middlebury.edu/mview).

laser scanning process, to be used as a baseline for measuring accuracy and completeness. The ground truth cannot be downloaded directly though; usually one performs reconstruction, and then the evaluation testing is done per request. The multiview images are provided with different number of cameras and are of the 640 × 480 resolution. The images were captured using the Stanford Spherical Gantry[12] which enables moving a camera on a sphere to specified latitude/longitude angles. In order to obtain ground-truth model, the object was scanned from several orientations using a Cyberware Model 15 laser scanner.

An additional standardized 3D dataset is provided by the Stanford 3D Scanning Repository.[13] The purpose of this repository is to make some range data and detailed reconstructions available to the public for benchmarking. It provides different 3D models that can be used, for example, in the virtual environment to render the images and subsequently use them for 3D reconstruction purposes. The virtually generated images with known 3D model represent a valuable case scenario because it can be used for different use-cases exploration and algorithm evaluation and tunings. For generating the 3D models, different kind of 3D scanners have been used, which provide different quality of 3D viewing and 3D visualization for the target application.

On the other hand, most of the other available datasets provide only multiview images along with calibration data and silhouettes that can be used for 3D reconstruction evaluation, 3D modeling, and visualization. Selected examples of relevant datasets are listed below.

EPFL—Computer Vision Group CVLAB

The Computer Vision Group CVLAB at EPFL provides extensive datasets related to multi-camera visualization of the 3D content and 3D registration, in indoor and outdoor environment, for various applications, such as tree structure reconstruction, multiview evaluation, stereo face database, multiview stereo, multi-camera pedestrians video, multiview car dataset, deformable surface reconstruction, etc. The multiview evaluation dataset[14] represents one of the most important available dataset for evaluation of the multi-camera calibration and 3D reconstruction algorithms based on multiview imaging [12]. It consists of six multiview datasets with ground-truth 3D point cloud and rendered 3D model. Moreover, it also provides results of different structure from motion algorithms for the given data.

Next, stereo face database[15] is provided, which consists of 100 faces in eight positions captured by two cameras. These datasets are generated for the purpose of validation for the proposed approach for face modeling and face recognition from a pair of calibrated stereo cameras [13]. However, it can be used more generally for

[12]Stanford Spherical Gantry (https://graphics.stanford.edu/projects/gantry/).

[13]Stanford 3D Scanning Repository (http://graphics.stanford.edu/data/3Dscanrep/).

[14]CVLAB multiview evaluation dataset (https://cvlab.epfl.ch/).

[15]CVLAB stereo face database (http://cvlab.epfl.ch/data/stereoface).

stereo 3D reconstruction, 3D face modeling, and motion tracking. The dataset contains the camera parameter file including intrinsic matrix K, radial distortion, rotation matrixes, and translation vector. The camera images size is $640 \times 480 \times 3$, captured in controllable indoor environment.

The additional two datasets are the multiview stereo set of buildings and multiview car dataset. The multiview stereo set of buildings in outdoor environment for dense depth and 3D reconstruction[16] contains images of mid-resolution size (approx. $3000 \times 2000 \times 3$) and has been generated for validation of the proposed stereo from multiple views method [14, 15]. The original images have been compensated for radial distortion, and external and internal calibration parameters have been provided along. Additionally, initial 3D points from calibration have also been provided. The multiview car dataset[17] contains 20 sequences of cars as they rotate by 360°. There is one image approximately every 3°–4°. Using the time of capture information from the photos, it is possible to calculate the approximate rotation angle of the car. The dataset has been used for the multiview object pose estimation algorithm [16] but represents a good dataset also for general 3D reconstruction and 3D registration validation purpose.

TUM—Computer Vision Group

There are multiple datasets available capturing objects from various vantage points.[18] Each entry contains an image sequence, corresponding silhouettes, and full calibration parameters. The camera configuration setup consists of circular configuration with special lighting in indoor conditions. In this setup, five different objects were captured from various positions ("bird", "beethoven", "bunny", "head", "pig"). This dataset is specifically generated for validation of the proposed multiview camera 3D reconstruction [17], but it represents a valuable dataset to be used for 3D reconstruction benchmarking and performance comparison between different methods.

Cornell 3D Location Recognition Datasets

The 3D Location Recognition Datasets[19] contain a large amount of multiview images of Rome and Dubrovnik [18], which can be used for 2D-to-3D matching, i.e., 3D reconstruction from multiple views, based on which 3D point cloud can be obtained and used for 3D modeling and visualization evaluation.

[16]CVLAB stereo dataset of buildings (http://cvlab.epfl.ch/data/strechamvs).

[17]CVLAB multiview car dataset (http://cvlab.epfl.ch/data/pose).

[18]TUM—Computer Vision Group (http://vision.in.tum.de/data/datasets/3dreconstruction).

[19]Cornell 3D Location Recognition Datasets (http://www.cs.cornell.edu/projects/p2f).

Washington University Photo Tourism Dataset

The dataset[20] represents 715-image reconstruction of Notre Dame Cathedral in Paris, which can be used for 3D reconstruction, modeling, and visualization evaluation.

Photometric Stereo Datasets

Besides multiview stereo reconstruction algorithms, substantial progress has been made in the development of photometric stereo methodologies, which can deal with general materials and unknown illumination conditions. The main idea here is to use a single camera and capture multiple images with changeable lighting conditions, where one usually uses controllable lighting conditions. This approach is particularly valuable and important for performing fine detail 3D reconstruction that cannot be obtained with only multiview stereo correspondences. However, due to the lack of suitable benchmark data with ground-truth shapes (normals), quantitative comparison and evaluation is difficult to achieve. Related to these approaches the corresponding databases have been generated and are made available.

Photometric Harvard Stereo Dataset

Photometric Harvard Stereo Dataset[21] provides data for normal and 3D surface reconstruction. Each object in the dataset is illuminated under 20 different directional lightings, which are calibrated with two chrome spheres. The lighting strength is estimated by a simple normalization on image intensities (99 percentile) followed by a nonlinear optimization. The albedo and normal vectors of the object are solved with a least squares system, and the depth map is integrated with the Frankot-Chellappa algorithm [19]. The reconstruction error is measured by re-rendering the estimated normal map into a shading image and comparing that with the actual captured one. The data, as well as the code for normal and surface reconstruction, are provided.

"DiLiGenT" Photometric Stereo Dataset

"DiLiGenT" Photometric Stereo Dataset[22] is photometric stereo image dataset provided with calibrated directional lightings, objects of general reflectance, and "ground-truth" shapes (normals) for orthographic projection and single-view setup. In addition to the first dataset for such a purpose, a photometric stereo taxonomy is provided as well, emphasizing on non-lambertian and uncalibrated methods. Based on the dataset, state-of-the-art photometric stereo methods are quantitatively evaluated for general non-lambertian materials and unknown lightings to analyze their strengths and limitations [20].

---

[20]Washington University Photo Tourism Dataset (http://phototour.cs.washington.edu/datasets/).

[21]Photometric Harvard Stereo Dataset (http://vision.seas.harvard.edu/qsfs/Data.html).

[22]'DiLiGenT' Photometric Stereo Dataset (https://sites.google.com/site/photometricstereodata/).

Standford Computer Vision and Geometry Lab Datasets

The 3D dataset for other more advanced computer vision applications such as multiview 3D reconstruction, registration, and recognition applications are provided by Stanford Computer Vision and Geometry Lab[23]: (1) PASCAL3D + dataset [21], which is a novel and challenging dataset for 3D object detection and pose estimation. PASCAL3D + augments 12 rigid categories of the PASCAL VOC 2012 [22] with 3D annotations. This dataset represents a rich test bed to study 3D detection and pose estimation; (2) Stanford 2D-3D-Semantics Dataset, 2D-3D-S[24] [23], provides a variety of mutually registered modalities from 2D, 2.5D, and 3D domains, with instance-level semantic and geometric annotations. It covers over 6000 m^2 and contains over 70,000 RGB images, along with the corresponding depths, surface normal, semantic annotations, global XYZ images (all in forms of both regular and 360° equirectangular images) as well as camera information. It also includes registered raw and semantically annotated 3D meshes and point clouds. The dataset enables development of joint and cross-modal learning models and potentially unsupervised approaches utilizing the regularities present in large-scale indoor spaces.

11.3 Characterization and Selection of Light-Field Content for Perceptual Assessment

Many efforts have been devoted to the design of image and video quality assessment methods. In order to evaluate the quality of processed images, to compare the performance of different algorithms, or to determine the quality criteria in system optimization, the availability of test data is of primary importance. The Source Sequences (SRCs) selection is not a trivial task, especially for special content, such as light-field. In fact, the quality, the dataset cardinality, and the content of the selected SRCs may affect the performance assessment.

Concerning the content, to be as general purpose as possible, SRCs should span a wide range of content typologies. To characterize image content, low-level and high-level features can be used. In particular, low-level features, such as spatial information, color information, and brightness are considered important parameters that help in measuring the distortions suffered by data compression or transmission over a bandwidth-limited channel.

Among others, Spatial Information (SI) [24], colorfulness (CF) [25], contrast, correlation, homogeneity, brightness, hue, and saturation are related to image quality attributes and Human Visual System (HVS) characteristics [26]. In more details, SI is a perceptual indicator of spatial information of a scene, colorfulness is

[23]Stanford Computer Vision and Geometry Lab (http://cvgl.stanford.edu/resources.html).

[24]Stanford 2D-3D-Semantics Dataset 2D-3D-S (http://buildingparser.stanford.edu/dataset.html).

a perceptual attribute tied with image quality and naturalness of the images, while contrast, color information, and brightness are features strictly related to HVS features.

Several SI filters have been proposed in the literature. In [27] a method based on long edge detection is presented. Separate horizontal and vertical filters are applied, and the total edge energy is computed as Euclidean distance. Similarly, an SI filter for video is presented in [1], as a perceptual indicator of the spatial information of the scene. It measures the amount of spatial details for each frame; the SI value is higher for spatially complex scenes. SI filter has been applied to LF data in [28]. The authors show that the correlation between the SI scores estimated by using the cited methods is very high, since both the methods exploit the classical Sobel filtering. Another approach based on the ITU recommendation [29] has been adopted. The luminance component of the image is first filtered by using a Sobel filter. Then, the standard deviation over the pixels in each filtered component is computed as SI.

Colorfulness and aesthetic are important visual features having a significant impact on the perceptual quality of a scene. In literature, many efforts have been devoted to study the color impact and its assessment. A CF metric for natural images is presented in [25] based on the distribution of image pixels in CIELab color space [30]. In [31], aesthetics (e.g., "the principles of the nature and the appreciation of beauty") in photographic images is addressed by exploiting several metrics, such as light, CF, saturation, hue, and texture, to understand the human emotions with respect to the visual content.

Dealing with LF images, the inner structure of the LF must be considered. A light-field camera provides information about depth dependence and Lambertian lighting. Depth dependence implies multiple depths of semitransparent objects and the Lambertian surface reflects light with the equal intensity in all directions [32].

The depth dependence information can be exploited during coding, and the variation in depth of field information could give different compression levels at the same quality level.

Reflections and transparency are prevalent in natural images, that is, reflected and transmitted lights are superimposed on each other. The image can be modeled as a linear combination of the transmitted layer, which contains the scene of interest, and a secondary layer, which contains the reflection or transparency [33, 34]. The decomposition of the images into two layers is an ill-posed problem in the absence of additional information about the scene [35]. The light-field camera recorded information, particularly multiple views of a single scene, can be exploited to solve the problem. Therefore, in a test dataset images with transparency, reflections, and wide Depth of Field (DoF) variation are needed.

Depth Properties

One of the main properties of LF imaging is the possibility of obtaining depth information of the captured scene, offering both horizontal and vertical parallax.

As observed with 3D content, depth properties are crucial for an appropriate description and characterization of LF content.

Depth map and depth histogram: Obtaining the depth information from data captured by the acquisition systems (e.g., camera arrays, plenoptic cameras, etc.) is a challenging issue as demonstrated by the number of different approaches that are being proposed to deal with this problem. Different approaches should be considered when dealing with sparse LFs (e.g., captured by camera arrays) and dense LFs (e.g., captured by plenoptic cameras), due to the different acquisition properties, such as baseline and spatial aliasing. On one hand, depth estimations from sparse LFs can be obtained by using traditional multiview methods [36], as well as some specific techniques, for instance, based on sweeping [37] or multi-resolution matching [38]. On the other hand, for dense LF in [39], a simple technique based on computing block-wise cross-correlation is proposed. More recently, approaches taking into account multiview stereo correspondences [37, 40] have been introduced.

Disparity range: While the majority of the methods mentioned above to estimate depth provide a normalized map here only relative disparities can be obtained, the absolute disparities in terms of pixels are more important for QoE aspects, such as content characterization [41], have been presented. However, to obtain a reliable content characterization, it is required the estimate of the depth range of the scene regarding distances to the nearest and furthest objects, or the camera calibration parameters. When these data are not available, estimation algorithms, such as the multiview stereo algorithm described in [42], could return pixel disparities. This algorithm has been used (over the subaperture images of the LF images) for characterizing LF data in [28].

Occlusions are one of the most important problems to deal with in-depth estimations for LFs. However, until now, only few depth estimation algorithms specifically manage and model occlusions, i.e., the occlusion model for depth map estimation in [43]. This algorithm is applied over the LF data structure, and the amount of possible occluded pixels are computed and considered in the content characterization [28].

Refocusing Features

As aforementioned, one of the main applications of LF images is the possibility of changing the focused elements of the images. Therefore, it is important to find appropriate descriptors that could help in the characterization of LF content, providing an estimation of their possible performance in this particular use case. One alternative is to analyze the properties and shape of the disparity histogram since it provides information about the distribution in depth of the elements of the scene.

Other possibilities to deal with the use of blur metrics, such as the technique proposed in [44] taking into account the perceived image quality induced by blur. In addition, some approaches have been proposed to measure focuses specifically in LF images, such as the Multifocal Scene Defocus Quality (MSDQ) metric, which quantifies the perceptual visual quality of rendering LF images [45].

Finally, as shown in [28], the refocusing range of the LF images can be computed using the "shift and sum" algorithm that is based on the digital refocusing approach proposed in [46], which reveals that refocused images can be obtained by adding shifted subaperture images. In particular, the refocusing range is determined by the slope parameters of the algorithm used to obtain images refocused on the nearest and the furthest elements of the scene.

Selected Light-field Datasets

Several LF datasets have been proposed in the literature. The main features are reported in Table 11.1. Stanford LF Archive [47] is widely used; however, the images are captured by using a multi-camera system including gantry, microscope, etc. Nowadays, different LF cameras have been realized [48], (e.g., Lytro, Lytro Illum, and Raytrix), thus allowing the consumers to exploit such a technology. Lytro Illum is the newer version of the Lytro plenoptic camera, characterized by increased resolution and processing capabilities, while Raytrix is a so-called focused plenoptic camera. As can be noticed, the dataset [47] is not sufficient to deal with new challenges, perceptual quality evaluation, performance testing for processing algorithms, etc., which arose with the advancement of the LF technology. Other recently proposed datasets listed in Table 1 have been designed for specific purposes and the images have been acquired mostly by the Lytro plenoptic camera. In the dataset [49], the Lytro Illum camera has been used. However, most of the images have similar features and motivations behind the particular image content selection have not been reported. In [48], a LF image dataset is proposed. The dataset creation methodology using Lytro Illum, description of LF images, and analysis of LF image content is tailored. The SRCs image content selection criteria is defined, a comprehensive LF image quality dataset is proposed and made freely available to the research community, a spatial information estimation metric is exploited, an analysis of the features of the proposed dataset is provided.

11.4 Special Point-Cloud and Holographic Content Datasets

The overview of publicly available 3D visual content datasets mentioned in the previous paragraphs is far from complete since the number of relevant databases is continuously growing. For completeness, it is important to note that there are datasets used recently in the frame of development and standardization of image and video compression techniques within the Joint Photographic Experts Group (JPEG)[25] committee and the Moving Picture Experts Group (MPEG).[26]

[25]Joint Photographic Experts Group (https://jpeg.org/index.html).

[26]Moving Picture Experts Group (https://mpeg.chiariglione.org/).

Table 11.1 Overview of light-field datasets with corresponding features

Dataset	Year	Purpose	Features	Acquisition devices	Depth map
GUC light-field face and iris database [71]	2016	Face and iris recognition	Two biometric image databases collected by using a Lytro camera on multiple faces and visible iris (112 subjects for faces and 55 subjects for eye pattern)	Lytro	No
Lytro dataset [72]	2015	Light-field Reconstruction	30 images, with indoor and outdoor, motion blur, long exposure time, and flat image	Lytro	No
EPFL light-field image dataset [49]	2015	General	118 Lytro images with different categories: buildings, landscapes, people, etc.	Lytro Illum	No
LCAV-31 [73]	2014	Object recognition	Light-field images of 31 object categories captured from ordinary household objects and designed for object recognition purpose	Lytro	No
Light-field saliency Dataset (LFSD) [74]	2014	Saliency map estimation	100 light-field images with 60 indoor scenes and 40 outdoor scenes	Lytro	Yes (estimated)
Synthetic light-field archive [75]	2013	General	Artificial light-field images including images with transparencies, occlusions, and reflections	Camera (artificial light field)	No
Light-field analysis [76]	2013	Depth map estimation	Seven Blender and Six Gantry images; however, images do not cover the wide range of natural scenes	Blender Software and Gantry device	Yes
Stanford Light-Field Archive [47]	2008	General	20 light fields sampled using a camera array, a gantry, and a light-field microscope.	Gantry, light-field microscope, and camera array	No
SMART [48]	2016	General	15 Lytro Illum images with different categories	Lytro Illum	No

Notable progress is being made in the frame of JPEG Pleno[27] [50], which intends to provide a standard framework to facilitate the capture, representation, and exchange of omnidirectional, depth-enhanced, point-cloud, light-field, and holographic imaging modalities. JPEG Pleno is planned to provide an efficient compression format that will guarantee the highest quality content representation with reasonable resource requirements.

The JPEG Pleno Database[28] contains images from multiple plenoptic imaging modalities, e.g., light-field, point-cloud, and holographic imaging. There are five point-cloud datasets in the JPEG Pleno Database, one light-field dataset, and two datasets of holographic images. The light-field dataset [49] is addressed in the previous paragraph, thus only point-cloud, and holographic JPEG Pleno datasets are overviewed below. There is also one additional 3D point-cloud dataset [51] included in the overview.

11.4.1 JPEG Pleno Database: Point-Cloud Datasets

8i Voxelized Full Bodies (8iVFB v2) dataset [52] contains dynamic voxelized point cloud, i.e., sequence of frames with sets of points constrained to lie on a regular 3D grid. The dataset includes four sequences named "longdress", "loot", "redandblack", and "soldier". The human subjects' full bodies are captured by 42 RGB cameras configured in 14 clusters, at 30 fps with 10 s length. One spatial resolution is provided for each sequence: a cube of $1024 \times 1024 \times 1024$ voxels. The attributes of an occupied voxel are the red, green, and blue components of the surface color.

There are upper bodies of five subjects captured in the Microsoft Voxelized Upper Bodies dataset, named "Andrew", "David", "Phil", "Ricardo", and "Sara". The capturing was done using four frontal RGBD cameras, at 30 fps, over a 7–10 s period for each. Two spatial resolutions are provided for each sequence: a cube of $512 \times 512 \times 512$ voxels and a cube of $1024 \times 1024 \times 1024$ voxels.

ScanLAB Projects acquired and provide two datasets, namely, the Science Museum Shipping Galleries point-cloud dataset and Biplane point-cloud dataset. For the first dataset, the Shipping Galleries at the Science Museum were 3D scanned before their decommissioning in 2012 by ScanLAB Projects. A total of 256 scans were taken of the space and its exhibits to create a digital model of over two billion precisely measured points. This digital replica has been used to create a virtual flythrough of the gallery spaces providing detailed narration about the key exhibits and artefacts. The second dataset, Biplane, consists of the scan of a Handley Page Gugnunc, wooden biplane from 1920s exhibited at the Science Museum, Wroughton.

[27]JPEG Pleno (https://jpeg.org/jpegpleno/index.html).

[28]JPEG Pleno Database (https://jpeg.org/plenodb/).

The GTI-UPM Point-cloud dataset includes a directory structure consisting of several 3D models (both point clouds and naked/textured meshes) reconstructed from 2D pictures by GTI-UPM within the activities of the EU-funded research project BRIDGET (BRIDging the Gap for Enhanced broadcasT).

11.4.2 JPEG Pleno Database: Holographic Datasets

There are two holographic datasets available, namely ERC Interfere Holograms (data set 1) and B-com Holograms. Holography allows for recording and reproduction of wavefields of light. It is able to fully capture the three-dimensional structure of objects. Holograms represent interference patterns and their signal properties are very different from natural photography and video. The Interfere[29] database [53] contains five computer generated holograms created from 2D and 3D objects using an algorithm capable of handling self-occlusion for 3D objects. B-com Holograms [54] were synthesized using the algorithms developed by the Institute of Research & Technology (IRT) b-com.[30]

Oakland 3D Point-Cloud Dataset

This repository[31] contains labeled 3D point-cloud laser data collected from a moving platform in an urban environment. This dataset was used to produce the results presented in [51]. The data was collected using Navlab11 equipped with side-looking SICK LMS laser scanners and used in push-broom. The data was collected around CMU campus in Oakland. Data are provided in ASCII format: x y z label confidence, one point per line, space as separator. Corresponding VRML files (*.wrl) and label counts (*.stats) are also provided. The dataset is made of two subsets (part2, part3) with each its own local reference frame, where each file contains 100,000 3D points. The training/validation and testing data was filtered and labeled remapped from 44 into five labels.

11.5 Datasets Annotated with Ratings from Subjective Experiments

This section describes selected publicly available datasets that have been annotated in a subjective study. Most of the studies result in Mean Opinion Scores (MOS) quantifying the quality of each stimulus in the set. Such databases are

[29]ERC-funded Interfere project (http://www.erc-interfere.eu/).

[30]b-com hologram repository (https://hologram-repository.labs.b-com.com).

[31]Oakland Dataset (http://www.cs.cmu.edu/∼vmr/datasets/oakland_3d/cvpr09/doc/).

essential for design and evaluation of objective quality metrics, described in the respective chapter in this book.

Since visual attention is of very high importance for understanding human perception in 3D applications, some effort has also been dedicated to track and record observers' gaze when exposed to the content. The datasets annotated with data from eye-tracking experiments are very useful for modeling perceptual mechanisms of the human visual system.

The described databases are further divided into image quality datasets, video quality datasets, 3D models quality datasets, and eye-tracking datasets.

11.5.1 3D Image Quality Databases

In the following paragraphs, there are six selected 3D image quality databases listed and described in detail.

IRCCyN/IVC 3D Images

The dataset[32] comprises six original stereoscopic images (with mean resolution of 512×448 pixels) and 90 distorted versions, annotated with respective Differential MOS (DMOS) values [55]. The used distortions include blur (Gaussian or downscale and upscale), JPEG, and JPEG2000 each on five different levels.

The subjective experiment was performed on 21″ Samsung SyncMaster 1100 MB display with 1024×768 pixels resolution and the frequency of 120 Hz. The viewing conditions were according to ITU-R Rec. BT.500 [56] and the viewing distance was set to four times the height of the images. The images were displayed in the center without upscaling. The observers were equipped with crystal shutter glasses.

There were 19 participants of sufficient visual acuity enrolled in the test. Their average age was 28.2. The images were evaluated using SAMVIQ [24] procedure in two sessions of 30 min per observer. The resulting DMOS scores range from 0 to 100.

LIVE 3D Image Quality Database Phase I

This database[33] contains 20 stereoscopic source images of 640×360 pixels [57]. From these scenes, 365 distorted images were created. 80 images were distorted by JPEG, 80 by JPEG2000, 80 by white noise, 80 by JPEG2000 transmitted over Rayleigh fading channel with various signal to noise ratio, and 45 by Gaussian blur. All the distortions are applied symmetrically, i.e., to both left and right image in each stereo pair.

[32]IRCCyN/IVC 3D Images dataset (http://ivc.univ-nantes.fr/en/databases/3D_Images/).

[33]LIVE 3D Image Quality Database Phase I (http://live.ece.utexas.edu/research/quality/live_3dimage_phase1.html).

A 22″ passive stereoscopic display IZ3D with the resolution set to 800 × 600 pixels was employed for subjective assessment. Each image was viewed by 17 subjects for 8 s and then assessed according to single stimulus continuous quality evaluation (SSCQE) procedure with hidden reference [56]. Two subjects were eliminated by outlier removal. The results are provided in the form of DMOS ranging from −10 to 100 (negative DMOS meaning an image evaluated better than reference).

LIVE 3D Image Quality Database Phase II

Despite the similarity in name and certain overlap in source content, this dataset [58][34] can be considered independent from the Phase I described above. Here, the distortions were applied both symmetrically and asymmetrically, and a different subjective study was conducted.

There are eight stereoscopic source images of 640 × 360 pixels and 360 distorted versions available. The applied distortions are similar to the Phase I, i.e., JPEG, JPEG2000, white Gaussian noise, Gaussian blur, and Rayleigh fading channel. From each combination of source image and distortion, three symmetrically, and six asymmetrically distorted stereo pairs were created.

The experiment was performed on 58″ Panasonic 3D television with active shutter glasses from the distance of 116 in., i.e., four times the screen height. 33 observers (22–42 years old) participated in the test which comprised of two 30 min long sessions. The procedure and data processing was the same as in case of Phase I described above.

MMSPG 3D Image Quality Assessment Database

The dataset[35] deals with the impact of distance of cameras during acquisition on the final stereoscopic image quality [59]. The set contains nine full HD (1920 × 1080) source scenes, each captured by cameras with six different distances, ranging from 10 to 60 cm.

In the test, the stereoscopic images were displayed on 46″ polarized stereoscopic full HD display Hyundai S465D. The viewing distance was three times the screen height, and the conditions were conforming to ITU-R Rec. BT.500 [56].

The content was assessed by 17 observers (22 to 53 years old, 30 on average). Single Stimulus (SS) methodology with five level discrete scale (Bad, Poor, Fair, Good, and Excellent) has been adopted. No outliers have been detected, thus the dataset provides raw scores from all of the observers, together with respective MOS and confidence intervals.

[34]LIVE 3D Image Quality Database Phase II (http://live.ece.utexas.edu/research/quality/live_3dimage_phase2.html).

[35]MMSPG 3D Image Quality Assessment Database (http://mmspg.epfl.ch/3diqa).

IRCCyN/IVC DIBR Images

This dataset,[36] described in detail in [60], focuses on depth image based rendering (DIBR). Three multiview sequences are considered—Book Arrival (1024 × 768, 16 cameras with 6.5 cm spacing), Lovebird1 (1024 × 768, 12 cameras with 3.5 cm spacing), and Newspaper (1024 × 768, nine cameras with 5 cm spacing). For each of them, four new viewpoints are generated using seven different algorithms thus obtaining 96 sequences in total. For the purpose of this study, a key-frame has been extracted from each of the sequences and compared to the others.

The results of two subjective experiments are available for the above-described images. In the first one, Absolute Category Rating (ACR) methodology [29] with five level discrete scale was used, while Pair Comparison (PC) procedure was adopted in the second. The conditions for the two tests were identical.

The content was displayed on a full HD TVLogic LVM401 W display. The viewing conditions were according to ITU-R Rec. BT.500 [56]. 43 subjects participated in both tests. Raw scores coming from both procedures are provided together with MOS, in case of ACR, and Thurstone-Moesteller scores [61] for PC methodology.

MCL-3D Database

The last image dataset to be described is MCL-3D [62] [37] and deals with DIBR as well. It is based on nine source scenes, provided in image plus depth form. Six of the scenes are in full HD (1920 × 1080) resolution, while the rest is in 1024 × 768. Six types of distortion, namely Gaussian blur, additive white noise, downsampling blur, JPEG, JPEG2000, and transmission errors, are applied on four different levels. Moreover, four types of rendering algorithms are employed. Overall, the dataset comprises of 693 stereoscopic pairs.

Pair comparison methodology has been adopted. The stimuli were displayed on 46.9″ LG 47LW5600 screen. The viewing distance was 3.2 meters, and each observer was given polarized glasses. The cubic function has been used to resize the images to fit the screen, and the gap between them was filled with gray pixels.

270 observers took part in the experiment in order to collect 30 opinion scores for each stimulus. The results were transformed into MOS ranging from 0 to 9.

11.5.2 3D Video Quality Databases

In the following paragraphs, there are three selected 3D video quality databases listed and described in detail.

[36]IRCCyN/IVC DIBR Images (http://ivc.univ-nantes.fr/en/databases/DIBR_Images/).

[37]MCL-3D Database (http://mcl.usc.edu/mcl-3d-database/).

IRCCyN/IVC NAMA3DS1

The video dataset described in [41][38] contains 10 progressive full HD stereo source sequences with 25 frames per second (fps) and 10 versions of each source symmetrically distorted by processing. The used algorithms include compression by H.264/AVC on three levels and by JPEG2000 on four levels, downsampling, sharpening, and a combination of downsampling and sharpening. Overall, there are 110 videos in the dataset. The duration of 99 sequences is 16 s while the other 11 videos are 11 s long.

The content was evaluated in the conditions defined by ITU-R Rec. BT.500 [56] on a 46″ full HD 50 Hz LCD Philips 46PFL9705H with shutter glasses from 172 cm which corresponds to three times the picture height.

ACR with hidden reference [29] was selected as an appropriate subjective procedure. 29 observers (12 females and 17 males of age between 18 and 63) participated in the study. One of the observers was eliminated by outlier removal. The publicly available data, therefore, include (apart from the video sequences) raw scores, MOS, and standard deviations computed from 28 observers.

MMSPG 3D Video Quality Assessment Database

Similarly to the previously described Image Quality Database [59],[39] MMSPG 3D Video Quality Database [63] studies the impact of the camera distance. The dataset comprises of six different source scenes captured by full HD cameras in six distances (10–50 cm) from each other with 25 fps.

The procedure and the conditions were similar to the experiment conducted for MMSPG 3D Image Quality Database. 20 subjects (24–37 years old, 27 on average) participated in the test, but three of them have been recognized as outliers by the post-screening procedure. The final analysis was, therefore, performed on data from 17 observers.

IRCCyN/IVC DIBR Videos

The database described in [64][40] is an extension of the previously introduced IRCCyN/IVC DIBR Image dataset [60]. The same three reference sequences have been used. The first one has 15 fps while the other two 30 fps. Apart from the three different unprocessed views, seven view interpolation algorithms were included in the test, together with three different levels of H.264/AVC compression applied on the first view. Altogether, the dataset contains 102 video sequences.

The display, the room, and the viewing conditions were the same as in the case of still images, however, only one study using ACR methodology was conducted. The resulting MOS values are obtained from 32 observers.

[38]IRCCyN/IVC NAMA3DS1 (http://ivc.univ-nantes.fr/en/databases/NAMA3DS1_COSPAD1/).

[39]MMSPG 3D Video Quality Assessment Database (http://mmspg.epfl.ch/cms/page-58395.html).

[40]IRCCyN/IVC DIBR Videos (http://ivc.univ-nantes.fr/en/databases/DIBR_Videos/).

11.5.3 3D Models Quality Databases

In the following paragraphs, there are two selected 3D models quality databases listed and described in detail.

LIRIS 3D Model Masking Database

The first dataset evaluating the quality of 3D models was described in [65].[41] There are four models included in the dataset. Three levels of noise and smoothing are applied to rough and intermediate areas, as well as to the whole model. The noise is also added to the smooth areas. Final set, therefore, contains four reference and 84 distorted objects.

First, the subjects were trained by showing the original and the worst cases. After that, the models (including the original) were shown sequentially, each for 20 s, and scored from 0 to 10 according to the perceived impairment (0 meaning no impairment). The observers were allowed to interact with the objects (rotation, scaling, and translation).

12 participants performed the test. The authors provide their raw scores, together with MOS and objective metrics values.

LIRIS/EPFL 3D Model General-Purpose Database

The second 3D models quality database [66][42] also include four original models, although two of them are different than in the previous dataset. Here, only three levels of noise are applied either on smooth or rough regions. This gives four original and 24 distorted objects in total.

In the subjective experiment, each reference object and all of its versions were displayed together. The observers rated each of the distorted models on the scale from 0 to 4 according to similarity to the original (4 meaning completely identical). The objects were displayed for 3 min and participants could interact with them (rotation, scaling, and translation).

The study was carried out with 11 observers. With the dataset, raw scores and MOS are made publicly available, as well as some objective perceptual metrics values [67].

11.5.4 Eye-Tracking 3D Databases

In the following paragraphs, there are three selected eye-tracking 3D databases listed and described in detail.

[41]LIRIS 3D Model Masking Database (http://liris.cnrs.fr/guillaume.lavoue/data/datasets.html).

[42]LIRIS/EPFL 3D Model General-Purpose Database (http://liris.cnrs.fr/guillaume.lavoue/data/datasets.html).

IRCCyN/IVC 3D Gaze

The 3D Gaze dataset [68][43] includes 18 stereo images (provided as two 2D images in png format). 10 of them are obtained from the Middlebury stereo database,[44] while the rest were obtained by the authors. The dataset focuses on the influence of content features on the visual attention deployment, therefore no distortions are added, and the observers are given a free viewing task (i.e., they were instructed to freely observe the images without any particular task).

The images were displayed on a Panasonic BT-3DL2550 polarized screen with the frequency of 60 Hz and full HD resolution. SMI Hi-Speed eye-tracker was used in binocular mode. The acquisition frequency was 500 Hz. The viewing conditions were according to the ITU-R Rec. BT.500 [56] and the viewing distance was set to three times the screen height.

35 observers between 18 and 46 years old (24.23 on average) participated in the eye-tracking experiment. Raw data from the eye-tracker for each observer are provided, along with fixation density maps, the original stereoscopic pairs, depth maps, and disparity maps. Additional information and all the files for download are publicly available.

EyeC3D: 3D Video Eye-tracking Dataset

Unlike the previous database, EyeC3D [69][45] provides the visual attention information in videos. Eight stereo sequences of 8–10 s were watched in a free viewing task. 46″ polarized stereoscopic full HD display Hyundai S465D was used together with Smart Eye Pro 5.8 eye-tracker with the accuracy less than 0.5 degrees and sampling frequency of 60 Hz.

21 subjects participated in the test (18–31 years old with average of 21.8). Each sequence was watched twice by every observer. Fixation density maps were computed for each frame. The database also provides a list of all fixation points.

IRCCyN/IVC Eye-tracking Database for Stereoscopic Videos

The last dataset to be described is also dealing with task-free visual attention in stereoscopic videos [70].[46] It is also much larger than the previously described eye-tracking datasets with 47 stereo sequences composed by two 2D videos merged on a side by side avi files. Disparity map for each frame is also provided.

Panasonic BT-3DL2550 polarized screen with frequency of 60 Hz and full HD resolution was employed along with SMI RED binocular eye-tracker with acquisition frequency of 50 Hz. The room conditions were compliant with ITU-R Rec. BT.500 [56] and the viewing distance was set to three times the screen height.

---

[43]IRCCyN/IVC 3D Gaze (http://ivc.univ-nantes.fr/en/databases/3D_Gaze/).

[44]Middlebury stereo database (http://vision.middlebury.edu/stereo/data/).

[45]EyeC3D: 3D Video Eye-tracking Dataset (http://mmspg.epfl.ch/eyec3d).

[46]IRCCyN/IVC Eye-tracking Database for Stereoscopic Videos (http://ivc.univ-nantes.fr/en/databases/Eyetracking_For_Stereoscopic_Videos/).

The duration of one session was approximately 20 min. 40 observers (19–44 years old with average of 26) took part in the experiment. Fixation density map for each frame is provided as a png image. Higher pixel value (i.e., more white) means higher visual saliency.

11.6 Conclusions

This chapter presents an overview of datasets, their categorization, creation, and typical applications in development and performance evaluation of methods for processing, coding, transmission, and quality assessment of 3D visual content. As for the content types, stereoscopic and multiview image and video content datasets are introduced. Then, light-field content characterization and selection is addressed. Also, selected special point-cloud and holographic content datasets are reviewed. The most popular datasets annotated with ratings from subjective experiments are presented. Development of publicly available 3D visual content datasets, recently including also special visual content, e.g., point cloud and holographic, was largely promoted also by the standardization bodies, namely JPEG and MPEG. Datasets used within selected standardization efforts are also described in this chapter.

The aim is not to provide an exhaustive listing and description of all existing 3D visual content datasets, but more to give examples of the most commonly used publicly available datasets. Any effort of this type captures the current status. However, numerous new datasets are introduced every year. It is related to the fact that novel techniques for coding, transmission, and quality assessment are being continuously developed. Description of the most recent datasets can be found in regularly updated online resources, e.g., Qualinet Databases, which were also presented in this chapter.

References

1. Winkler, S.: Analysis of public image and video databases for quality assessment. IEEE J. Sel. Top. Signal Process. **6**(6), 616–625 (2012). https://doi.org/10.1109/JSTSP.2012.2215007
2. Winkler, S., Savoy, F.M., Subramanian, R. X-Eye: a reference format for eye tracking data to facilitate analyses across databases. In: Proceedings of IS&T/SPIE Human Vision & Electronic Imaging (2014). https://doi.org/10.1117/12.2042433
3. Winkler, S., Subramanian, R. Overview of eye tracking datasets. In: Proceedings of 5th International Workshop on Quality of Multimedia Experience (QoMEX) (2013). https://doi.org/10.1109/qomex.2013.6603239
4. Schöps, T., Schönberger, J., Galliani, S., Sattler, T., Schindler, K., Pollefeys, M., Geiger, A.: A multi-view stereo benchmark with high-resolution images and multi-camera videos. In: Proceedings of IEEE Computer Conference on Computer Vision and Pattern Recognition 2538–2547 (2017). https://doi.org/10.1109/CVPR.2017.272

5. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? The KITTI vision benchmark suite. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 3354–3361 (2012). https://doi.org/10.1109/cvpr.2012.6248074
6. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. Comput. Vis. **47**(1), 7–42 (2002). https://doi.org/10.1023/A:1014573219977
7. Butler, D., Wulff, J., Stanley, G., Black, M.: A naturalistic open source movie for optical flow evaluation. In: Proceedings of the European Conference on Computer Vision, pp. 611–625 (2012). https://doi.org/10.1007/978-3-642-33783-3_44
8. Johnson-Roberson, M., Barto, C., Rounak, M., Sharath, N., Ram, V.: Driving in the matrix: can virtual worlds replace human-generated annotations for real world tasks? In: Proceedings of IEEE International Conference on Robotics and Automation (2017). https://doi.org/10.1109/icra.2017.7989092
9. Menze, M., Geiger, A.: Object scene flow for autonomous vehicles. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015). https://doi.org/10.1109/cvpr.2015.7298925
10. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multi-view stereopsis. IEEE Trans. Pattern Anal. Mach. Intell. **32**(8), 1362–1376 (2010). https://doi.org/10.1109/TPAMI.2009.161
11. Seitz, S., Curless, B., Diebel, J., Scharstein, S., Szeliski, R.: A Comparison and evaluation of multi-view stereo reconstruction algorithms. In: Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR) (2006). https://doi.org/10.1109/cvpr.2006.19
12. Strecha, C., von Hansen, W., Van Gool, L., Fua, P., Thoennessen, U.: On benchmarking camera calibration and multi-view stereo for high resolution imagery. In: Proc IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2008). https://doi.org/10.1109/cvpr.2008.4587706
13. Fransens, R., Strecha, C., Van Gool, L.: Parametric stereo for multi-pose face recognition and 3D-face modeling (2005). In: Proceedings of ICCV 2005 Workshop Analysis and Modeling of Faces and Gestures, vol. 3723, pp. 109–124 (2005). https://doi.org/10.1007/11564386_10
14. Strecha, C., Fransens, R., Van Gool, L.: Wide-baseline stereo from multiple views: a probabilistic account. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2004)
15. Strecha, C., Fransens, R., Van Gool, L.: Combined depth and outlier estimation in multi-view stereo. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2006). https://doi.org/10.1109/cvpr.2006.78
16. Ozuysal, M., Lepetit, V., Fua, P.: Pose estimation for category specific multiview object localization. In: Proceedings of Conference on Computer Vision and Pattern Recognition (2009). https://doi.org/10.1109/cvprw.2009.5206633
17. Cremers, D., Kolev, K.: Multiview stereo and silhouette consistency via convex functionals over convex domains. IEEE Trans. Pattern Anal. Mach. Intell. **33**(6), 1161–1174 (2011). https://doi.org/10.1109/TPAMI.2010.174
18. Li, Y., Snavely, N., Huttenlocher, D.P.: Location recognition using prioritized feature matching. In: Proceedings of ECCV (2010). https://doi.org/10.1007/978-3-642-15552-9_57
19. Frankot, R., Chellappa, R.: A method for enforcing integrability in shape from shading algorithms. IEEE Trans. Pattern Anal. Mach. Intell. **10**(4), 439–451 (1988). https://doi.org/10.1109/34.3909
20. Shi, B., Wu, Z., Mo, Z., Duan, D., Yeung, S.-K., Tan, P.: A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
21. Xiang, Y., Mottaghi, R., Savarese, S.: Beyond PASCAL: a benchmark for 3D object detection in the wild. In: Proceedings of IEEE Winter Conference on Applications of Computer Vision (WACV) (2014). https://doi.org/10.1109/wacv.2014.6836101

22. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The pascal visual object classes (VOC) challenge. Int. J. Comput. Vis. **88**(2), 303–338 (2010). https://doi.org/10.1007/s11263-009-0275-4
23. Armeni, I., Sax, A., Zamir, A., Savarese, S. Joint 2D-3D-semantic data for indoor scene understanding. In: Computer Vision and Pattern Recognition (2017, to appear)
24. ITU-R Recommendation BT.1788: Methodology for the subjective assessment of video quality in multimedia applications, Jan 2007
25. Hasler, D., Suesstrunk, S.E.: Measuring colorfulness in natural images. Hum. Vis. Electron. Imaging VIII **2003**, 87–95 (2003). https://doi.org/10.1117/12.477378
26. Faloutsos, C., Barber, R., Flickner, M., Hafner, J., Niblack, W., Petkovic, D., Equitz, W.: Efficient and effective querying by image content. J. Intell. Inf. Syst. **3**(3–4), 231–262 (1994). https://doi.org/10.1007/BF00962238
27. Pinson, M.: Spatial information (SI) filter (2016). https://www.its.bldrdoc.gov/resources/video-quality-research/guides-and-tutorials/spatial-information-si-filter.aspx. Accessed 29 Sept 2017
28. Paudyal, P., Gutiérrez. J., Le Callet, P., Carli, M., Battisti, F.: Characterization and selection of light field content for perceptual assessment. In: Proceedings of QoMEX (2017). https://doi.org/10.1109/qomex.2017.7965635
29. ITU-T Recommendation P.910: Subjective video quality assessment methods for multimedia applications, Apr 2008
30. Tkalcic, M., Tasic, J.F.: Colour spaces: perceptual, historical and applicational background. In: Proceedings of EUROCON (2003). https://doi.org/10.1109/eurcon.2003.1248032
31. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Studying aesthetics in photographic images using a computational approach. Proc. Eur. Conf. Comput. Vis. **3953**, 288–301 (2006). https://doi.org/10.1007/11744078_23
32. Bishop, T.E., Zanetti, S., Favaro, P.: Light field superresolution. In: Proceedings of IEEE International Conference on Computational Photography (ICCP 09) (2009). https://doi.org/10.1109/iccphot.2009.5559010
33. Szeliski, R., Avidan, S., Anandan, P.: Layer extraction from multiple images containing reflections and transparency. In: IEEE Conference on Computer Vision and Pattern Recognition (2000)
34. Wang, Q., Lin, H., Ma, Y., Kang, S.B., Yu, J.: Automatic layer separation using light field imaging, arXiv preprint (2015). arXiv:1506.04721
35. Levin, A., Weiss, Y.: User assisted separation of reflections from a single image using a sparsity prior. IEEE Trans. Pattern Anal. Mach. Intell. **29**(9), 1647–1654 (2007). https://doi.org/10.1109/TPAMI.2007.1106
36. Denker, K., Umlauf, G.: Accurate real-time multi-camera stereo-matching on the GPU for 3D reconstruction. J. WSCG **19**(1–3), 9–16 (2011)
37. Jeon, H.-G., Park, J., Choe, G., Park, J., Bok, Y., Tai, Y.W., Kweon, I.S.: Accurate depth map estimation from a lenslet light field camera. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015). https://doi.org/10.1109/cvpr.2015.7298762
38. Dabała, Ł., Ziegler, M., Didyk, P., Zilly, F., Keinert, J., Myszkowski, K., Seidel, H.-P., Rokita, P., Ritschel, T.: Efficient multi-image correspondences for on-line light field video processing. Comput. Graph. Forum **35**(7), 401–410 (2016). https://doi.org/10.1111/cgf.13037
39. Montilla, I., Puga, M., Luke, J.P., Marichal-Hernandez, J.G., Rodriguez-Ramos, J.M.: Design and laboratory results of a plenoptic objective: from 2D to 3D with a standard camera. J. Disp. Technol. **11**(1), 73–78 (2015). https://doi.org/10.1109/JDT.2014.2361257
40. Ziegler, M., Engelhardt, A., Müller, S., Keinert, J., Zilly, F., Foessel, S., Schmid, K.: Multi-camera system for depth based visual effects and compositing. In: Proceedings of European Conference on Visual Media Production (2015). https://doi.org/10.1145/2824840.2824845
41. Urvoy, M., Barkowsky, M., Cousseau, R., Koudota, Y., Ricordel, V., Le Callet, P., Gutierrez, J., Garcia, N.: NAMA3DS1-COSPAD1: subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences.

In: Proceedings of Fourth International Workshop on Quality of Multimedia Experience (2012). https://doi.org/10.1109/qomex.2012.6263847
42. Wang, Z.: Objective image quality assessment: facing the real-world challenges. In: Image Quality and System Performance (keynote speech paper) (2016)
43. Wang, T.-C., Efros, A.A., Ramamoorthi, R.: Depth estimation with occlusion modeling using light-field cameras. IEEE Trans. Pattern Anal. Mach. Intell. **38**(11), 2170–2181 (2016). https://doi.org/10.1109/tpami.2016.2515615
44. Liu, H., Wang, J., Redi, J., Le Callet, P., Heynderickx, I.: An efficient no-reference metric for perceived blur. In: Proceedings of European Workshop on Visual Information Processing (2011). https://doi.org/10.1109/euvip.2011.6045525
45. Wu, W., Llull, P., Tosic, I., Bedard, N., Berkner, K., Balram, N.: Content-adaptive focus configuration for near-eye multi-focal displays. In: Proceedings of IEEE International Conference on Multimedia and Expo (2016). https://doi.org/10.1109/icme.2016.7552965
46. Ng, R., Levoy, M., Brédif, M., Duval, G., Horowitz, M., Hanrahan, P.: Light field photography with a hand-held plenoptic camera. Comput. Sci. Techn. Rep. **2**(11), 1–11 (2005)
47. Vaish, V., Adams, A.: The (new) stanford light field archive. http://lightfield.stanford.edu/ (2008). Accessed 29 Sept 2017
48. Paudyal, P., Olsson, R., Sjostrom, M., Battisti, F., Carli, M.: SMART: a light field image quality dataset. In: Proceedings of International Conference on Multimedia Systems (MMSys 2016) (2016). https://doi.org/10.1145/2910017.2910623
49. Rerabek, M., Yuan, L., Authier, L.A., Ebrahimi, T.: EPFL light-field image dataset. ISO/IEC JTC 1/SC 29/WG1, Technical Report (2015)
50. Ebrahimi, T., Foessel, S., Pereira, F., Schelkens, P.: JPEG pleno: toward an efficient representation of visual reality. IEEE Multimed. **23**(4), 14–20 (2016). https://doi.org/10.1109/MMUL.2016.64
51. Munoz, D., Bagnell. J.A., Vandapel, N., Hebert, M.: Contextual classification with functional max-margin markov networks. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) (2009). https://doi.org/10.1109/cvprw.2009.5206590
52. d'Eon, E., Harrison, B., Myers, T., Chou, P.A.: 8i voxelized full bodies—a voxelized point cloud dataset. ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document WG11M40059/WG1M74006, Geneva, January (2017)
53. Blinder, D., Ahar, A., Symeonidou, A., Xing, Y., Bruylants, T., Schretter, C., Pesquet-Popescu, B., Dufaux, F., Munteanu, A., Schelkens, P.: Open access database for experimental validations of holographic compression engines. In: 7th International Workshop on Quality of Multimedia Experience (QoMEX) (2015). https://doi.org/10.1109/qomex.2015.7148145
54. Gilles, A., Gioia, P., Cozot, R., Morin, L.: Hybrid approach for fast occlusion processing in computer-generated hologram calculation. Appl. Opt. **55**(20), 5459–5470 (2016). https://doi.org/10.1364/AO.55.005459
55. Benoit, A., Le Callet, P., Campisi, P., Cousseau, R.: Quality assessment of stereoscopic images. EURASIP J. Image Video Process. (2008). https://doi.org/10.1155/2008/659024
56. ITU-R Recommendation BT.500–13; Methodology for the subjective assessment of the quality of television pictures, Jan 2012
57. Moorthy, A.K., Su, C.-C., Mittal, A., Bovik, A.C.: Subjective evaluation of stereoscopic image quality. Signal Process. Image Commun. **28**(8), 870–883 (2013). https://doi.org/10.1016/j.image.2012.08.004
58. Chen, M.-J., Cormack, L.K., Bovik, A.C.: No-reference quality assessment of natural stereopairs. IEEE Trans. Image Process. **22**(9), 3379–3391 (2013). https://doi.org/10.1109/TIP.2013.2267393

59. Goldmann, L., De Simone, F., Ebrahimi, T.: A comprehensive database and subjective evaluation methodology for quality of experience in stereoscopic video. In: Proceedings of Electronic Imaging (EI), 3D Image Processing (3DIP) and Applications (2010). https://doi.org/10.1117/12.839438
60. Bosc, E., Pépion, R., Le Callet, P., Köppel, M., Ndjiki-Nya, P., Pressigout, M., Morin, L.: Towards a new quality metric for 3-D synthesized view assessment. IEEE J. Sel. Top. Signal Process. **6029277**, 1332–1343 (2011). https://doi.org/10.1109/JSTSP.2011.2166245
61. Thurstone, L.L.: A law of comparative judgement. Psychol. Rev. **34**(4), 273–286 (1927). https://doi.org/10.1037/h0070288
62. Song, R., Ko, H., Kuo, C.C.: MCL-3D: a database for stereoscopic image quality assessment using 2D-image-plus-depth source. J. Inf. Sci. Eng. **31**(5), 1593–1611 (2015)
63. Goldmann, L., De Simone, F., Ebrahimi, T.: Impact of acquisition distortions on the quality of stereoscopic images. In: Proceedings of 5th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM) (2010)
64. Bosc, E., Le Callet, P., Morin, L., Pressigout, M.: Visual quality assessment of synthesized views in the context of 3D-TV. In: Zhu, C., Zhao, Y., Yu, L., Tanimoto, M. (eds) 3D-TV System with Depth-Image-Based Rendering Architectures, Techniques and Challenges. Springer, New York (2012). https://doi.org/10.1007/978-1-4419-9964-1_15
65. Lavoué, G., Drelie Gelasca, E., Dupont, F., Baskurt, A., Ebrahimi, T.: Perceptually driven 3D distance metrics with application to watermarking (2006). In: Proceedings of SPIE, vol. 6312. https://doi.org/10.1117/12.686964
66. Lavoué, G.: A local roughness measure for 3D meshes and its application to visual masking. ACM Trans. Appl. Percept. **5**(4) (2009). https://doi.org/10.1145/1462048.1462052
67. Lavoué, G., Corsini, M.: A comparison of perceptually-based metrics for objective evaluation of geometry processing. IEEE Trans. Multimed. **12**(7), 636–649 (2010). https://doi.org/10.1109/TMM.2010.2060475
68. Wang, J., Perreira Da Silva, M., Le Callet, P., Ricordel, V.: Computational model of stereoscopic 3D visual saliency. IEEE Trans. Image Process. **22**(6), 2151–2165 (2013). https://doi.org/10.1109/TIP.2013.2246176
69. Hanhart, P., Ebrahimi, T.: EyeC3D: 3D video eye tracking dataset. In: Proceedings of Sixth International Workshop on Quality of Multimedia Experience (QoMEX 2014) (2014). https://doi.org/10.1109/qomex.2014.6982290
70. Fang, Y., Wang, J., Li, J., Pépion, R., Le Callet, P.: An eye tracking database for stereoscopic video. In: Proceedings of Sixth International Workshop on Quality of Multimedia Experience (QoMEX 2014) (2014). https://doi.org/10.1109/qomex.2014.6982288
71. Raghavendra, R., Raja, K., Busch, C.: Exploring the usefulness of light field camera for biometrics: an empirical study on face and iris recognition. IEEE Trans. Inf. Forensics Secur. **11**(5), 922–936 (2016). https://doi.org/10.1109/tifs.2015.2512559
72. Mousnier, A., Vural, E., Guillemot, C.: Partial light field tomographic reconstruction from a fixed-camera focal stack. In: Computer Vision and Pattern Recognition, arXiv preprint (2015). arXiv:1503.01903
73. Ghasemi, A., Afonso. N., Vetterli, M.: LCAV-31: a dataset for light field object recognition. In IS&T/SPIE Electronic Imaging, pp. 902014–902014 (2014). https://doi.org/10.1117/12.2041097
74. Li, N., Ye, J., Ji, Y., Ling, H., Yu, J.: Saliency detection on light field. IEEE Trans. Pattern Anal. Mach. Intell. **39**(8), 1605–1616 (2017). https://doi.org/10.1109/TPAMI.2016.2610425
75. Wetzstein, G.: Synthetic light field archive (2016). http://web.media.mit.edu/~gordonw/SyntheticLightFields/, Accessed 29 September 2017
76. Wanner, S., Meister, S., Goldluecke, B.: Datasets and benchmarks for densely sampled 4D light fields. In: Annual Workshop on Vision, Modeling and Visualization: VMV (2013). https://doi.org/10.2312/pe.vmv.vmv13.225-226