Data Article

A free database of eye movements watching "Hollywood" videoclips

Francisco M. Costela [a, b, *], Russell L. Woods [a, b]

[a] Schepens Eye Research Institute, Massachusetts Eye and Ear, Boston, MA, USA
[b] Department of Ophthalmology, Harvard Medical School, Boston, MA, USA

ABSTRACT

The provided database of tracked eye movements was collected using an infra-red, video-camera Eyelink 1000 system, from 95 participants as they viewed 'Hollywood' video clips. There are 206 clips of 30-s and eleven clips of 30-min for a total viewing time of about 60 hours. The database also provides the raw 30-s video clip files, a short preview of the 30-min clips, and subjective ratings of the content of the videos for each in categories: (1) genre; (2) importance of human faces; (3) importance of human figures; (4) importance of man-made objects; (5) importance of nature; (6) auditory information; (7) lighting; and (8) environment type. Precise timing of the scene cuts within the clips and the democratic gaze scanpath position (center of interest) per frame are provided. At this time, this eye-movement dataset has the widest age range (22–85 years) and is the third largest (in recorded video viewing time) of those that have been made available to the research community. The data-acquisition procedures are described, along with participant demographics, summaries of some common eye-movement statistics, and highlights of research topics in which the database was used. The dataset is freely available in the Open Science Framework repository (link in the manuscript) and can be used without restriction for educational and research purposes, providing that this paper is cited in any published work.

* Corresponding author. Schepens Eye Research Institute, Massachusetts Eye and Ear, Boston, MA, USA.
E-mail address: Francisco_costela@meei.harvard.edu (F.M. Costela).

Specifications Table

Subject area	*Ophthalmology*
More specific subject area	Visual Science and Psychophysics area, interested in analyzing gaze data from people watching video with a wide spectrum of ages
Type of data	Gaze eye movements, videoclips, trial condition data, demographics tables, timing tables
How data was acquired	An EyeLink 1000 eye-tracker was used
Data format	Gaze unfiltered data (x, y, time coordinates, trial condition data), video, Excel files
Experimental factors	Raw data (from EDF files) were exported as Matlab (easy to access) files preserving all coordinates and adding experimental trial data.
Experimental features	Subjects viewed a subset of a total of 217 clips from professionally recorded video material ("Hollywood" movies) using a high-resolution, infrared-sensing eye tracker
Data source location	Data were collected and stored at Schepens Eye Research Institute, Boston, Massachusetts, USA.
Data accessibility	Data stored in public repository Open Science Framework. Link: https://osf.io/g64tk/.
Related research article	Costela, F. M., & Woods, R. L. (2018). When watching video, many saccades are curved and deviate from a velocity profile model. *Frontiers in neuroscience*, 12, 960, PMC6330331 [1]. Wang, S., Woods, R. L., Costela, F. M. and Luo, G. (2017) Dynamic gaze-position prediction of saccadic eye movements using a Taylor series. *Journal of Vision* 17 (14), 3, PMC5710308 [2].

Value of the data

- Watching television, movies, and other video content is a major source of entertainment, relaxation, education, infor- mation (e.g., current affairs), and supports personal identity, integration, and social interaction [1–3]. Watching video content is a common daily activity, as the average adult in the USA watches an estimated 8 hours of video content per day, of which, about 5 hours are traditional television and about 2 hours are on mobile devices [4]. With such directed content, viewers tend to look in the same place most of the time [5,6], such that there is a common, shared, or democratic video scanpath through the clip or center of interest (COI) per frame. The aim of this publication is to provide the scientific community with a large database of gaze recordings (eye movements with head constrained) from 95 observers with a wide age range. Each participant viewed a subset of clips from a total of 217 clips extracted from professionally-recorded video material ("Hollywood" movies) using a high-resolution, infrared-sensing, video-based, eye tracker. The eye- movement database consists of 2706 viewings by 76 participants of 206 video clips of 30 seconds duration (about 22 hours) and 76 viewings by 19 participants of eleven video clips of 30 minutes duration (about 38 hours) for a total of about 60 hours of viewing. The task of the observers was simply to watch the clip. For the 30-s clips, participants were asked to describe the content of the clip [2]. For each of the 217 video clips, the content of the clip was subjectively classified, in 30- s increments, in eight categories: (1) genre; (2) importance of human faces; (3) importance of human figures; (4) importance of man-made objects; (5) importance of nature; (6) auditory information; (7) lighting; and (8) environment type. Also, in those 217 video clips, the timing of the scene cuts was identified, which may be useful for analyses such as changes in eye movements due to cut scenes.
- At this time, the database is, to our knowledge, the largest video dataset with balanced age distribution among partici- pants (up to 85 years; see Table 1) and the third largest (after Mital et al. [3] and Mathe & Sminchisescu [4]) with a natural- viewing task, made available to the vision research community. Further, it is one of the few databases with an eye-tracking sampling frequency of 1000 Hz and also the one providing the largest number of videoclips (217 total) and the longest sessions for some of the clips (30 minutes). See Table 1 for comparisons with other freely-available video datasets. Eye- movement characteristics when watching video can differ substantially from datasets in which the visual task has been experimentally controlled, therein the importance of this database.
- Most publicly-available eye-movement datasets have static-images as stimuli, such as images of scenes or faces. (see also the MIT Saliency Benchmark: http://saliency.mit.edu/datasets.html). Moving images (video) are more representative of the normal environment and thus this database may be more representative of natural eye movements than those obtained with static images or constrained fixation targets (e.g., stimuli that instantaneously change location). The dynamic nature of the stimulus in a video can produce more complex eye movements. For example, a saccade may be initiated during a pursuit or the saccade target may be gone before the saccade has landed due to a scene cut or movement of the target within the video, whether from object motion or camera changes (e.g., panning).
- This database could be put to use under a variety of research circumstances, such as: gaze statistics, developing and testing models of visual salience, models of ocular motor control, reduction of band-width by restricting high resolution to the video scan path, and certainly others that were not considered. The dataset could be used as stimuli for neurons or computational neurons or networks. It has been proposed that neurons and neuronal systems may respond differently to natural stimuli as compared to manufactured stimuli. The dataset contains natural scenes that are in motion and for which there is gaze data, so may further enhance our understanding of neural mechanisms. For testing of neural arrays, a model of the peripheral degradation in the retinal image quality could be incorporated by modifying the video based on the gaze data. The database may make a significant contribution to the image processing community, for both educational and research purposes.

1. Data

This database includes:

(1) gaze recordings (eye movements with head constrained) from 95 observers as they watched "Hollywood" clips. The raw gaze data is stored in Matlab files that contain a structure ('EyetrackRecord') with the x and y gaze coordinates, sample time, pupil size, and an array indicating which samples were missing data.
(2) 206 video clips (MOV format) of 30 seconds duration viewed by 76 participants (about 21 viewing hours).
(3) the first 30 seconds of eleven 30-min video clips. 19 participants viewed video clips of 30 minutes duration (about 38 viewing hours).
(4) subjective ratings based on the content of the 217 videos for eight different categories, fixation data, democratic center of interest (COI) location per frame, precise timing of the scene cuts within the clips. The subjective ratings are provided for the 30-s video clips and for 30-s segments of the 30-min video clips.
(5) demographic information about the 95 participants (summarized in Table 2).
(6) details about the experimental apparatus, video database, data collection procedure, visual task, explanation of the COI determination, and statistics of the fixations and saccades (Fig. 1) extracted from the dataset.

To corroborate our data:

(1) the Gaussian distribution of gazes for a frame is shown in Fig. 2.
(2) a supplementary video is provided which plots the gaze locations of the participants superimposed on every frame of a 30-sec clip (the video scan paths).
(3) fixations and saccade metrics were compared with other studies.
(4) normalized scanpath saliency scores [5] were calculated across our pool of participants (Fig. 3).
(5) some publications in which subsets of this dataset has been used are mentioned.

The database can be accessed following this hyperlink: https://osf.io/g64tk/.

2. Experimental design, materials, and methods

A. Participants

Ninety-five normally-sighted fluent English speakers (median age: 56.2 years) participated in two studies that were approved by the Institutional Review Board of the Schepens Eye Research Institute in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Informed consent was obtained from each participant prior to data collection.

Preliminary screening of the participants included self-report of ocular health, measures of visual acuity and contrast sensitivity for a 2.5°-high letter target, and evaluation of fixation and central retinal health using retinal videography (Nidek MP-1, Nidek Technologies, Vigonza, Italy or Optos OCT/SLO, Marlborough, MA, USA). All the participants had visual acuity of 20/25 or better, letter contrast sensitivity of 1.67 log units or better, steady central fixation and no evidence of retinal defects. For each participant, self-reported gender, age, and education level (categorized within seven levels based on World Health Organization categories) was obtained. A summary of the demographics of the participants is shown in Table 2. Education information was missing from 24 of the 95 orphan.

Table 1

Comparison between freely-available video-viewing eye-movement datasets. Bold indicates the largest amount in each category.

Database	#subjects	Age range	Sampling rate	#videos	Duration (secs)	Total hours
Actions [4]	20	21−41	500	**1707**	~240	92
ASCMN [5]	13	23−35	30	24	30	2.6
Coutrot Database [6]	72	20−35	1000	60	17	20.4
DIEM [7]	31−218	18−36	1000	85	27−217	~90
GazeCom Video [8]	54	18−34	250	18	20	5.4
Eye-2-I [9]	51	Students	60	4	600	34
IRCCyN Video 1 [10]	37	N/A	50	51	10	5.2
IRCCyN Video 2 [11]	30	~23	50	100	6	5
IRCCyN Stereoscopic [12]	40	19−44	50	41	~70	32
IRCCyN/HD UHD [13]	34	19−44	30	37	~11	3.9
SAVAM [14]	50	18−27	500	41	20	11.4
SFU [15]	15	18−30	30	12	10	0.5
TUD Task [16]	12	Students	250	25	20	1.7
USC CRCNS Orig. [17]	8	23−32	240	50	30	3.4
USC MTV [18]	16	23−32	240	50	30	6.7
USC VAGBA [19]	14	23−32	240	50	30	5.8
Our 30 Secs Dataset	76	22−85	1000	206	30	21.5
Our 30 Mins Dataset	19	22−58	1000	11	**1800**	37.5

B. Studies

The gaze of each subject was tracked monocularly (best or preferred eye) as they viewed an iMac (Apple, Cupertino, CA), mid-2010, 27″ display (60 × 34 cm; 60Hz; 2560 by 1440 pixels; 16:9 aspect ratio) or a Barco F50 projector (Gamle Fredrikstad, Norway) and Stewart Filmscreen (Torrance, CA) Aeroview 70 rear-projection screen to create a 27" display (60 × 34 cm; 120 Hz; 1920 x 1080 pixels 16:9 aspect ratio) from 1 m away for a $33° × 19°$ potential viewing area. The maximum luminance was 128.9 and 723.5 cd/m^2, and the minimum luminance was 0.14 and 0.26 cd/m^2 for the iMac and Barco-Aeroview displays respectively. Sessions were conducted in a dark room with no illumination other than the displays. Illumination at the display surface was 11 lux.

Gaze was tracked with a table-mounted, video-based EyeLink 1000 system (SR Research Ltd., Mississauga, Ontario, Canada). Subjects' head movements were restrained for the duration of the experiment using a SR Research head and chin rest. The clips were displayed and data collected with a MATLAB program using the Psychophysics, Video, and EyeLink Toolboxes. At the beginning of the experiment, the eye tracker was calibrated using a nine-point calibration procedure. If the average error exceeded 1.5° during this step, the calibration was repeated. The Eyelink 1000 has a reported spatial resolution of 0.01° RMS and tracks dark pupil and corneal reflection (first Purkinje image). The

Table 2

Self-reported demographic characteristics of participants.

Gender	Male	52 (54%)
	Female	43 (46%)
Age (median, min-max)		56.2 y (22-85y)
Race/Ethnicity	Black	5 (5%)
	White	87 (91%)
	Asian	4 (4%)
	Hispanic	1 (1%)
	Not registered	3 (3%)
Highest education	High school diploma	4 (5%)
	Some college	6 (8%)
	Bachelor's degree	23 (32%)
	Associate degree	2 (2%)
	Master's degree	18 (24%)
	Professional degree	7 (9%)
	Doctoral degree	11 (17%)

EyeLink file sample filter was set to standard (which reduced the noise level by a factor of 2). The percentage of samples without a reported gaze position was 4.1%.

At the beginning of each trial, participants were instructed to watch the stimulus "normally, as you would watch television or a movie program at home." At the end of each 30-s clip, the participant was asked to describe the contents of that clip [2]. In that first study, 61 participants watched 40 to 46 of 206 thirty-second "Hollywood" video clips in one session. In the second study, 19 participants watched two to nine 30-min movie clips (with calibration repeated every 10 min) across multiple sessions. For those 30-min clips, no further questions were asked after the viewing.

C. Video clips

The 217 "Hollywood" (directed) video clips were chosen to represent a range of genres and types of depicted activities. The genres included nature documentaries (e.g., BBC's Deep Blue, The March of the Penguins), cartoons (e.g., Shrek, Mulan), and dramas (e.g., Shakespeare in Love, Pay it Forward). Progressive AVI format files were obtained from the original NTSC DVDs (some interlaced material could not be converted without artifacts, so those were not used). The aspect ratios of the clips varied widely. When the aspect ratio was lower (less tall) than the 16:9 of the display, the video was shown at full width and in the vertical center with black regions above and below.

Of the 217 clips, 206 clips were 30 seconds long and were selected from parts of the films that had relatively few scene cuts, which was reflected in the average of nine cuts per minute in our clips, as compared to the approximately 12 per minute in contemporary films. The clips included conversations, indoor and outdoor scenes, action sequences, and wordless scenes where the relevant content was primarily the facial expressions and body language of one or more actors. The other eleven clips were 30 minutes long and were extracted from Hollywood movies with similar content and genres (e.g., Inside Job, Bambi, Juno, K-pax, Flash of Genius).

The video content of all 206 30-s clips and for each 30-s segment of the 11 30-min clips was rated for eight categories: (1) importance of human faces; (2) importance of human figures; (3) importance of man-made objects; (4) importance of nature; (5) auditory information; (6) lighting; (7) environment type and (8) genre, where "importance" relates to the ability to follow the story (comprehension). Three observers who did not participate in the studies in which gaze was tracked were asked to rate based on the importance of the categories human faces, human figures, man-made objects, nature and auditory information for understanding of the video content. Each 30-s segment was rated by at least one observer. In a previous study of image enhancement that measured preference, a difference between the importance rating of human faces and other importance ratings (e.g. nature, human figures) on preference was shown. Each rating scale ranged from 0 to 5, where 0 was absent (or unimportant) and 5 was present (or important for understanding), except for environment type that was binary (outdoor or indoor). The responses of the observers were averaged for each rating category for each 30-s video segment.

The database includes the 206 30-s video clips and 11 30-s previews of the 30-min video clips that were used for data collection. Information about the source of the clips is provided. Provision of the 30-s video clips is considered fair use (1976 Copyright Act in the USA) and is not an infringement of copyright since the clips are only a small portion of the content of the material from which they were taken, were used for research, and would have no impact on the commercial value of the source material.

D. Scene cut detection

For each clip, each frame was converted to grayscale, and the sum of the between-frame, pointwise differences in log-grayscale value at each pixel was calculated, after normalizing by the size of the clip in each frame. A threshold based on the average of this metric plus one standard deviation was used to discriminate those transitions that relate to scene cuts. This threshold provided a better sensitivity (0.7) than using two (0.55) or more standard deviations when compared with the number of cuts reported by the same observers that rated the video categories at no expense of specificity.

For those clips where the number of cuts calculated by our algorithm was lower than the number of cuts reported by observers, the time of the missed cuts was added manually. The algorithm found some additional "scene cuts" in clips. Some of these related to fast transitions (i.e., revolving doors or fast-motion panning) that were kept, given the original purpose of a previous study on the effect of distractors on saccades. Those cases where the transitions were due to slow fading between scenes were removed. This algorithm was applied to detect scene cuts within the eleven 30-min clips, but no manual checking was performed. The specific scene cut timings for each clip in the dataset is provided.

E. Gaze analysis

Users of our database can implement their own eye-movement classification algorithm using the provided raw eye movement data. Detecting saccades and fixations in this dataset is more difficult than in datasets with artificial stimuli or static images because the eye movement sequences made while watching video are often more complex. In particular, saccades may not start or end with a fixation and the inter-saccadic interval is often short (<200 ms). The viewed scene may have moving objects of interest and other objects in the scene may also be moving, all of which may act as a stimulus that affects eye movements.

Sampling frequency was 1 kHz for all examples. For our analyses, blinks were identified and removed using Eyelink's online data parser. Periods preceding and following these missing data were removed if they exceeded a speed threshold of 30°/s. Then, interpolations were applied over the removed blink data by applying cubic splines. Data were smoothed and then velocity criteria and additional criteria (e.g. removing glissades) were employed. Additional details about our saccade classification algorithm can be found in a recent publication [1]. Other methods, such as finding the peak velocity and its acceleration window or k-means cluster analysis, could be employed. Using our saccade categorization, the pool of extracted saccades followed the saccadic peak velocity-magnitude relationship ("main sequence"; Fig. 1A), abd there was a skewed distribution of magnitudes (median 4.02°; Fig. 1A), with horizontal and vertical saccades being most common (Fig. 1B).

Fixations per clip and subjects (including gaze location in degrees and duration) are available in the dataset. These results are consistent with the saccadic amplitudes (median 5.5°) and fixation durations (median 326 ms) reported by Dorr et al. (2008) [7].
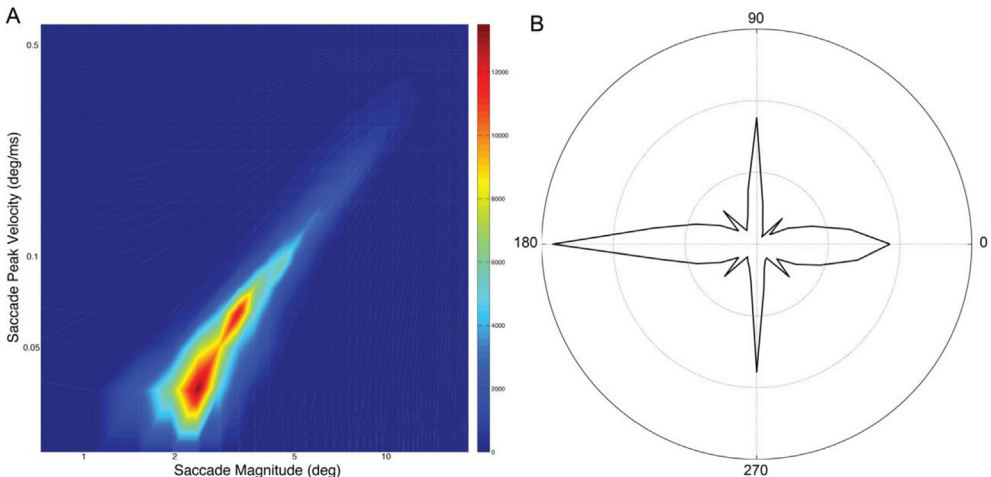
Fig. 1. Saccadic features for all 95 subjects and 296,249 saccades. A) Saccadic peak velocity–magnitude main sequence. The distribution is plotted on logarithmic scale where peak velocity is indicated on the y-axis and magnitude indicated on the x-axis. Colors represent the number of samples (right-hand scale). B) Polar distribution of saccade directions (bin size 4.5°). Distance from the origin represents frequency.

F. COI determination

Each viewer of a video has a unique video scanpath, the sequence of gaze locations across time. Within each frame, when there are multiple viewers with normal sight, most people look in about the same place most of the time. We refer to those gaze locations with a frame as the center of interest (COI). We provide a supplementary video with the video scanpaths for one clip.

Different subsets of 30-s video clips were randomized and assigned to each of the 61 participants in Study 1. Overall, each of the 206 30-s video clips was viewed by 24 participants. For each video clip, blinks, saccades, and other lost data were removed, which should leave fixations and pursuits. For each video frame (33 ms), for each participant, the gaze position data points (1−33) were averaged. While the gaze of all participants is often in one location, it can be distributed across more than one location, particularly in scenes when there are two people speaking (Fig. 2A). When the distribution of gaze locations is not unimodal, methods such as average or median location may not represent the full distribution well. A novel method to determine the democratic COI was used. First, for each frame in each clip, for all the available data from all participants (up to 24), a kernel density estimate of the average gaze positions was computed (Fig. 2B). Then, the area under the region of the density estimate for all potential positions of a rectangular box (a quarter of the dimensions of the original frame) was integrated across the frame interpolating with a symmetrical Gaussian function. The democratic COI


Fig. 2. Example of democratic COI determination. (A) Gaze locations of the 24 participants during one frame of one video clip. (B) Kernel density estimate of those gaze locations in that video frame shown as a heat map, with red representing a higher density. The green rectangle represents the box used to determine the democratic COI.

was defined as the location of the center of the box with the highest integral value (Fig. 2B). The benefit of this approach over averaging or taking the median of the gaze points is that it better accounts for multimodal gaze distributions.

G. Eye movement similarity

A variety of methods has been proposed in the literature to assess the consistency of eye movements across different observers. Eye movement similarity was assessed using a normalized scanpath saliency (NSS) approach [7]. The advantage of this approach, in contrast to methods involving clustering algorithms or string-editing, is that it sums smooth Gaussian functions for each fixation, so that nearby fixations are weighted more than far away fixations, and no discrete bins or regions of interest need to be defined. It is robust to outliers, and also incorporates temporal smoothing.

NSS maps are normalized and scaled to a unit standard deviation. Consequently, NSS scores close to zero indicate a degree of similarity between scanpaths no greater than would be expected from a random sampling of points. Scores greater or less than zero are roughly analogous to positive or negative correlation scores, but have no minimum or maximum bounds. Parameters specified for this analysis by Dorr et al. [7] were used. In that study, gaze patterns were compared between natural movies (i.e., natural scenes without direction of action) and directed (professionally cut) Hollywood trailers. Those gaze patterns were found to be systematically different. In particular, the Hollywood trailers evoked very similar eye movements among observers (higher coherence) and showed the strongest bias for the center of the screen. NSS scores can vary enormously depending on the content and they are meaningful only within the same study. Here, a similar pattern was found wherein most of the participants with normal vision who participated in Study 1 had a high NSS coherence score (see Fig. 3). Two participants showed an unusual behavior with low NSS scores but we found no reason to exclude them based on missing data during recording. Our results are also in agreement with Dorr et al. [7] confirming that gaze during Hollywood movies are highly coherent.
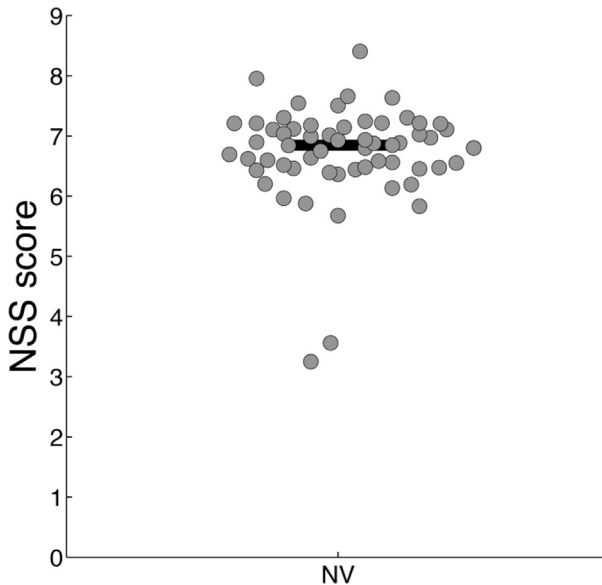
Fig. 3. Mean NSS score for participants. Gray filled circles represent the average NSS score for each of 61 participants who viewed the 30-s clips. The flat horizontal black line corresponds to the global average for the group.

Acknowledgments

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dib.2019.103991.

References

[1] F.M. Costela, R.L. Woods, When watching video, many saccades are curved and deviate from a velocity profile model, Front. Neurosci. 12 (960) (2018).
[2] S. Wang, R.L. Woods, F.M. Costela, G. Luo, Dynamic gaze-position prediction of saccadic eye movements using a Taylor series, J. Vis. 17 (14) (2017) 3, https://doi.org/10.1167/17.14.3. PubMed PMID: 29196761; PubMed Central PMCID: PMCPMC5710308.
[3] H. Terasaki, H. Hirose, Y. Miyake, et al., S-cone pathway sensitivity in diabetes measured with threshold versus intensity curves on flashed backgrounds, A, Mar; SI-O-V-S, 0146-0404 I, Investig. Ophthalmol. Vis. Sci. 37 (4) (1996), 680-04.
[4] S. Mathe, C. Sminchisescu, Actions in the eye: dynamic gaze datasets and learnt saliency models for visual recognition, IEEE Trans. Pattern Anal. Mach. Intell. 37 (7) (2015) 1408−1424, https://doi.org/10.1109/TPAMI.2014.2366154. PubMed PMID: 26352449.
[5] N. Riche, M. Mancas, D. Culibrk, V. Crnojevic, B. Gosselin, T. Dutoit, Dynamic saliency models and human attention: a comparative study on videos, in: Asian Conference on Computer Vision, Springer, Berlin, Heidelberg, 2012, pp. 586−598.
[6] A. Coutrot, N. Guyader, How saliency, faces, and sound influence gaze in dynamic social scenes, J. Vis. 14 (8) (2014) 5, https://doi.org/10.1167/14.8.5. PubMed PMID: 24993019.
[7] P.K. Mital, T.J. Smith, R. Hill, J.M. Henderson, Clustering of gaze during dynamic scene viewing is predicted by motion, Cognitive Computation 3 (1) (2011) 5−24.
[8] M. Dorr, T. Martinetz, K.R. Gegenfurtner, E. Barth, Variability of eye movements when viewing dynamic natural scenes, J. Vis. 10 (10) (2010) 28, https://doi.org/10.1167/10.10.28. PubMed PMID: 20884493.
[9] K.-T. Ma, X. Qianli, L. Rosary, L. Liyuan, S. Terence, K. Mohan, Eye-2-I: eye-tracking for just-in-time implicit user profiling, in: IEEE 2nd International Conference on Signal and Image Processing (ICSIP), 2017, pp. 311−315.
[10] S. Péchard, P. Pépion, P. Le Callet, Region-of-interest intra prediction for H.264/AVC error resilience, in: Proc ICIP, 2009 (Cairo, Egypt).
[11] U. Engelke, P. Pepion, H. Le Callet, Z. Zepernick, Linking distortion perception and visual saliency in H.264/AVC coded video containing packet loss, in: Proc VCIP, Huang Shan, China, 2010.
[12] Y. Fang, J. Wang, J. Li, R. Pépion, P. Le Callet, An eye tracking database for stereoscopic video. In Quality of Multimedia Experience (QoMEX), in: Sixth International Workshop on IEEE, 2014, pp. 51−52.
[13] T. Vigier, J. Rousseau, M.P. Da Silva, P. Le Callet, A new HD and UHD video eye tracking dataset, in: Proceedings of the 7th International Conference on Multimedia Systems, ACM, 2016, p. 48.
[14] M. Gitman YE, D. Vatolin, A. Bolshakov, A. Fedorov, Semiautomatic visual-attention modeling and its application to video compression, ICIP (2014).
[15] H. Hadizadeh, M.J. Enriquez, I.V. Bajic, Eye-tracking database for a set of standard video sequences, IEEE Trans. Image Process. 21 (2) (2012) 898−903, https://doi.org/10.1109/TIP.2011.2165292. PubMed PMID: 21859619.
[16] H. Alers, J. Redi, I. Heynderickx, Examining the effect of task on viewing behavior in videos using saliency maps, in: Proc SPIE, 2012, pp. 22−26, 8291(San Jose, CA).
[17] L. Itti, Automatic foveation for video compression using a neurobiological model of visual attention, IEEE Trans. Image Process. 13 (10) (2004) 1304−1318. PubMed PMID: 15462141.
[18] R. Carmi, L. Itti, The role of memory in guiding attention during natural vision, J. Vis. 6 (9) (2006) 898−914, https://doi.org/10.1167/6.9.4. PubMed PMID: 17083283.
[19] Z. Li, S. Qin, L. Itti, Visual attention guided bit allocation in video compression, Image Vis Comput. 29 (1) (2011) 1−14.