X-Eye: A Reference Format For Eye Tracking Data To Facilitate Analyses Across Databases

Stefan Winkler, Florian M. Savoy, Ramanathan Subramanian

Advanced Digital Sciences Center, University of Illinois at Urbana-Champaign, Singapore

ABSTRACT

Datasets of images annotated with eye tracking data constitute important ground truth for the development of saliency models, which have applications in many areas of electronic imaging. While comparisons and reviews of saliency models abound, similar comparisons among the eye tracking databases themselves are rare. In an earlier paper,[1] we reviewed the content and purpose of over two dozen databases available in the public domain and discussed their commonalities and differences. A major issue is that the formats of the various datasets vary a lot owing to the nature of tools used for eye movement recordings, and often specialized code is required to use the data for further analysis. In this paper, we therefore propose a common reference format for eye tracking data, together with conversion routines for 16 existing image eye tracking databases to that format. Furthermore, we conduct a few analyses on these datasets as examples of what *X-Eye* facilitates.

Keywords: Eye tracking, visual attention, saliency, gaze, fixations, common format, comparison, center bias

1. INTRODUCTION

Modeling saliency, which relates to detection/identification of the scene information that attracts visual attention, has been an active topic of interest among the computer vision, graphics, human-computer interaction, and other communities such as advertising. A number of saliency models taking into account bottom-up (or stimulus-based) and top-down (or perception-based) factors have been proposed in the literature – Borji and Itti recently published an exhaustive review.[2] Eye-tracking databases typically provide ground truth for saliency models to learn image regions of interest based on eye-movement patterns observed with human subjects. Recently, we reviewed over two dozen publicly available eye tracking databases to help researchers identify the appropriate dataset for their saliency studies.[1] Nevertheless, a thorough comparison of eye-tracking data from different datasets has not yet been attempted.

A major impediment to analyzing and comparing eye tracking databases is that they tend to provide data in different formats, owing to the nature of hardware and software used for sampling and recording eye movements. Eye movements comprise *fixations*, denoting stationary phases where scene information is absorbed by the human visual system, and *saccades*, representing ballistic motion of the eyes to sample different scene regions. Even as information regarding fixations and saccades (e.g. fixation start and end times, saccade begin and end coordinates) can be extracted from raw eye tracking data, it can be quite tedious and confusing for researchers to make sense of the data for further use. For instance, some eye trackers output gaze data with reference to the stimulus, while others compute gaze positions with respect to screen coordinates. Therefore, in many cases, specialized code is necessary for interpreting and converting the raw eye tracker output into a known reference format.

To this end, we propose *X-Eye*, a reference format for describing eye movement data to facilitate comparison and analysis of eye tracking databases. The conversion routines for 16 existing image-based eye tracking databases to the *X-Eye* format can be downloaded http://vintage.winklerbros.net/x-eye.html. Also, in order to provide a flavor of how a common format can facilitate data analysis, we look at some basic eye tracking statistics and compare the center bias across datasets.

The paper is organized as follows. Section 2 presents related work and databases. Section 3 introduces the proposed *X-Eye* reference format. Section 4 illustrates several example use cases for cross-database analysis. Section 5 concludes the paper.

Send correspondence to S. Winkler, E-mail: stefan.winkler@adsc.com.sg

2. EYE TRACKING DATASETS

2.1 Related Work

We presented an overview and comparison of over two dozen eye tracking databases in an earlier paper,[1] reviewing parameters such as the type of stimuli analyzed or constraints during data acquisition. That study was essentially designed to facilitate future researchers in identifying and using the right dataset for their analyses. However, the comparisons we presented there were based on "meta-data", such as the number of images and subjects, whereas we did not conduct any comparative analysis using the actual eye tracking data provided by these databases.

Recently, a few other works have also analyzed characteristics of eye tracking datasets, albeit as an aside to the evaluation of saliency algorithms. Borji and Itti present a brief overview of image and video-based eye tracking databases used for evaluation of saliency models as part of a survey of state-of-the-art saliency methods.[2] Borji et al. also compare the agreement between eye fixation maps and saliency predictions for 35 saliency models for 3 image-based and 2 video-based datasests using three different types of evaluation scores.[3] The analysis concludes that in general, there is a gap to bridge in order to make saliency predictions human-like, and discusses the need for incorporating top-down factors in saliency approaches as one of the key requirements to this end.

In a subsequent work[4] analyzing indices used for evaluating saliency models and datasets employed for visual attention prediction, the authors systematically consider the effect of factors such as center-bias on saliency modeling, and compare the performance of 32 saliency models on 4 image-based eye tracking datasets. Their analysis identifies the *MIT CSAIL*[5] and *NUSEF*[6] datasets as the most suitable for understanding visual attention, given the large number of stimuli and subjects for which eye movement recordings are available, and the *KTH*[7] and *NUSEF*[6] datasets as the hardest for saliency modeling.

2.2 Summary of databases

An overview of the test material, subjects, viewing setup, and other experimental details of each database is provided in Table 2. Additional specifics of those eye tracking datasets that were not covered in our earlier paper[1] are discussed below. An up-to-date list of eye tracking databases is available on the author's home page, http://stefan.winkler.net/resources.html.

- *DUT-OMRON*[8] database contains eye movement recordings for 5172 natural, high-resolution images selected from the SUN[9] dataset. All images are of maximum 400 × 400 pixels resolution and contain one or more salient objects with a complex background. A total of 25 subjects were involved in annotating both salient objects through rectangles (5 such annotations were acquired per stimulus), and eye fixation ground truth. During the experiment, each participant was instructed to draw rectangles around salient objects in the image, as determined by their own perception. This helps to obtain a better understanding of which salient scene objects were fixated by users, as the rectangle annotations facilitate the removal of outliers commonly observed with eye fixation data.

- *IRCCyN LIVE*[10] dataset is part of a study comparing eye fixation density maps (FDMs) acquired from different eye tracking systems. The authors analyzed the effect of stimulus presentation time and image semantics and evaluated the impact of FDM differences on three applications, namely saliency modeling, image quality assessment, and image retargeting. To this end, they compiled eye movement data in three different laboratories (in different geographical locations) employing different eye tracking hardware, with identical protocol. 29 images from the LIVE database[11] comprising natural images were shown to 15-21 subjects. This eye tracking study confirms that despite significant differences in data acquisition conditions (including human factors such as cultural differences), the resulting FDMs are very similar and can be used as reliable ground truth.

- *Memorability*[12] database was compiled to examine the relationship between image memorability and visual attention. To this end, the authors recorded eye movements from seventeen subjects (10 male, 7 female) for 135 images from the image memorability dataset.[13] The eye tracking data was used to demonstrate that attention-related features such as scene coverage can better account for image memorability as compared to low-level image features.

3. X-EYE REFERENCE FORMAT

As mentioned in Section 1, a key impediment in analyzing eye movement data across different databases is that the recorded data format varies for different datasets. To cite a few examples:

- Some eye trackers produce visual attention data in the form of fixations (coordinates plus durations), while others generate only the raw gaze coordinates from which the above information can be extracted.
- The manner in which the gaze coordinates are output also varies depending on the eye tracking system used – some output gaze positions with respect to the stimulus coordinate system, while others output these positions with reference to screen coordinates.
- The data organization strongly varies from one dataset to another. While most of them provide separate data for different images and observers, it can be non-trivial to establish which image or user some data refers to. A few datasets provide one *.mat* file and organize the information using *MATLAB* structures, while others come with multiple files and make use of meta-data fields or folder/file names to identify the contents.

The above factors mean that each database comes in its own data representation and format. Other researchers who want to make comparisons across databases have to spend significant time and effort converting all the data to a canonical format. The main contribution of this work is that the proposed *X-Eye* format can significantly facilitate the analysis and evaluation of eye tracking data.

The *X-Eye* eye movement descriptor format consists of the following information (an example is shown in Figure 1):

1. Stimulus ID (name of the image stimulus used).
2. Stimulus dimensions (image width and height).
3. Subject ID. Together with stimulus ID, this allows for easier analysis on a per-stimulus and per-subject basis (e.g. the computation of inter-observer agreement).
4. Number of eye fixations recorded for the subject.
5. Mean fixation duration for the subject. Together with the previous item, this enables the computation of basic fixation-related statistics without having to actually parse the fixation data.
6. Fixation number denoting sequence of fixations made.
7. Fixation (x, y) positions. We universally adopt image coordinates here, which is advantageous on two counts: no other information (such as screen resolution or scaling) is required to determine user-fixated locations, and the need for writing additional code to center the stimulus (which is typical of eye tracking systems) based on screen and stimulus dimensions is eliminated.
8. Fixation begin and end times. This information is useful when the temporal sequence of fixations (*where* did the observers look during early and late fixations?) is of interest. The begin and end times are output in milliseconds, and the quantities are computed assuming that the image stimulus was presented at time $t = 0$.
9. Fixation duration – the difference between fixation begin and end times in milliseconds.
10. Inter-fixation duration – the time interval between the end time of the current fixation and the begin time of the next, which can comprise one or more saccades.

```
image name = automan_06.png
image width = 1024
image height = 768

user name = subject_17

number of fixations = 13

average fixation duration = 238.769231

Fix no, Xpos, Ypos, Begintime, Endtime, Duration, Interfix
1       480.4   598.9   330     502     172     36
2       377.6   729.2   538     718     180     1212
3       113.9   494.2   1930    2134    204     52
4       61.9    197.2   2186    2470    284     52
5       493     160.5   2522    2806    284     28
6       533.3   208.9   2834    3102    268     48
7       158.4   495.2   3150    3382    232     24
8       123     539.7   3406    3566    160     44
9       212.6   273.5   3610    3766    156     40
10      381.2   103.2   3806    4214    408     56
11      840.8   421     4270    4558    288     44
12      668.3   704.5   4602    4878    276     44
13      808.2   578.8   4922    5114    192     0
```

Figure 1: Sample data for a given subject and image of the *KTH* dataset in *X-Eye* format.

Dataset	Fixation locations	Fixation durations	Inter-fixation durations	Raw eye-tracking data	Coordinate system
DUT-OMRON[8]	**Yes**	No	No	No	Image
FiFA[14]	Yes	Yes	No	**Yes**	Image
GazeCom Image[15]	No	No	No	**Yes**	Image
IRCCyN Image 1[16]	**Yes**	**Yes**	No	No	Image
IRCCyN Image 2[17]	No	No	No	**Yes**	Screen
IRCCyN LIVE[10]	No	No	No	**Yes**	Screen
KTH[7]	**Yes**	**Yes**	**Yes**	No	Image
LIVE DOVES[18]	Yes	Yes	No	**Yes**	Image
McGill ImgSal[19]	Yes	No	No	**Yes**	Image
Memorability[12]	**Yes**	**Yes**	No	No	Image
MIT CSAIL[5]	No	No	No	**Yes**	Image
MIT CVCL[20]	Yes	Yes	No	**Yes**	Image
MIT LowRes[21]	No	No	No	**Yes**	Image
NUSEF[6]	**Yes**	**Yes**	**Yes**	No	Screen
Toronto[22]	**Yes**	**Yes**	**Yes**	No	Screen
VAIQ[23]	No	No	No	**Yes**	Image

Table 1: Information provided by the various datasets, and data used for the conversion to *X-Eye* (highlighted in bold). Raw eye tracking data are used whenever possible. Screen coordinates are converted to image coordinates.

Table 1 shows the type of data provided by each dataset. It also describes the strategy our conversion routines follow. Some of the datasets only provide raw eye tracking data. In those cases, we adopt the acceleration-based algorithm employed by Judd et al.[5] to detect fixations given the raw gaze data and gaze sampling frequency. Other datasets provide fixation locations and durations, but no information about inter-fixations. If those also provide raw eye-tracking data, we only use the latter and process the data with the above-mentioned algorithm. This allows us to recover the inter-fixation durations uniformly across all databases that provide raw data. However, the resulting fixation information might differ from the one in the databases due to differences in the extraction algorithm. Some datasets do not provide enough data to populate all the variables of the *X-Eye* format, in which cases the missing fields are set to zero. As can be seen from Table 1, most databases provide either raw eye tracking data or all the required timing information, with the exception of *DUT-OMRON*, *IRCCyN Image 1*, and *Memorability* databases.

Furthermore, the *IRCCyN Image 2*, *IRCCyN LIVE*, *NUSEF*, and *Toronto* datasets provide fixation locations in screen coordinates. The conversion process to retrieve the image coordinates is often described in their documentation. For some experiments, the images are displayed at the center of the screen, while others resize them to cover the full screen. In ambiguous cases, we plot the fixations points on top of the images and choose the conversion whose result makes the most sense.

Raw-to-*X-Eye* conversion routines (in *MATLAB* code) as well as *README* files describing how to retrieve the data for the 16 image-based eye movement databases listed in Table 2 can be downloaded from http://vintage.winklerbros.net/x-eye.html. The conversion routines write the output to `.txt` and `.mat` files. We decided not to release the converted eye tracking data as such, as this may be against the terms of use of some datasets.

4. EXAMPLE USE CASES

We now demonstrate various types of analysis that are made possible by our common reference format *X-Eye*. We first compare a number of basic eye tracking statistics across datasets and then study the phenomenon of center bias in more detail.

4.1 Basic statistics

To complement the meta-data analysis of our earlier paper,[1] we compare the average number of viewers and the number of images or videos in each database. This is shown in Figure 2. Clearly, there is a trade-off between the amount of test material and the number of viewers, due to the amount of time needed for the experiments. For example, the recent *DUT-OMRON* database has the most images by far, but only 5 viewers per image. Differences between the data shown here and our earlier meta-data analysis are due to the fact that not all subjects viewed all images in every experiment.

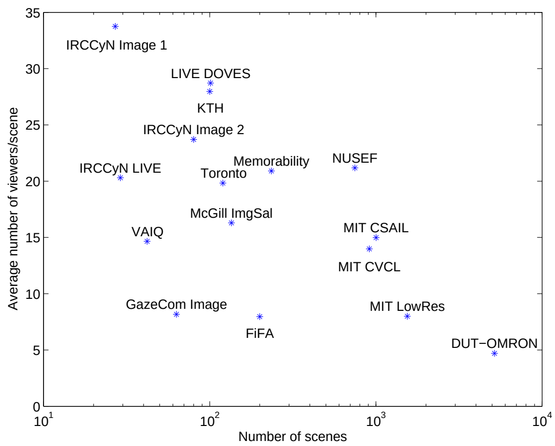


Figure 2: Average number of viewers vs. number of scenes.

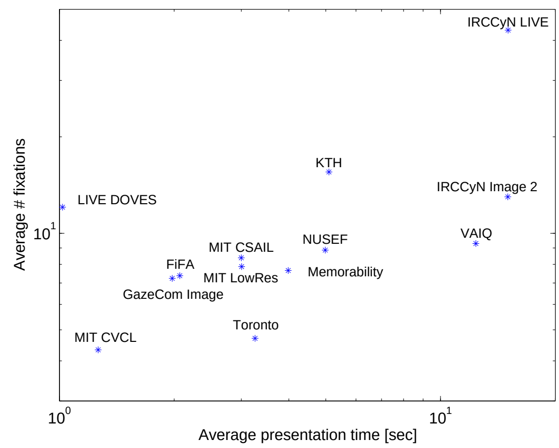


Figure 3: Average number of fixations vs. average presentation time.

The average number of fixations per image and subject roughly increases with the average presentation time, as shown in Figure 3. Figure 4 shows the proportion of time spent in fixations (as opposed to saccades), which is approximately 80-90% of viewing time for most databases. Differences between datasets are expected, as they differ in terms of the image content as well as the tasks given to subjects.

Finally, Figure 5 shows the total viewing time aggregated over all subjects and scenes, as an indication of the overall amount of eye tracking data and fixations provided in each dataset. *NUSEF* has a clear lead with nearly 22 hours of eye tracking data, whereas the datasets at the opposite end (*FiFA*, *GazeCom*) have less than one hour.* For certain databases (including e.g. *NUSEF*), there is a substantial difference between the viewing time estimated from the meta-data[1] and the actual viewing time as computed from the eye tracking data, highlighting the need to use the actual eye tracking data for accurate comparisons.

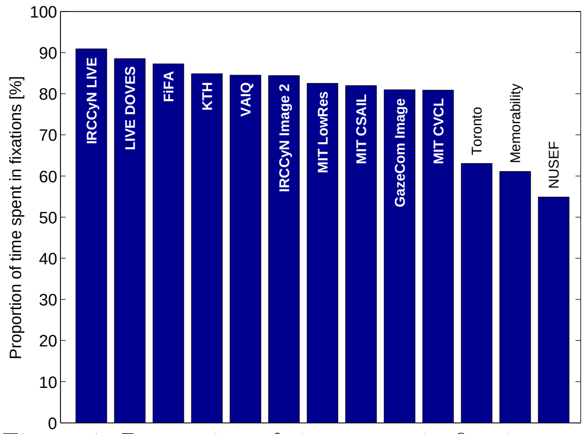


Figure 4: Proportion of time spent in fixations as a percentage of total.

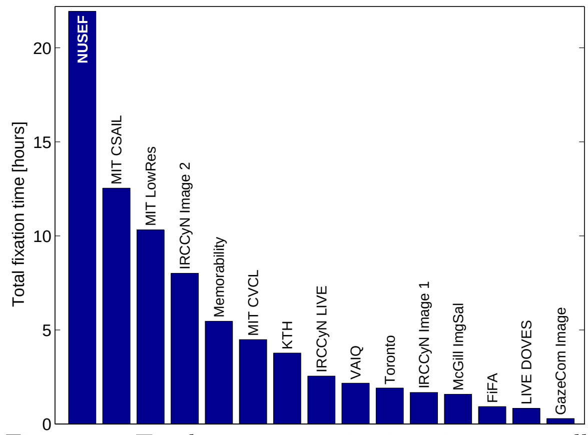


Figure 5: Total aggregate viewing time over all subjects and images.

4.2 Center bias

To further demonstrate the utility of the *X-Eye* reference descriptor format, we analyze the extent of center bias for the considered datasets. Center bias refers to the combined effect of two biases: the propensity of viewers to preferentially attend to details around the image center (usually) before moving on to decode the remaining scene details, and the tendency of photographers (content creators) to place the object(s) of interest near the center of the scene being imaged. The phenomenon of center bias has been extensively discussed in a number of saliency studies.[3–5] At least two of them[3,4] have investigated the influence of center bias on indices denoting saliency prediction accuracy, but this analysis was restricted to only a few datasets.

For our analysis of center bias, we consider overlapping rectangular image regions. We define the center bias as the amount of fixations falling inside the rectangle divided by the total amount of fixations. We weight the fixations by their duration if this data is available. In the first step we compare the center bias for two rectangle sizes, one containing the central 11% of the image area, and one containing the central 25% (see Figure 6). We find the results to be highly correlated, as shown in Figure 7a (correlation coefficient of 0.9477, regression line: $y = 0.91x - 17$). Therefore, we use the 25% region for the remainder of the analysis. The large variations in center bias across databases are also evident from this plot, ranging from 40-80% (more on this below).

We then compare early center bias (i.e. percentage of fixations within the first 500 ms falling into the central region) vs. overall center bias (i.e. percentage of overall fixations within the central region). This is shown in Figure 7b. Early fixations can exhibit a higher center bias for a number of reasons. However, early center bias is not very pronounced for most databases, except for two (*KTH* and *IRCCyN LIVE*), whose early center bias is in the 80-90% range, compared to their overall center bias of only 40-50%.

* The *DUT-OMRON* database does not contain fixation timing information, so we cannot compute the actual viewing time, but according to the meta-data it should be about 14 hours.

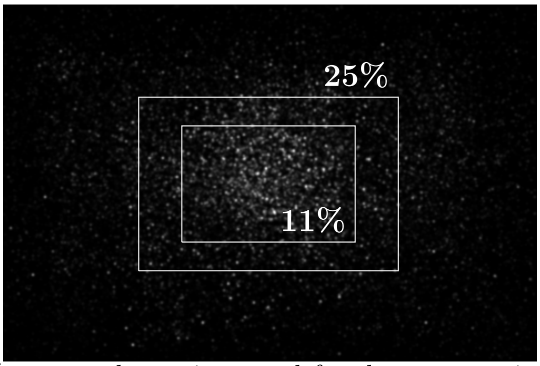
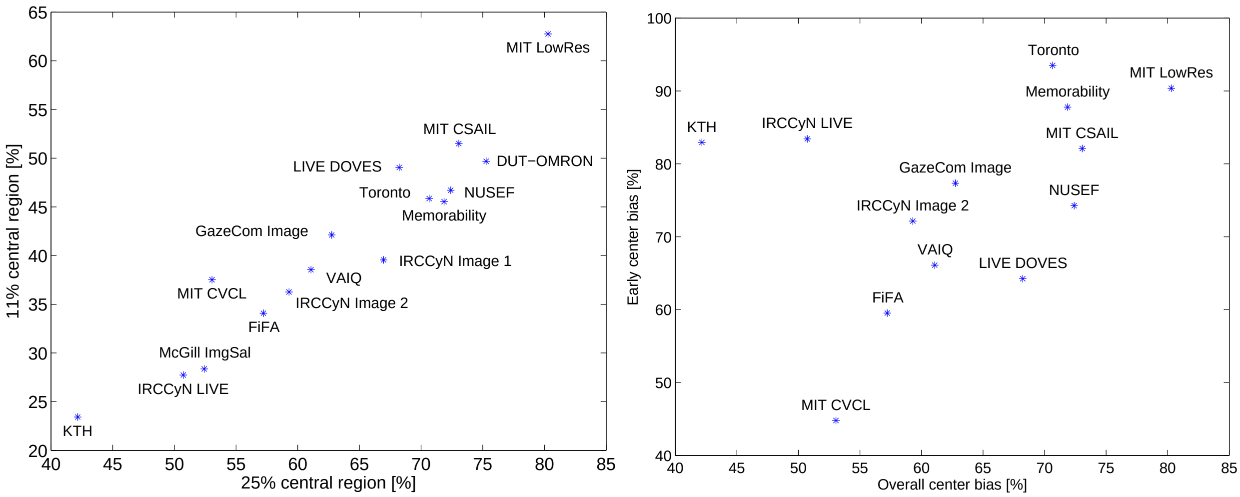


Figure 6: Central 11% and 25% rectangular regions used for the computation of center bias, overlaid on the heatmap of all fixations of the *IRCCyN Image 2* dataset.



(a) 11% vs. 25% regions (b) Early vs. overall

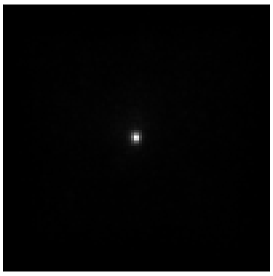
Figure 7: Analysis of center bias.

To demonstrate the distribution of fixations across the image area, we generate heat maps for each dataset. We resize all the images to 100×100 pixels and modify their fixation locations accordingly. We create a matrix of the same size, in which we add the durations of all fixations according to their location. We finally filter the matrix with a Gaussian smoothing kernel ($\sigma = 0.5$) and normalize the results. Figure 8 shows the heat maps of the *LIVE DOVES* and *KTH* datasets, which are each representative of several others. They illustrate the significant differences between datasets in terms of spatial fixation distribution. The fixations of some are concentrated near the center of the image, while for others they are spread out more evenly across the entire image area.

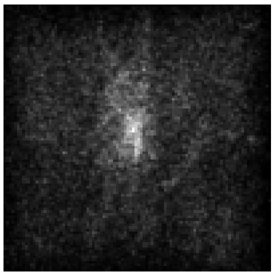
Figure 9 shows the histogram of fixations locations for those two datasets. The proportion of fixations within a given percentage range of the total image area represents the amount of fixations (weighted by their durations if available) falling inside the central rectangle of that size divided by the total number of fixations. The histograms are not cumulative; all fixations are thus counted only once (inside the smallest possible rectangle). The first bar, which represents the fixations lying in the central rectangle covering 10% of the total image size, is much larger for *LIVE DOVES* than *KTH*, consistent with the two corresponding heat maps shown in Figure 8.

5. CONCLUSIONS

We proposed *X-Eye*, a common reference format for eye tracking data. Conversion routines for 16 existing image eye tracking databases to that format can be downloaded from http://vintage.winklerbros.net/x-eye.html.

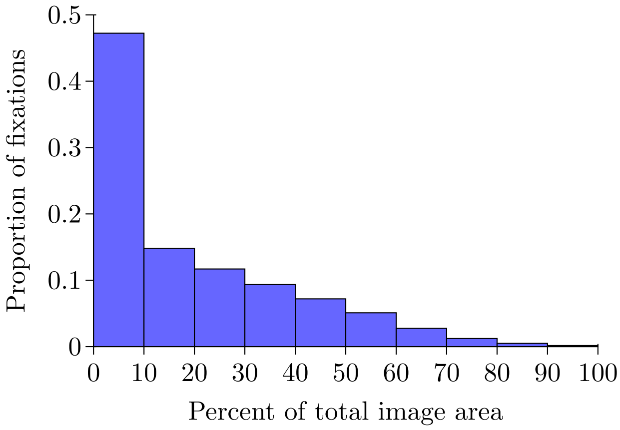


(a) *LIVE DOVES*

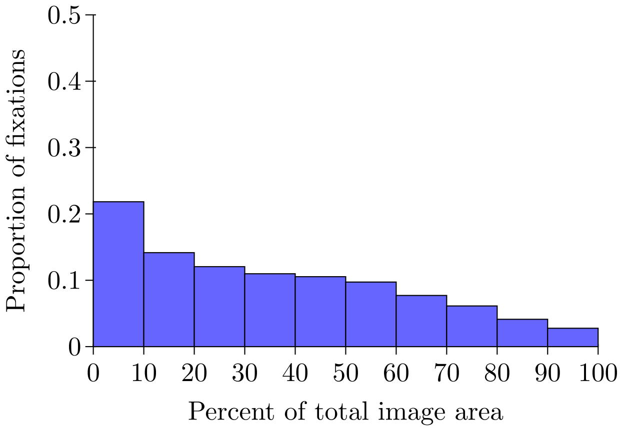


(b) *KTH*

Figure 8: Heat maps of *LIVE DOVES* and *KTH* datasets.



(a) *LIVE DOVES*



(b) *KTH*

Figure 9: Histograms of fixation locations for *LIVE DOVES* and *KTH* datasets.

We demonstrate the utility of such a common reference format by conducting various comparative analyses on these datasets. We find significant differences among datasets in terms of several basic statistics, and most notably center bias. We hope *X-Eye* will facilitate further quantitative cross-database comparisons.

6. ACKNOWLEDGMENTS

This study is supported by the research grant for ADSC's Human Sixth Sense Programme from Singapore's Agency for Science, Technology and Research (A*STAR).

REFERENCES

[1] Winkler, S. and Ramanathan, S., "Overview of eye tracking datasets," in [*Proc. International Workshop on Quality of Multimedia Experience (QoMEX)*], 212–217 (July 3–5, 2013).

[2] Borji, A. and Itti, L., "State-of-the-art in visual attention modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**(1), 185–207 (2013).

[3] Borji, A., Sihite, D. N., and Itti, L., "Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study," *IEEE Transactions on Image Processing* **22**(1), 55–69 (2013).

[4] Borji, A., Tavakoli, H. R., Sihite, D. N., and Itti, L., "Analysis of scores, datasets, and models in visual saliency modeling," in [*Proc. International Conference on Computer Vision (ICCV)*], (Dec. 1–8, 2013).

[5] Judd, T., Ehinger, K., Durand, F., and Torralba, A., "Learning to predict where humans look," in [*Proc. International Conference on Computer Vision (ICCV)*], (2009). http://people.csail.mit.edu/tjudd/WherePeopleLook/.

[6] Ramanathan, S., Katti, H., Sebe, N., Kankanhalli, M., and Chua, T.-S., "An eye fixation database for saliency detection in images," in [*Proc. European Conference on Computer Vision (ECCV)*], (2010). http://mmas.comp.nus.edu.sg/NUSEF.html.

[7] Kootstra, G., de Boer, B., and Schomaker, L. R. B., "Predicting eye fixations on complex visual stimuli using local symmetry," *Cognitive Computation* **3**(1), 223–240 (2011). http://www.csc.kth.se/~kootstra/index.php?item=215&menu=200.

[8] Yang, C., Zhang, L., Lu, H., Ruan, X., and Yang, M.-H., "Saliency detection via graph-based manifold ranking," in [*Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*], 3166–3173 (June 23–28, 2013). http://ice.dlut.edu.cn/lu/DUT-OMRON/Homepage.htm.

[9] Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., and Torralba, A., "SUN database: Large-scale scene recognition from abbey to zoo," in [*Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*], 3485–3492 (2010).

[10] Engelke, U., Liu, H., Wang, J., Le Callet, P., Heynderickx, I., Zepernick, H.-J., and Maeder, A., "Comparative study of fixation density maps," *IEEE Transactions on Image Processing* **22**(3), 1121–1133 (2013). http://www.irccyn.ec-nantes.fr/spip.php?article841.

[11] H.R. Sheikh, Z.Wang, L. C. and Bovik, A., "LIVE image quality assessment database release 2." [http://live.ece.utexas.edu/research/quality].

[12] Mancas, M. and Le Meur, O., "Memorability of natural scenes: The role of attention," in [*Proc. International Conference on Image Processing (ICIP)*], (Sept. 15-18, 2013). http://www.tcts.fpms.ac.be/attention/?categorie13/databases.

[13] Isola, P., Xiao, J., Torralba, A., and Oliva, A., "What makes an image memorable?," in [*Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*], 145–152 (2011).

[14] Cerf, M., Harel, J., Einhäuser, W., and Koch, C., "Predicting human gaze using low-level saliency combined with face detection," in [*Proc. Neural Information Processing Systems*], **20**, 241–248 (Dec. 3–8, 2007). http://www.fifadb.com/.

[15] Dorr, M., Martinetz, T., Gegenfurtner, K., and Barth, E., "Variability of eye movements when viewing dynamic natural scenes," *Journal of Vision* **10**(10) (2010). http://www.inb.uni-luebeck.de/tools-demos/gaze.

[16] Le Meur, O., Le Callet, P., Barba, D., and Thoreau, D., "A coherent computational approach to model bottom-up visual attention," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(5), 802–817 (2006). http://www.irccyn.ec-nantes.fr/spip.php?article553.

[17] Wang, J., Chandler, D. M., and Le Callet, P., "Quantifying the relationship between visual salience and visual importance," in [*Proc. SPIE Human Vision and Electronic Imaging*], **7527** (Jan. 17–21, 2010). http://www.irccyn.ec-nantes.fr/spip.php?article494.

[18] van der Linde, I., Rajashekar, U., Bovik, A. C., and Cormack, L. K., "DOVES: A database of visual eye movements," *Spatial Vision* **22**(2), 161–177 (2009). http://live.ece.utexas.edu/research/doves/.

[19] Li, J., Levine, M. D., An, X., Xu, X., and He, H., "Visual saliency based on scale-space analysis in the frequency domain," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**(4), 996–1010 (2013). http://www.cim.mcgill.ca/~lijian/database.htm.

[20] Ehinger, K., Hidalgo-Sotelo, B., Torralba, A., and Oliva, A., "Modelling search for people in 900 scenes: A combined source model of eye guidance," *Visual Cognition* **17**(6/7), 945–978 (2009). http://cvcl.mit.edu/searchmodels/.

[21] Judd, T., Durand, F., and Torralba, A., "Fixations on low-resolution images," *Journal of Vision* **11**(4) (2011). http://people.csail.mit.edu/tjudd/SaliencyBenchmark/.

[22] Bruce, N. D. B. and Tsotsos, J. K., "Saliency based on information maximization," in [*Proc. Neural Information Processing Systems*], **19**, 155–162 (Dec. 4–9, 2006). http://www.cs.umanitoba.ca/~bruce/datacode.html.

[23] Engelke, U., Maeder, A. J., and Zepernick, H.-J., "Visual attention modeling for subjective image quality databases," in [*Proc. Workshop on Multimedia Signal Processing (MMSP)*], (Oct. 5–7, 2009). http://www.sea-mist.se/tek/rcg.nsf/pages/vaiq-db.

Table 2: Eye tracking datasets at a glance (T is viewing time, D is viewing distance, d is screen diagonal, f is frequency).

Dataset	**Year**	**Scenes**	**Resolution**	**Users**	**Age**	T [sec]	D [cm]	d [in]	**Screen**	**Eye Tracker**	f [Hz]	**Restraint**
DUT-OMRON[8]	2013	5172	$<400\times400$	5		2				Tobii X1 Light	$\approx$30	None
FiFA[14]	2007	250	1024×768	7		2	80		CRT	EyeLink 1000	1000	Chin rest
GazeCom Image[15]	2010	63	1280×720	11	18-34	2	45	22	CRT	EyeLink II	250	Chin rest
IRCCyN Image 1[16]	2006	27	$\approx$768×512	40		15			CRT	Cambridge Research	50	
IRCCyN Image 2[17]	2010	80	481×321	18	19-45	15	40	17	LCD	Cambridge Research	50	
IRCCyN LIVE[10]	2013	29	1280×1024	54	18-60	15	70		both	various		
KTH[7]	2011	99	1024×768	31	17-32	5	70	18	CRT	Eyelink I		Headmount
LIVE DOVES[18]	2009	101	1024×768	29	μ =27	5	134	21	CRT	Fourward Tech. Gen. V	200	Bite bar
McGill ImgSal[19]	2013	235	640×480	21			70	17	LCD	Tobii T60	60	
Memorability[12]	2013	135	384×384	17	students	5	65	19		faceLAB 5	60	
MIT CSAIL[5]	2009	1003	$\approx$1024×768	15	18-35	3	61	19				Chin rest
MIT CVCL[20]	2009	912	800×600	14	18-40		75	21	CRT	ISCAN RK-464	240	Head rest
MIT LowRes[21]	2011	1544	1024×860	8	18-55	3	61	19		ETL 400 ISCAN	240	Chin rest
NUSEF[6]	2010	758	1024×860	13	18-35	5	76	17	LCD	ASL	30	
Toronto[22]	2006	120	681×511	20		4	75	21	CRT			
VAIQ[23]	2009	42	varying	15	20-60	12	60	19	LCD	EyeTech TM3		