Annotated 360-Degree Image and Video Databases: A Comprehensive Survey

Majed Elwardy, Hans-Jürgen Zepernick, and Veronica Sundstedt
Blekinge Institute of Technology, SE-37179 Karlskrona, Sweden
E-mail: {majed.elwardy}@bth.se

Introduction

- The advances in 5G mobile networks are expected to enable immersive interconnected mobile multimedia systems.
- As humans are the final judges of the quality of immersive multimedia, it is essential to engage a suitable ground truth in the design of such systems.
- Databases annotated with results from subjective tests constitute such ground truth given as opinion scores, head movements, eye tracking data, psychophysiological data, and data related to the viewers' behavior.
- On this basis, a comprehensive survey of publicly available annotated 360° image and video databases is provided.

Discussion

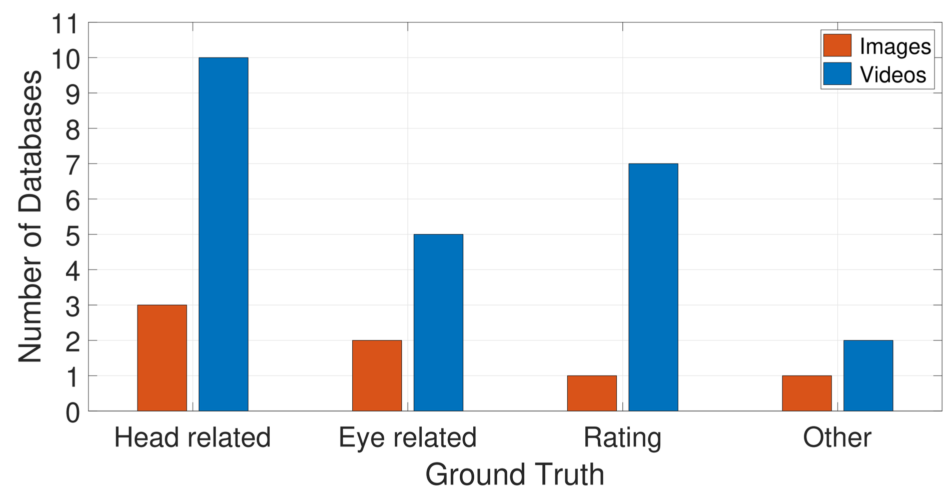

Figure 1: Number of databases with respect to ground truth classes.

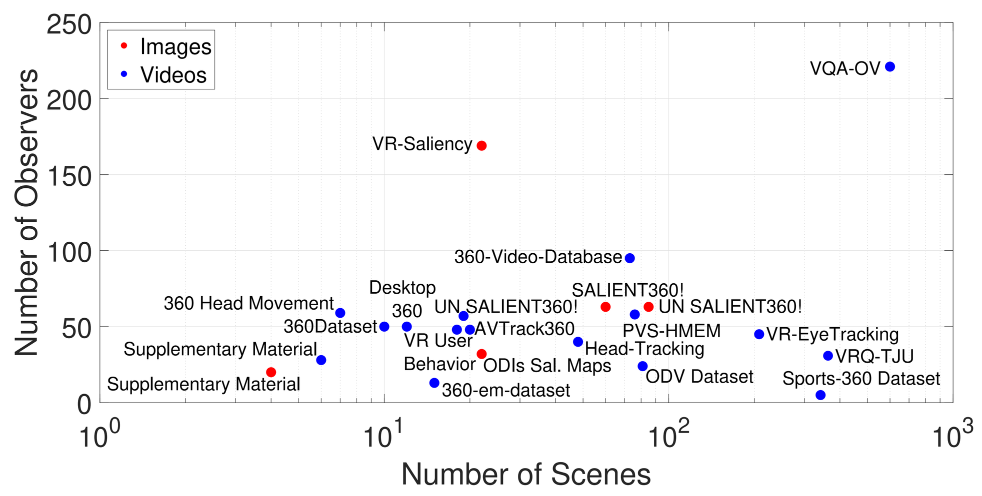

Figure 2: Scatter plot of number of observes versus number of scenes.

- Table 1 reveals that most of the databases focus on 360° videos as compared to 360° images.
- Figure 1 provides the number of databases with respect to four ground truth classes.
- Figure 2 shows a scatter plot illustrating the trade-off between the number of observers and the number of scenes.

Conclusions

The paper has given a comprehensive survey of publicly available annotated 360° image and video databases in terms of:

- Test material, subjects, ground truth, and purpose.
- The equipment used in the subjective tests.
- The execution of the subjective tests.

This survey may serve as a resource for selecting suitable ground truth for algorithm development and benchmarking, quality assessment and modeling tasks.

- Public ground truth in terms of opinion scores is still rare.

Acknowledgments

The work was supported by The Knowledge Foundation, Sweden, through the ViaTecH project (Contract 20170056).

Databases

Table 1: Comprehensive Survey of 360° Image and Video Databases

	Database [Year]	Scenes	Resolution [PX]	Observers	Age	Categories	Ground Truth	Purpose
Images	**SUN360** (2012)	67583	1024 × 512 9104 × 4552	—	—	80	Place category labels of panoramas	Recognize place category and viewpoint in scenes
	SALIENT360! (2017)	60	5376 × 2688 18332 × 9166	63 40-42/stimuli (24 F)	19-52 $\mu = 30$	5	Eye and head movement Scan-path data Saliency maps	Study visual attention
	Supplementary Material (2017)	4	3840 × 1920 7500 × 3750	20	22-35	—	Opinion scores	Maximize immersion and engagement for 360° images
	ODIs Saliency Maps (2017)	22	4096 × 2048	32 (16 in each group)	—	—	Viewpoint center trajectories Viewport limits	Estimate saliency maps for omnidirectional images without eye tracking
	VR-Saliency (2018)	22	8192 × 4096	169 (39 F)	17-55	—	Gaze direction Head orientation	Understand viewing behaviour and saliency in VR
Videos	**360-Video-Database** (2017)	73 (29-668 s) ($\mu = 118$ s)	—	95 (56 F)	18-24 Undergrads	19	Rating of arousal and valence Rotational head movements	Study emotions in VR
	360 Head Movement Dataset (2017)	7 (70 s) (25-60 fps)	3840 × 2048	59 (12 F) (36 FT)	min = 6 max = 62 $\mu = 34.12$	—	Head movement (Navigation)	Analyze users' navigation patterns
	VR User Behavior Dataset (2017)	18 (164-655 s)	1920 × 1080	48 (24 F)	10: ≤ 20 31: 20-26 7: ≥ 26	5	Head motions Head rotations Head positions	Explore user behavior in VR to improve applications
	Head-Tracking (2017)	48 (20-60 s)	2880 × 1440 up to 7680 × 3840	40 (11 F)	—	8	Head movement MOS DMOS	Develop subjective video quality assessment methods for quality loss of impaired videos
	360Dataset (2017)	10 (60 s) (30 fps)	4K	50 (24 F) (28 FT)	20-48 most are early 20's	3	Head positions Head orientations Saliency and motion maps	Support optimizing and developing 360° video applications
	Sports-360 dataset (2017)	342	—	5	—	5	Viewing angle for objects	Train and evaluate a deep 360° pilot agent that captures the objects of interest
	Supplementary Material (2017)	6 (60-65 s) (25-30 fps)	4K FHD	28 (15 F)	20-38 $\mu = 26.25$ $Md = 25$	3	Head rotations Opinion scores SSQ	Evaluate and compare the integral quality of VR contents of two different HMDs
	PVS-HMEM (2017)	76 (10-80 s) ($\mu = 26.9$ s)	3K 8K	58 (17 F)	18-36	8	Head movement Eye movement (Heat maps, scan-path data)	Predict head movement in panoramic video
	VR-EyeTracking (2018)	208 (20-60 s) (25 fps)	4K 3840 PX width	45 (20 F)	20-24	6	Eye movement	Explore gaze prediction in dynamic contents based on history scan-path and content
	Desktop360 (2018)	12 (60-120 s)	2048 × 1024 3840 × 2048	50	20-35	5	Yaw trajectory Pitch trajectory	Study viewers navigation behaviors when watching VR videos
	Supplementary Material (2017)	6 (60-65 s) (25-30 fps)	4K FHD	28 (15 F)	20-38 $\mu = 26.25$ $Md = 25$	3	Head rotations Opinion scores SSQ	Evaluate and compare the integral quality of VR contents of two different HMDs
	PVS-HMEM (2017)	76 (10-80 s) ($\mu = 26.9$ s)	3K 8K	58 (17 F)	18-36	8	Head movement Eye movement (Heat maps, scan-path data)	Predict head movement in panoramic video
	VR-EyeTracking (2018)	208 (20-60 s) (25 fps)	4K 3840 PX width	45 (20 F)	20-24	6	Eye movement	Explore gaze prediction in dynamic contents based on history scan-path and content
	Desktop360 (2018)	12 (60-120 s)	2048 × 1024 3840 × 2048	50	20-35	5	Yaw trajectory Pitch trajectory	Study viewers navigation behaviors when watching VR videos
	AVTrack360 (2018)	20 (30 s)	3840 × 1920	48 (25 F)	18-65 $\mu = 27.9$ $Md = 25$	—	Head movement SSQ	Develop a framework to evaluate head rotation data
	VQA-ODV (2018)	600 (10-23 s)	3840 × 1920 7680 × 3840	221 (78 F)	19-35	10	MOS Head movement Eye movement	Develop an objective quality metric that takes HM and EM data into account
	VRQ-TJU (2018)	377 (30 fps)	2560 × 2560	30 (15 F)	—	—	MOS	Develop an objective quality metric using convolutional neural networks
	360_em_dataset (2019)	15 (43-85 s)	3840 × 1920	13	—	—	Eye movement Gaze recordings	Gaze behaviour analysis Develop gaze event detection algorithms
	ODV Dataset (2019)	81 (10 s)	2023 × 1016 3600 × 1800 8128 × 4064	24 (4 F)	22-38 $\mu = 29.71$	—	Opinion scores	Develop objective quality metrics for omnidirectional video
Both	**UN SALIENT360!** (2018)	Training 85 img/19 vid Benchmarking 26 img/21 vid (20 s) (24-30 fps)	3840 × 1920	63/img (24 F) 57/vid (25 F)	19-52/img $\mu = 30$ 19-44/vid $\mu = 25.7$	5	Head movement Eye movement Saliency maps Scan-path data	Support research on visual attention for 360°images and videos and its application to algorithm development and quality assessment

(μ=mean, Md=median, SD=standard deviation, F=female, FT=first time, s=second, PX=pixel, fps=frames per second)

Table 2: Equipment Used in the Subjective Tests

	Database	HMD	Resolution [PX]	f [Hz]	FOV	Sensors/Devices	Software/Platform
Images	**SALIENT 360!**	Oculus DK2 driver version 2	960 × 1080 per eye	75	100°	SMI eye tracker (60 Hz) Gyroscope (75 Hz)	Unity
	ODIs Saliency Maps	Oculus Rift DK2	960 × 1080 per eye	75	100°	Head tracking	WebVR, ThreeJS APIs for testbed design
	VR-Saliency	Oculus Rift DK2	960 × 1080 per eye	75	100°	Pupil-labs eye tracker (120 Hz)	Unity game engine for testbed design
Videos	**360-Video-Database**	Oculus Rift CV1	1080 × 1200 per eye	90	110°	Oculus Rift remote, Gyroscope Magnetometer, Accelerometer	Vizard 5 (rating system) NVIDIA GTX 1080 GPU
	360 Head Movement dataset	Razer OSVR HDK2	960 × 1080 per eye	60	100°	Accelerometer, Gyroscope and Compass	Developed video player capturing head movement
	VR User Behavior Dataset	HTC Vive	1080 × 1200 per eye	90	110°	Two tracking stations Handheld controllers	Developed a Unity3D program
	Head-Tracking	HTC Vive	1080 × 1200 per eye	90	110°	—	Virtual desktop as panoramic video player
	360Dataset	Oculus Rift DK2	960 × 1080 per eye	75	100°	Open track (head tracking tool) records viewers orientations	SDK, Oculus video player NVIDIA GTX 970 GPU
	Supplementary Material	HTC Vive Oculus Rift CV1	1080 × 1200 per eye	90	110°	—	Whirligig player NVIDIA GTX980
	VR-EyeTracking	HTC Vive	1080 × 1200 per eye	90	110°	'7invensun a-Glass' eye tracker	Unity game engine NVIDIA Tesla M40 GPU
	AVTrack360	HTC Vive	1080 × 1200 per eye	90	110°	—	Whirligig player GTX1080 Graphics card
	VQA-ODV	HTC Vive	1080 × 1200 per eye	90	110°	Eye-tracking module, aGlass DKI	High-performance computer
	360_em_dataset	FOVE	2560 × 1440 per eye	70	100°	120 Hz eye tracker SteamVR	—
	ODV Dataset	HTC Vice	1080 × 1200 per eye	90	110°	—	Virtual desktop as panoramic video player
Both	**UN SALIENT360!**	HTC Vive	1080 × 1200 per eye	90	110°	SMI eye tracker (250Hz)	Unity3D NVIDIA GTX 1080 GPU

Table 3: Execution of the Subjective Tests

	Database	Prior the Test	Stimuli Duration	Between Stimuli	Pause (Rest)	Total Duration	Calibration/Reporting (During/After the Test)	Experiment Environment
Images	**SALIENT 360!**	Vision test Color test	25 s	6 s Grey screen	5 min. middle of test	35 min.	Eye tracker calibration after 1st stimulus + every 3 - 4 min.	Subjects watch stimuli sitting on swivel chair
	ODIs Saliency Maps	Training session	10 s 20 s	—	—	—	—	Subjects freely explore image scene
	VR-Saliency	Explanation	30 s	—	—	10 min.	Eye tracker calibration	Two scenarios, i.e., standing and seated
Videos	**360-Video-Database**	Learn the study steps	Varies with video	5 s (Preparation)	5 min. between groups	40 min.	Participants rated using SAM	Subjects watch stimuli sitting on swivel chair
	360 Head Movement Dataset	Explanation + Training	70 s grey screen	5-10 s	—	7 min.	Oral instruction to adjust HMD vision	Default is subject stands but some asked to sit
	VR User Behavior Dataset	Demo video	Varies with video	Subjects can choose	Varies between subjects	—	Short test afterwards about content being memorized	Subjects watch stimuli sitting on swivel chair
	Head-Tracking	Training session	Varies with video	3 s Grey screen	5 min. after 16 stimuli	30 min. maximum	Rate the videos after each sequence	Subjects watch stimuli sitting on swivel chair
	360Dataset	Learn the study steps	60 s	—	—	30 min. maximum	Questionnaire at the end of the experiment	Subjects stand and turn around freely
	Supplementary Material	Vision test Color test	60 s	60 s Rate the video	60 s between stimuli	90 min. maximum	Opinion scores SSQ	Subjects watch stimuli sitting on swivel chair
	VR-EyeTracking	Vision test	Varies with video	20 s	3 min. between groups	30 min. per group	Eye tracker calibration after each group	Subjects freely explore the video scenes
	AVTrack360	Vision test Color test	30 s	10 s Grey screen	2-3 min. between stimuli	45 min.	SSQ	Subjects watch stimuli sitting on swivel chair
	VQA-ODV	Vision test	10-23 s	6 s Grey screen	5 min. between training/test	60 min.	5 min. eye tracker calibration before training	Subjects freely explore the video scenes
	360_em_dataset	—	43-85 s	—	Arbitrary length	17 min.	Eye tracker calibration before each test session	Subjects watch stimuli sitting on swivel chair
	ODV Dataset	Vision test	10 s	3 s	—	30 min.	—	Subjects watch stimuli sitting on swivel chair
Both	**UN SALIENT360!**	Vision test	10-23 s	—	—	20 min. maximum	Eye tracker calibration every fifth stimulus	Subjects watch stimuli sitting on swivel chair