

Journal of Visualized Experiments

Combining eye-tracking data with an analysis of video content from free-viewing a video of a walk in an urban park environment --Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE58459R3
Full Title:	Combining eye-tracking data with an analysis of video content from free-viewing a video of a walk in an urban park environment
Keywords:	Urban parks Eye-tracking Machine learning Computer Vision Areas of interest Walking Video content analysis
Corresponding Author:	Marco Amati
Corresponding Author's Institution:	
Corresponding Author E-Mail:	marco.amati@rmit.edu.au
Order of Authors:	Marco Amati Chris McCarthy Ebadat Ghanbari Parmehr Jodi Sita
Additional Information:	
Question	Response
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (US\$2,400)
Please indicate the city, state/province, and country where this article will be filmed . Please do not use abbreviations.	RMIT University, Melbourne; Swinburne University Melbourne, Fitzroy Gardens

TITLE:

Combining Eye-tracking Data with an Analysis of Video Content from Free-viewing a Video of a Walk in an Urban Park Environment

AUTHORS AND AFFILIATIONS:

Marco Amati^{1,2}, Chris McCarthy³, Ebadat Ghanbari Parmehr⁴, Jodi Sita⁵

¹Visiting Professor, Dipartimento di Scienze Agro-Ambientali e Territoriali, Università degli Studi di Bari, Italy

² Centre for Urban Research, Royal Melbourne Institute of Technology (RMIT University), Melbourne, Australia

³School of Software and Electrical Engineering, Swinburne University of Technology, John Street, Hawthorn, 3122 Australia

⁴Faculty of Civil Engineering, Babol Noshirvani University of Technology, Babol, Iran

⁵School of Science, Australian Catholic University, 115 Victoria Pde, Fitzroy, Melbourne, Australia

Corresponding Author:

Marco Amati (Marco.Amati@rmit.edu.au)

Tel: +61(3)-9925-9887

Email Addresses of Co-authors:

Chris McCarthy (cdmccarthy@swin.edu.au)

Ebadat Ghanbari Parmehr (parmehr@nit.ac.ir)

Jodi Sita (Jodi.Sita@acu.edu.au)

KEYWORDS:

Urban parks, Eye-tracking, Machine learning, Computer Vision, Areas of interest, Walking, Video content analysis

SUMMARY:

The objective of the protocol is to detail how to collect video data for use in the laboratory; how to record eye-tracking data of participants looking at the data and how to efficiently analyze the content of the videos that they were looking at using a machine learning technique.

ABSTRACT:

As individuals increasingly live in cities, methods to study their everyday movements and the data that can be collected becomes important and valuable. Eye-tracking informatics are known to connect to a range of feelings, health conditions, mental states and actions. But because vision is the result of constant eye-movements, teasing out what is important from what is noise is complex and data intensive. Furthermore, a significant challenge is controlling for what people look at compared to what is presented to them.

The following presents a methodology for combining and analyzing eye-tracking on a video of a natural and complex scene with a machine learning technique for analyzing the content of the

video. In the protocol we focus on analyzing data from filmed videos, how a video can be best used to record participants' eye-tracking data, and importantly how the content of the video can be analyzed and combined with the eye-tracking data. We present a brief summary of the results and a discussion of the potential of the method for further studies in complex environments.

INTRODUCTION:

Our daily lived experiences of urban environments greatly impact on our health and well-being. Our well-being can depend on the amount of green spaces that we view and experience¹⁻³, and these views can be quantified using eye-tracking equipment to guide decision making about park design. However, an issue arises with the volume of eye tracking data that is generated and making sense of this data. As the equipment for recording gaze data in a lab or natural setting becomes easier to use and more powerful, researchers need to consider how we can collect and analyze data validly to help with decision-making questions.

So far, a great deal of eye tracking research has used photographs in a survey or laboratory setting⁴. While this methodology allows for a great deal of reproducibility and control over results, it is unable to take advantage of the latest advances in eye-tracking technology which include the use of video and wearable mobile eye-trackers. Furthermore, we would argue that the act of walking and relaxing is necessarily dynamic particularly when oriented towards a task such as wayfinding⁵. Therefore a fully scientific understanding of these settings should take place outside of the laboratory. However, at the moment, eye-tracking in a real-life naturalistic setting makes comparing the experience between subjects very difficult. For example, if we wanted to compare whether one respondent looks at trees more than another, how could we control for the fact their point of view would be constantly changing compared to others or that their head might have turned. Detailed analysis in these conditions is near impossible with current analysis techniques. We would argue that it is important to control the viewing areas available to the individual being studied and in the analysis to be able to account for the total scene being viewed at any one point in time.

There are a set of theories linking stress levels and perceptions of safety to landscape views and well-evolved measures of stress⁶⁻⁷. There has also been a rapid increase in the sophistication of eye-tracking equipment to measure gaze⁸. Eye-tracking is important because involuntary eye-movements may be more reliably connected to preference, stress and other traditional measures than surveys and intrusive, physiological tests such as salivary cortisol levels. The objective of this research is to develop tools that enable a more precise measurement of eye-tracking data applied to more naturalistic settings, so as to provide further evidence for or refute long-standing landscape theories that have informed park design for decades.

The aim of this project is to develop and test a novel analysis technique that can generate relevant eye-tracking data for different videos of park walk simulations. Our work reported here and elsewhere⁹ represents a half-way point between the naturalistic setting of a fully mobile eye-tracking system and the lab-based photo studies referred to above. In particular, we concentrate on using videos as the stimulus material, exploring how this material can be used to test the amount of fascination that different parks generate in the City of Melbourne. Our work is based

on the assumption that detailed analysis of videos is a necessary step to breach before undertaking a fuller, more naturalistic assessment of the potential of parks to provide restoration from stress.

In this study, we employed a desktop eye-tracker with videos of walks through urban parks and asked participants to imagine that they were taking a relaxing walk through a park. We describe a method to allow the amount of time that participants spent looking at different objects to be comparable between parks. Desktop studies are generally easier to control compared to mobile ET studies and allow comparative analysis of each subject.

Standard eye-tracking software uses a manual area of interest tool in which an operator can manually draw boundaries around objects of interest in each scene. This enables the amount of time that participants spent looking at different objects to be automatically counted. For video data, this process is labour intensive and subject to operator subjectivity and error. In later versions of eye-tracking analysis software, AOIs can automatically track objects across frames when they are the same size in the video. This is an improvement, however, this is only intended to be used for a small number of stimuli in each image and each image must be checked and confirmed.

Manual labelling of objects in an image is common and supported by image editing software such as GNU Image Manipulation Program (GIMP). Given that 1 s produces 30 frames or images, the manual labelling of videos is impractical. In addition, AOI labelling by drawing vector polygons around the edge of complex objects such as tree canopies is very time consuming. Finally, while it is conceivably possible to calculate the size of objects in a field of view using vector labelling, this feature is not currently available.

The method that we report on below deals with these limitations. This study employed automatic labelling of objects. This is possible using an image processing technique known as semantic labelling, in which every pixel in every frame of the video is assigned a label indicating an object class. Machine learning is used to derive pixel classifiers for each object class of interest. These classifiers provide a probabilistic label for each pixel (known as unary potentials), which are then refined in a subsequent optimization process to achieve the final labelled output. These classifiers learn statistical decision boundaries between each object classes in the space of features extracted from the image, including texture, histogram of edge orientations, RGB color values, and the normalized image coordinates. An appropriate technique for this is implemented in the DARWIN machine learning toolbox¹⁰ and is described below.

PROTOCOL:

Ethical approval for this project was given by the Australian Catholic University ethics committee - approval number #201500036E. This ensured that informed consent was gained from all participants and all participants participated voluntarily, and that participants data remained anonymous and confidential. In addition the approval was given due the method and equipment

meeting Australian standards safety regulations.

1. Filming Urban Scenes that Can Be Used in an Eye-Tracking Study

1.1. Use a high-quality digital video camera attached to a gimbal to create a simulated 'walking' or 'sitting' video.

1.2. Record the film with smooth fluid movements in 4k resolution at 25 fps and save them at 1920 x 1080 HD.

1.3. Use a professional filming company to produce digital films if unfamiliar with this or lacking access to equipment.

1.4. Shoot film in one take, to a single length across all videos with precise predefined routes and characteristics.

NOTE: Example video details: Each simulated walk film was 3 min 50 s long and each simulated sitting film was 1 minute long. For both videos in this study, the videos started with a short stair climb leading to a path. The path was followed for exactly 1 min 30 s and then the camera stopped and panned left for 45-60 degrees and stopped for 15 seconds. The camera then continued along the walking path until the 3 min 25 s mark, panned left 40-70 degrees for 15 seconds, continued along the path until the 3 min 50 s mark, then faded to black.

1.5. Ensure that the films are comparable, in other words they do not introduce novel gaze data from different walk trajectories or camera movement. One film was taken of each walk, two in total.

1.6. As eye movements are affected by sound, record a file of generic urban park sounds from one of the walks for the same length of the video without interruptions from sudden loud noises outside of the park (*e.g.*, car alarms). This can be played instead of the existing soundtrack from both videos so that the visual impact of the video can be isolated.

NOTE: Excessive movement of vehicles and other objects can affect eye-movements and skew the results. If some of this intrusion does occur, the film can be played to the participants but the frames may need to be excluded from the analysis.

1.7. If interested in different activities in a walk, such as walking and then sitting on a bench, save or edit the sitting and walking scenes as separate files so as to test them independently.

2. Setup and Desktop Calibration of the Eye-tracking Equipment

2.1. Show the filmed sequences to individuals in a room where natural light can be excluded to avoid reflections on the screen. Show it to them on as large a screen as possible to occupy as much of the visual field, thereby avoiding distraction from outside the field of view. A widescreen

(16:9) PC monitor (22 inch) with participants approximately 60 cm away from the screen is an acceptable standard.

2.2. Perform a calibration to the screen for each participant using the inbuilt calibration tool on the eye-tracking software. Require participants to look at a red ball moving around the screen as part of this and stopping at 5 key calibration points.

2.3. During the calibration observe the participants and ensure that they are not moving their head as well. If they are too close to the screen and the screen is very large this can be a problem. Adjust the distance from the screen accordingly.

2.4. Edit each film to have a white cross situated in the top left corner of the early frames of the video. This can be done using a video editing software. Show this screen for 3 seconds, then begin to play the walk. This is to check the eye-tracking calibration and to ensure that the eye-tracking data collection times could be matched with individual frame numbers.

3. Recruitment and Ethics

3.1. Use a professional research recruitment company to ensure a spread of different genders, ages and occupations among the sample or recruit on site from among the student and staff body.

3.2. Screen participants for known ocular or neurological conditions and/or injuries. They must be asked whether they are taking medications that are known to affect eye movements (*e.g.* benzodiazepines).

3.3. Conduct near vision reading tests and a pen torch eye movement excursion test to demonstrate that they have full eye movement excursions.

3.4. Survey participants about their age, gender and frequency of park use.

4. Participant Experimental Setup

4.1. Ask participants to imagine themselves in need of restoration. Use a sentence that allows them to imagine the context of the eye-tracking video used such as: 'Imagine that it is midday and you are walking alone in Melbourne. You are mentally tired from intense concentration at work and are looking for somewhere to go for a walk, sit down and rest for a little while, before going back to work'.

4.2. Sit participants comfortably in front of the computer screen at a distance of 60–65 cm to view the films.

4.3. Using a spreadsheet program order the films randomly before playing in front of participants.

221
222 4.4. Use speakers to play the same audio for all films.

223
224 4.5. Record the gaze trajectory of the participants using a desktop eye-tracking system. Mount
225 the desktop eye-tracking device as per the manufacturer's instructions. This could be just below
226 the screen on a table, or it could clip to the edge of the screen, on top for example.

227
228 4.6. Use an eye-tracking unit that has a sampling rate of 120 Hz and an accuracy of 0.50°, which
229 allows large freedom of head movements enabling recording during natural head movements. A
230 lower frequency eye-tracker is also acceptable. At 57.3 cm, 1° of visual angle equates to ~1 cm
231 on the screen.

232
233 4.7. Allow participants to watch the footage on a monitor in high definition.

234
235 4.8. Record eye-movements using the eye-tracking software

236
237 4.9. Employ a fixation filter to convert the raw eye sample data for analysis. From the main menu,
238 click on **Tools > Settings**. Select the **Fixation** filters tab. Select the fixation filter according to the
239 manufacturer's specifications on the best type of filter to aggregate the raw eye-tracking data
240 into fixations.

241
242 4.10. If recording eye-tracking data using multiple films, give breaks between recording sessions
243 whenever participants request it.

244 245 **5. Connecting the Viewed Parks to Impressions of the Videos**

246
247 5.1. Compare the extent to which people viewed objects and their opinions of the videos by
248 asking participants to rate each of the parks on a scale of 1–10, firstly for whether they felt they
249 would be able to rest and recover in that environment (1, not very much, to 10, very much) and
250 secondly how much they liked the park (1, not very much, to 10, very much).

251
252 5.2. Ask the participants whether they would use the park to relax or relieve stress (Y/N) and
253 whether they recognized the park they were looking at (Y/N) to control for the potential impact
254 of a prior association with the park.

255
256 5.3. Record short answer responses from the participants to explain their impressions using a
257 voice recorder and then transcribe these.

258 259 **6. Automatic Video Analysis for Area of Interest Extraction**

260 261 **6.1. Selection of AOIs**

262
263 6.1.1. Choose items that are of interest to park designers, urban designers, planners or architects,
264 such as trees, shrubs, signposts, buildings, turf, paths, steps, etc.

6.1.2. For optimal performance and minimal training requirements (discussed further below), use elements that are easily visually distinguishable from each other to the naked eye, and/or consistently occupy different regions of each video frame. In general, sufficient training examples depicting visually distinguishing differences of each AOI should be sufficient for robust performance.

6.2. Training classifiers for AOI extraction

6.2.1. Selection of the training frames to be used, number and rationale

6.2.1.1. Choose an appropriate number of training frames (henceforth referred to as the Training Set). There is no fixed number that is appropriate.

NOTE: The frames must provide sufficient coverage of the range of visual appearance of each object class (*i.e.*, those objects to be labelled) throughout the video. For example, 40 frames out of a 15,000 frame video sequence depicting a single parkland scene were found to achieve sufficient accuracy. If a sequence contains significant variation throughout then more frames may be required.

6.2.1.2. Consider the frame content when choosing training frames. These include: lighting, shapes with respect to what is being classified (*e.g.*, not just one type of tree but a range of tree types, their position in the image, the way they are lit, *etc.*), texture, and color.

6.2.1.3. Consider also the number of frames to include. Specifically, the length and variation of visual conditions exhibited across the video being analyzed, as well as the number of object classes to identify, and the frequency of their appearance.

6.2.2. Manual pixel labelling for training frames

NOTE: Manual labelling the training frames will associate pixels with object classes of interest.

6.2.2.1. Pixel labelling of training frames

6.2.2.1.1. In turn, open each training frame from the video in the image editing software.

6.2.2.1.2. For each training frame from the video, overlay a transparent image layer on the loaded image for labelling and create a color palette, providing one color for each given object class of interest (*i.e.*, AOI).

6.2.2.1.3. Ensure that the color palette and the mapping of color to object classes is the same throughout the labelling process.

6.2.2.1.4. Select the color for the sample AOI.

6.2.2.1.5. Color regions of sample AOIs by selecting, with a mouse click and drag, pixels within the area to “color in” using the appropriate palette choice.

NOTE: It is possible to color large objects imprecisely. However, for narrow/small objects of only a few pixels width, take greater care to ensure that manual labelling accurately captures the object’s visual appearance.

6.2.2.1.6. Once labelling of a frame is complete, export the overlaid layer as a separate image file. Ensure the base file name matches the original frame base file name, but with a “c” appended to the end. For example, if the original frame filename was “1234.png”, then the labeled layer file name should be “1234c.png”.

6.2.2.1.7. Ensure all labeled images are stored in a single folder.

6.2.2.2. Pixel labelling of validation frames

NOTE: To quantitatively validate the accuracy of the trained classifier, an additional set of labeled frames should be created.

6.2.2.2.1. Select frames from the original video sequence not already chosen to be included in the training set. For a 5 minute video, these need not be more than 20 frames, but should be sampled uniformly from across the video sequence to ensure sufficient coverage.

6.2.2.2.2. Label pixels in each frame using the same procedure as outlined for preparing Training Frames (6.2.2). However, this time be as precise and as comprehensive as possible with labelling as this will be used as a ground truth comparison.

6.2.2.2.3. When labelling of a frame is complete, use the same naming convention as for training, however ensure files are saved in a separate validation frames folder.

6.2.3. Automatic pixel labelling of the video sequence

6.2.3.1. Download the DARWIN software library from <http://drwn.anu.edu.au>.

NOTE: The system described in this paper was implemented within a Linux environment, and so the Linux release of DARWIN should be downloaded and installed, following the instructions from the website.

6.2.3.2. Launch the Darwin GUI

6.2.3.3. In the Darwin GUI, click on **Load Training Labels**.

6.2.3.4. Using the file explorer dialogue box that appears, select the folder containing the labeled

training images for the relevant video sequence.

6.2.3.5. Click on **Load Video Frames** and follow the same procedure as 6.2.3.2 to select the folder containing all original frames of the video sequence. This should be a folder of images, in which each frame filename is the number of that frame in the video sequence (e.g., frame 1234 would be named 1234.png)

6.2.3.6. Click the **Train** button. The algorithm will now examine each labeled training frame and learn a model of appearance for classifying pixels into any of the specified object classes of interest.

6.2.4. Validating the trained classifier

6.2.4.1. Once training is complete, click the **Validate Training** button.

6.2.4.2. Using the file explorer dialogue box, select the folder containing all labeled validation images for the relevant video sequence. The trained classifier will now be used to classify pixels in each of the frames referred to in the validation set. Compare this to the ground truth labels provided in the validation folder.

NOTE: Once complete, the GUI will display accuracy statistics, including the percentage of pixels correctly labeled for each frame in the validation set, as well as across the entire validation set.

6.2.4.3. To visually validate the generated labels, click the **Visual Validation** button. If clicked, each generated labeled image is displayed next to the original validation frame.

NOTE: This may be instructive in identifying where mis-classifications are occurring on the object. For example, visual inspection may indicate errors are consistently occurring across a particular part of an object class, suggesting improved labelling in the training set is required. Alternatively, inspection may show that errors occur only at a particular time point in the video, suggesting more training examples are required from that time period in the video.

6.2.4.4. If the accuracy observed in either quantitative or qualitative validation falls below acceptable levels, then include further training examples. In this case, repeat all steps from 6.2.2 to include additional training frames, re-train the classifier following steps in 6.2.3, and re-validate following steps in 6.2.4.

6.2.5. Video pixel labelling

6.2.5.1. Once the classifier training and validation phase is complete, click **Label Frames** on the Darwin GUI to begin the full labelling of all frames in the video sequence using the trained classifier.

6.2.5.2. Follow the prompt to select a destination folder for all output frames, which will be in

the form of labeled images using the same color palette as used in training.

7. Registering the Eye-tracking Data to the Video Content

7.1. In the eye-tracking software click on **File > Export...** Export the eye-tracking file as a CSV file.

7.2. Open the file in a spreadsheet program.

7.3. From the video sequence identify the time at which the white cross on the top left hand of the frame (see 2.3) disappears. Use this time to identify the eye-tracking by using the relevant column on the eye-tracking result. This is the start of the eye-tracking data collection.

7.4. Add a column to the eye-tracking data.

7.5. Using this new column label each row or eye-tracking data point with a frame number starting at the number 1.

8. Displaying the Amount of Time that Participants Examined Different Classes of Objects in the Videos

NOTE: Due to the huge size of the eye-tracking data, Python programming language is better used for steps through to 8.4, although a data processing program can also be used.

8.1. Using the included Python code, calculate the amount of overlap between the objects of interest and the eye-tracking fixation time and frequency.

8.2. Sum this data to understand the amount of time that participants spent looking at different objects.

8.3. Employ a histogram to show the total amount of time that the objects occupied the screen.

8.4. Compare this with the total amount of time that participants looked at different objects.

8.5. To produce a heatmap in the eye-tracking software click on one of the park walk videos using the tick box.

8.6. Click the **Heat Map** tab.

8.7. Adjust the color and other features of the Heat Map as desired using the variables of 'Count' (total number of fixations made over the time window), 'Absolute duration' (accumulated fixation duration), and 'Relative duration' (the amount of time spent looking at an object divided by the amount of time spent looking at a scene).

8.8. Export the image or video that includes the heat map as a JPEG file or video file.

REPRESENTATIVE RESULTS:

Figure 1 and **Figure 2** show the result of taking all eye-tracking data for the whole video across all participants and producing a heat map; this is the standard approach available in eye-tracking software packages. By comparing **Figure 1** and **Figure 2** it is possible to identify that on average participants scanned left and right on the x coordinate of the video in **Figure 1** compared to **Figure 2**, which shows a rounder shape. This is because **Figure 1** had more horizontal elements throughout the video compared to the park in **Figure 2**. The image behind the heat map is a single frame and does not adequately represent the full content of the video.

The desktop eye-tracking system and software only counts results where both eyes can be located at the same time. In other words, if one or both eyes cannot be located the data is counted as lost. In our case eye-tracking data was captured for >80% of the time.

Figure 3 and **Figure 4** show the result of using the protocol and use of machine learning to analyze the content of the videos. **Figure 3** shows the dense vegetation of Fitzroy Gardens compared to the relatively sparse vegetation of Royal Park (**Figure 4**). In the latter, more sky is visible, more of the scenes are dominated by shrub vegetation. **Figure 5** shows the %fixation time on the different objects during the course of the video for one of the participants. It shows that although the path is clearly visible during the course of the video, the participant only looks at this feature occasionally but at key points. Similarly, as **Figure 6** shows, although a tiny fraction of the content of the video in Royal Park pertains to artificial objects, the participant in the study examines these features to a comparatively great extent.

The findings in **Figure 3**, **Figure 4**, **Figure 5**, and **Figure 6** can be summarized in **Figure 7** and **Figure 8** for all 39 participants whose data were used in this study. **Figure 7** shows the dwell time for all participants when looking at objects throughout the length of the video. **Figure 8** shows this same data divided by the amount of time and space that these different objects occupied in the video. A value of 1 indicates that the dwell time can be accounted for by the amount of object in the video. **Figure 7** shows that artificial objects such as street lamps and benches are dwelt on to a greater extent compared to other objects (>1). **Figure 7** also shows that objects that were less pertinent, such as the sky in both images, were viewed comparatively less (<1).

FIGURE AND TABLE LEGENDS:

Figure 1: Example heat map that is produced for the whole video of one of the parks. This shows where the majority of the eye tracks were located. Note the horizontal shape of the heat map because of the dominance of horizontal elements in the video.

Figure 2: Example heat map that is produced for the whole video of another one of the parks. Note the more rounded shape because of the presence of vertical and horizontal elements in the video.

Figure 3: Histogram of content in the video of Fitzroy Gardens analyzed using the machine learning technique.

Figure 4: Histogram of content in the video of Royal Park analyzed using the machine learning technique.

Figure 5: Eye tracking fixation time spent looking at objects in Fitzroy Gardens.

Figure 6: Eye tracking fixation time spent looking at objects in Royal Park.

Figure 7: Aggregated eye tracking dwell times for all participants and objects for both parks.

Figure 8: Aggregated eye tracking dwell times relative to content for all participants and objects for both parks.

DISCUSSION:

Generally, in standard software packages for analyzing eye-tracking data, a vector AOI is used. Even for a single still image, the size of the vector cannot be easily measured. Furthermore, including all AOIs in an image and calculating the relative amounts of AOIs is laborious. It is almost impossible to do this manually on a video without a machine learning technique such as the one described. This was a relatively simple statement that infers a free viewing situation. A much more precise scenario can be used and different scenario descriptions before the same stimulus can also affect eye movements¹¹.

Our research used a method for accurately classifying objects in a video and analyzing the extent to which these were visible to the eye. The method represents one means of accounting for the extent to which eye movements are linked to objects of interest. This enables the extent to which objects of interest are present in a field of view to be controlled when comparing the eye-tracking from different subjects with a mobile eye-tracking device, or different locations when using a desktop-based system. Considerations that can affect the automatic classification of objects using the machine learning we employ include lighting, shapes with respect to what is being classified (*e.g.* not just one type of tree but a range of tree types, their position in the image, the way they are lit *etc.*), texture, and color. Considerations on the number of frames include the length and variation of visual conditions exhibited across the video being analyzed, as well as the number of object classes to identify and the frequency of their appearance. The frames change every 1/25 seconds, but eye-tracking with the software occurs at 1/120 (120 Hz). For this reason eye-tracking data will be collected at 4.8 times the rate at which a new frame appears. It is not possible to label completely accurately but the labelling of the frames should change every 5 times. For example, eye-tracking data during 99 to 101 s has been considered for the frame of 100th second of the video.

A significant advantage of current eye-tracking packages is that they are set up to allow users to review a film of their own eye-tracking fixations and pathways and describe why they looked at certain objects. This results in a qualitative data set that can reveal why individual subjects think they have looked at certain objects. An extension of the project would be to also show them the amount of time they spent looking at objects at different times relative to the content, for

example the information in **Figure 8**. Yet, doing this by controlling for the number of objects in a scene rapidly enough is currently not possible.

For example, participants could be asked to view their own gaze paths that had been recorded and to describe why they had looked at the particular objects¹². In our case, at the end of each film participants were asked to rate each of the parks on a scale of 1–10, firstly for whether they felt they would be able to rest and recover in that environment (1, not very much, to 10, very much) and secondly how much did they like the park (1, not very much, to 10, very much).

ACKNOWLEDGEMENTS:

This work was financially supported by the City of Melbourne and partially by ARC DP 150103135. We would like to thank Eamonn Fennessy for his advice and collaborative approach. With special thanks to researcher assistants Isabelle Janecki and Ethan Chen whom also helped collect and analyze this data. All errors remain the authors.

DISCLOSURES:

None

REFERENCES:

1. Patrik, P., Stigsdotter, U.K. The relation between perceived sensory dimensions of urban green space and stress restoration. *Landscape and Urban Planning*. **94** (3-4), 264-275 (2010).
2. Bjørn, G., Patil, G.G. Biophilia: does visual contact with nature impact on health and well-being? *International Journal of Environmental Research and Public Health*. **6** (9), 2332-2343 (2009).
3. Velarde, Ma D., Fry, G., Tveit, M. Health effects of viewing landscapes—Landscape types in environmental psychology. *Urban Forestry & Urban Greening*. **6** (4), 199-212 (2007).
4. Polat, A. T., Ahmet A. Relationships between the visual preferences of urban recreation area users and various landscape design elements. *Urban Forestry & Urban Greening*. **14** (3), 573-582 (2015).
5. Peter, P., Giannopoulos, I., Raubal, M. Where am I? Investigating map matching during self-localization with mobile eye tracking in an urban environment. *Transactions in GIS*. **18** (5), 660-686 (2014).
6. Berto, R., Massaccesi, S. Pasini, M. Do Eye Movements Measured across High and Low Fascination Photographs Differ? Addressing Kaplan's Fascination Hypothesis. *Journal of Environmental Psychology*. **28** (2). 185-191 (2008).
7. Kaplan, S. The restorative benefits of nature: Towards an integrative framework. *Journal of Environmental Psychology*. **15**, 169-182 (1995).
8. Duchowski, A.T. *Eye Tracking Methodology: Theory and Practice*, 3rd ed. Springer International Publishing. London (2017).
9. Amati, M., Ghanbari Parmehr, E., McCarthy, C., Sita, J. How eye-catching are natural features when walking through a park? Eye- tracking responses to videos of walks? *Urban Forestry and Urban Greening*. **31**, 67-78 (2018)
10. Gould, S. DARWIN: A Framework for Machine Learning and Computer Vision Research and Development, *Journal of Machine Learning Research*. **13** (Dec), 3533–3537 (2012).

- 573 11. Richardson, D., Matlock, T. The integration of figurative language and static
574 depictions: an eye movement study of fictive motion. *Cognition*. **102** (1), 129–138 (2007).
575 12. Bojko, A. *Eye Tracking the User Experience: A Practical Guide to Research*. Rosenfeld, New
576 York (2013).
577

Figure 1

[Click here to access/download;Figure;Fig2AllBM.png](#) 

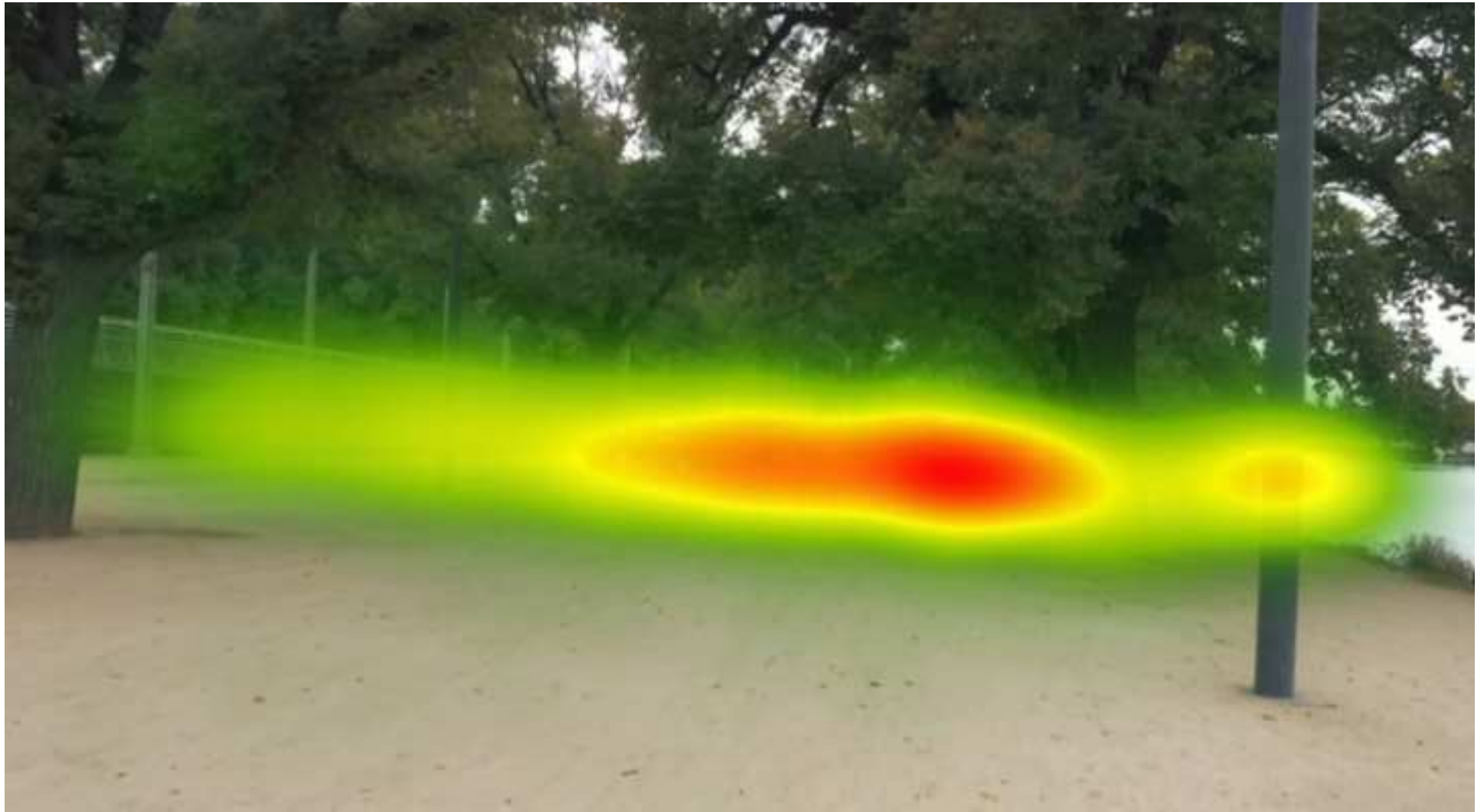


Figure 2

[Click here to access/download;Figure;Fig3AllFG.png](#) 

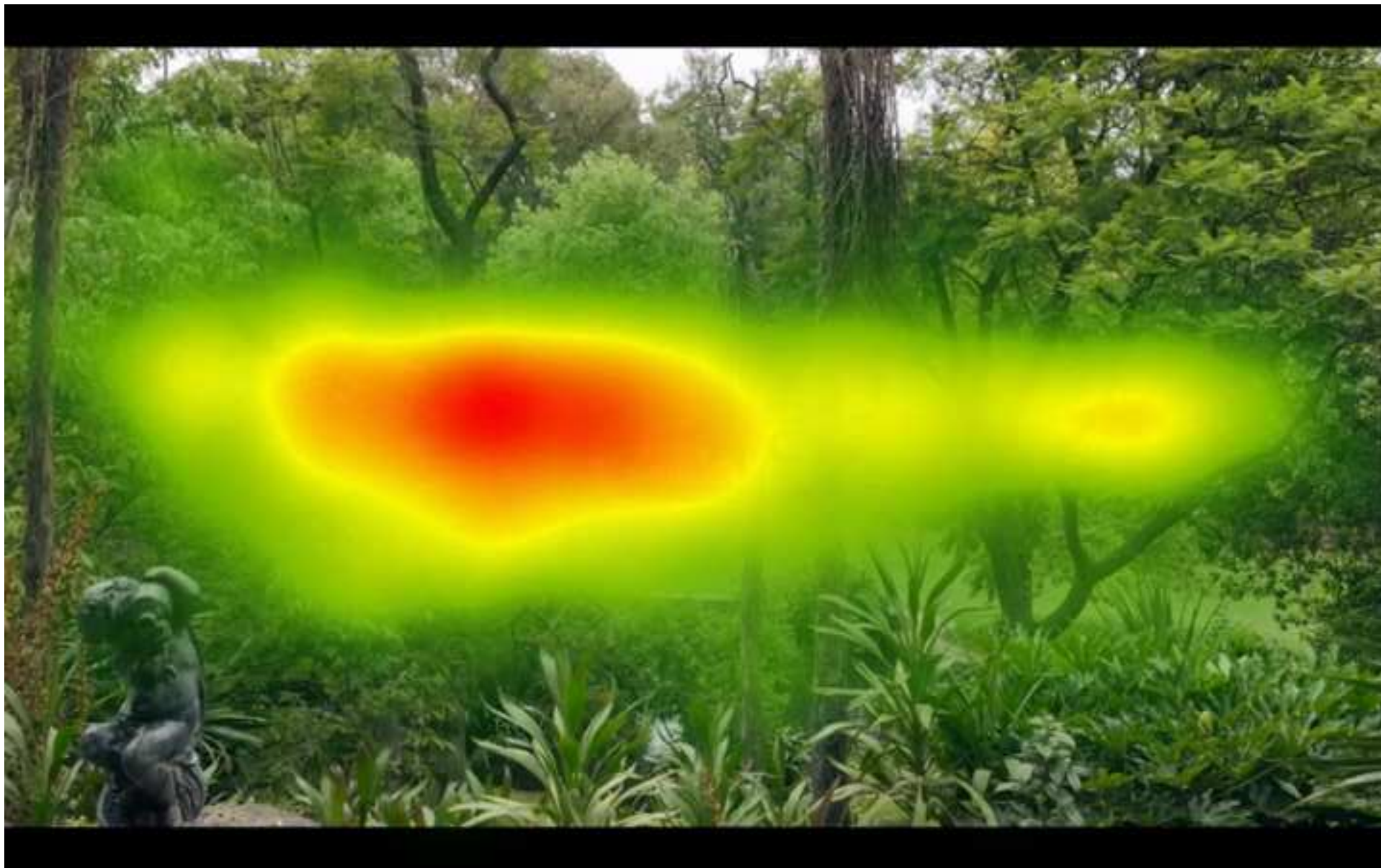


Figure 3

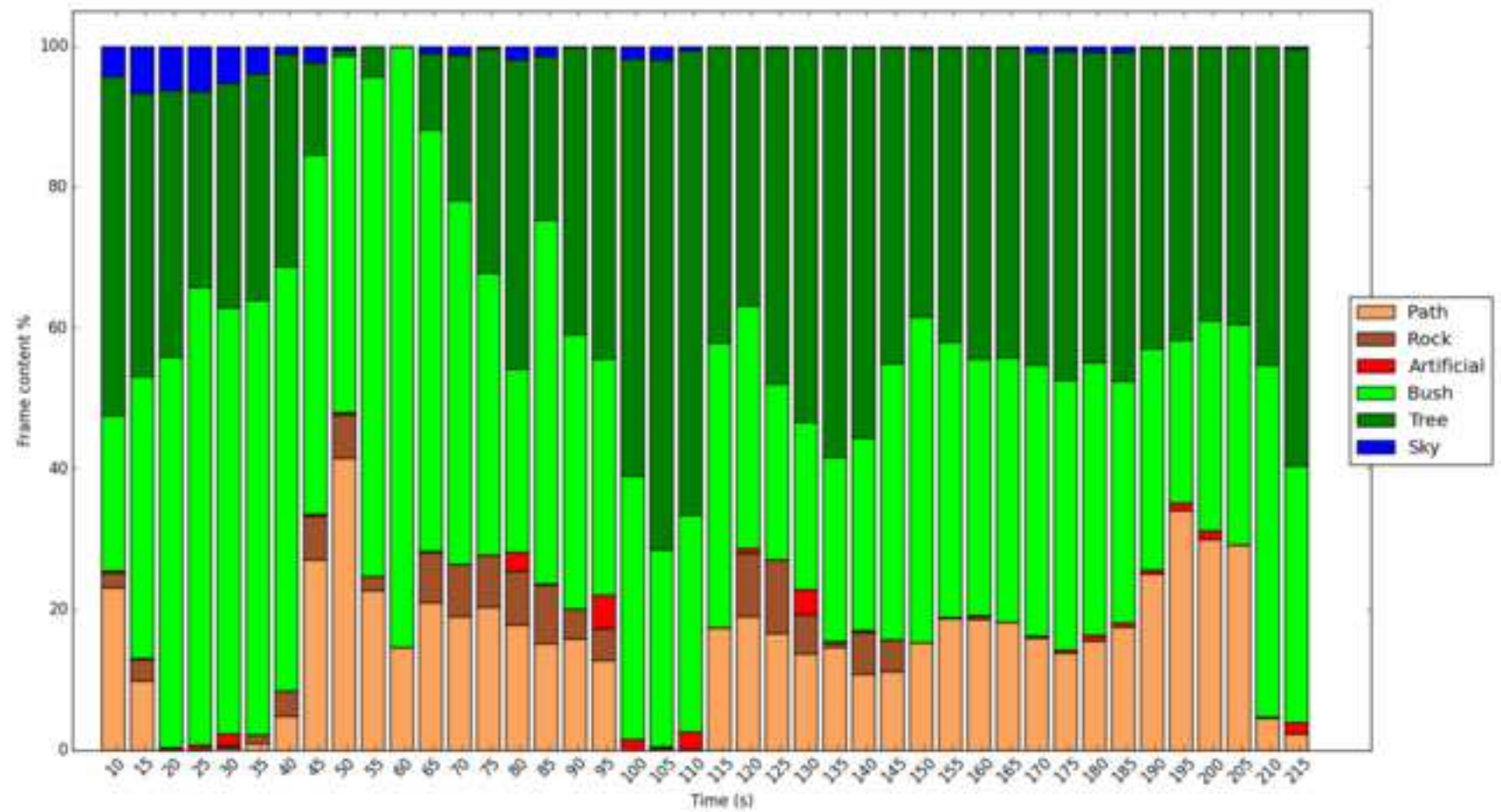


Figure 4

[Click here to access/download;Figure;Fig4RP_new.png](#)

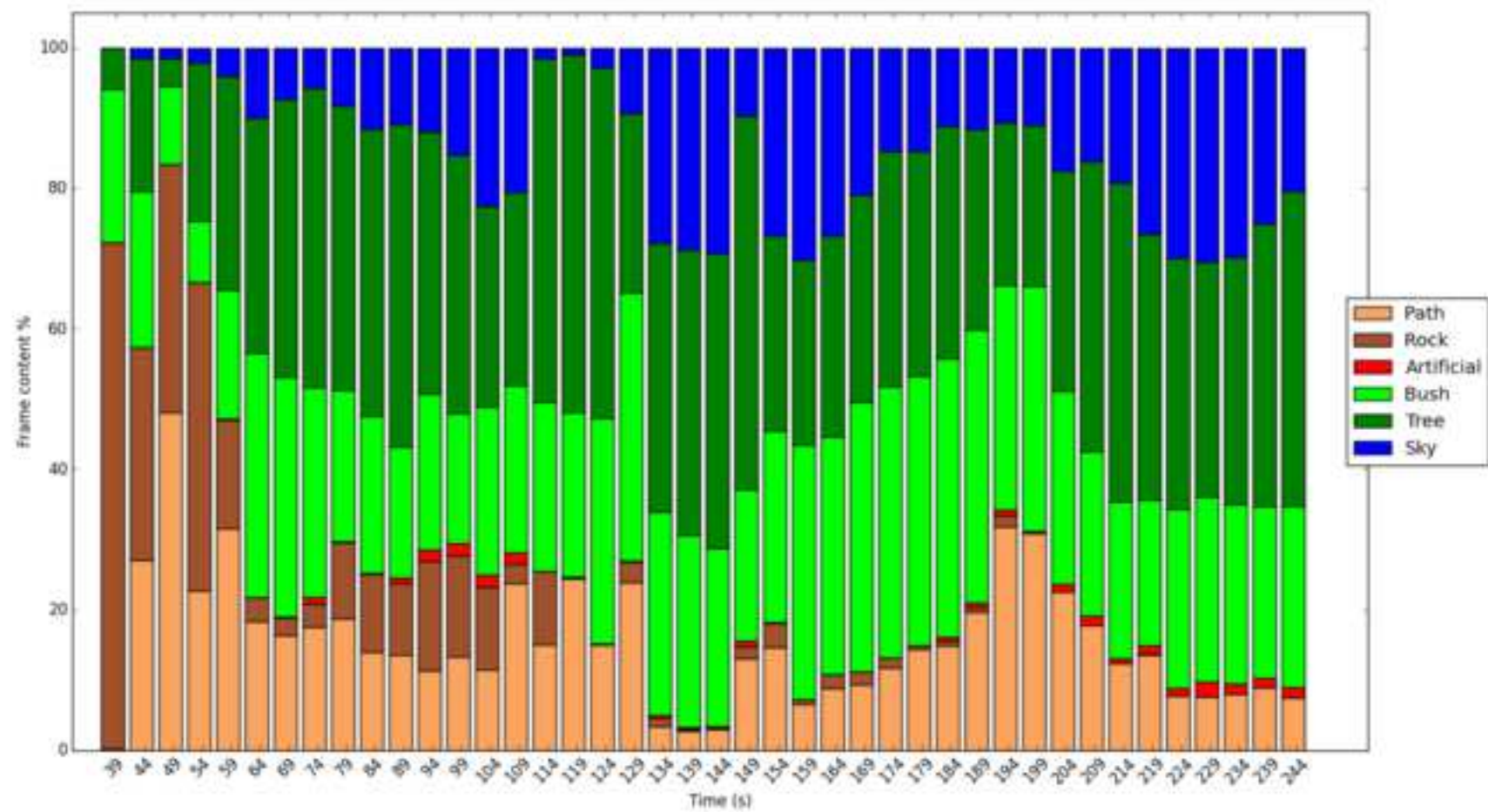


Figure 5

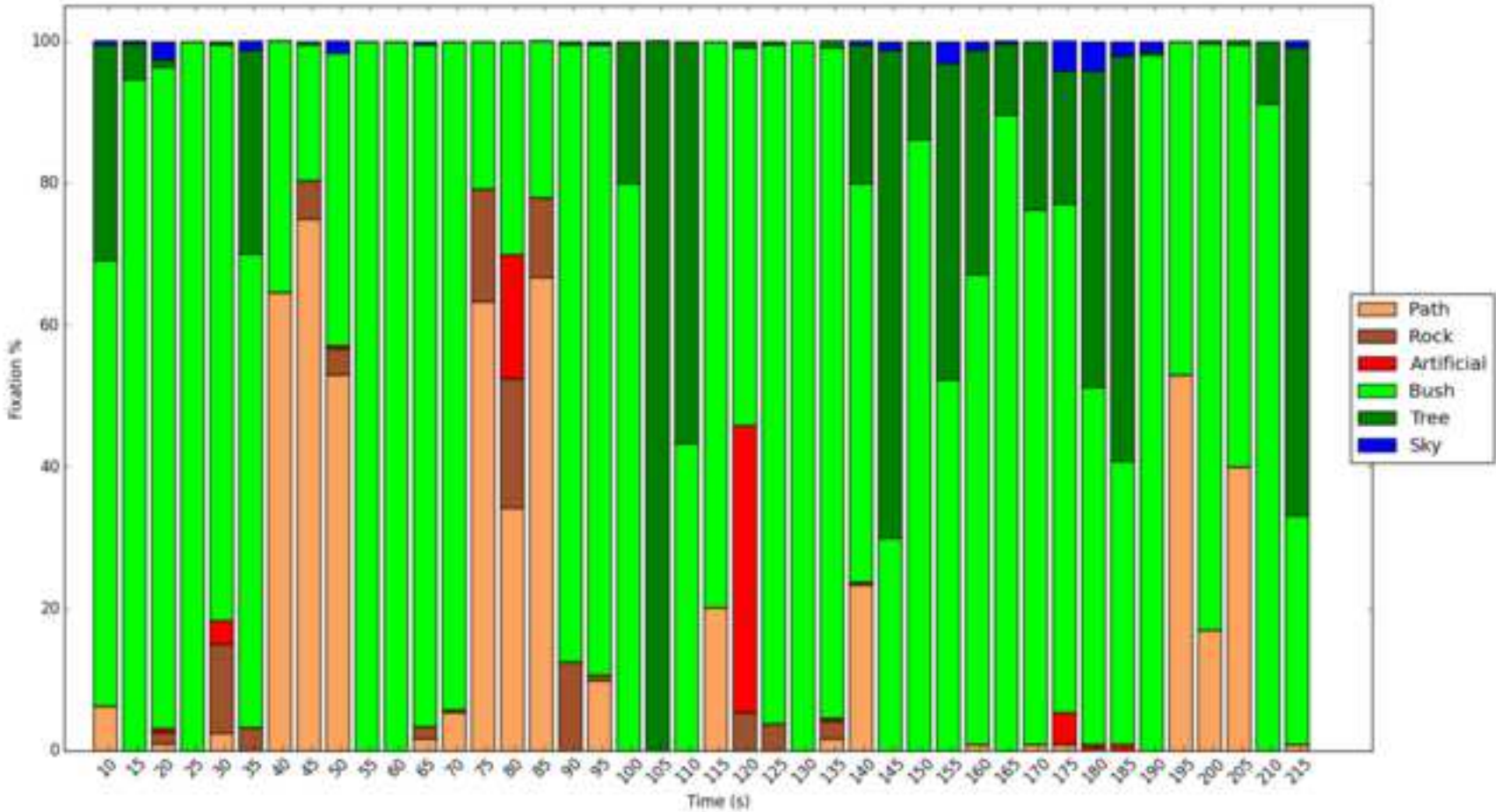


Figure 6

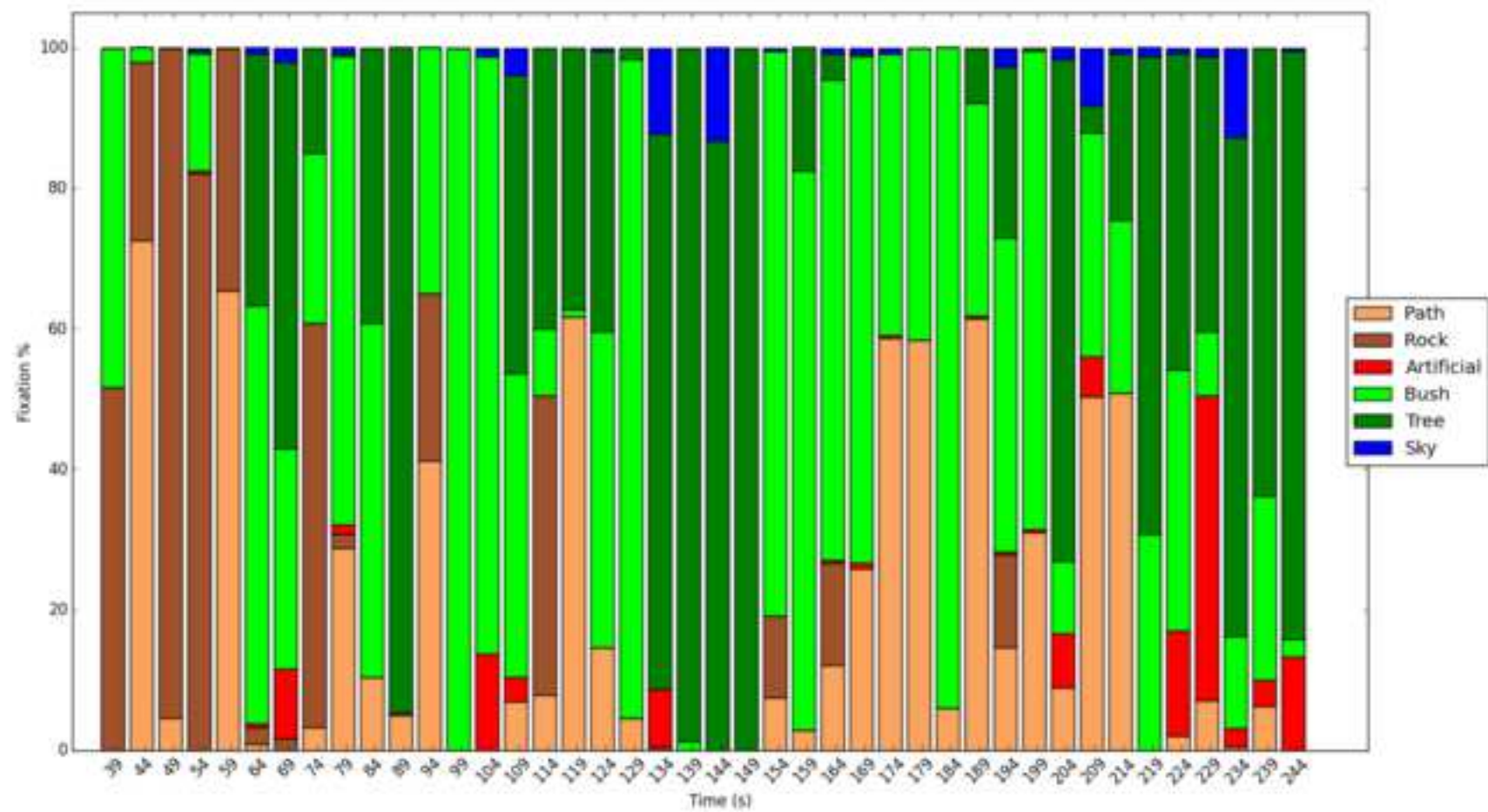


Figure 7

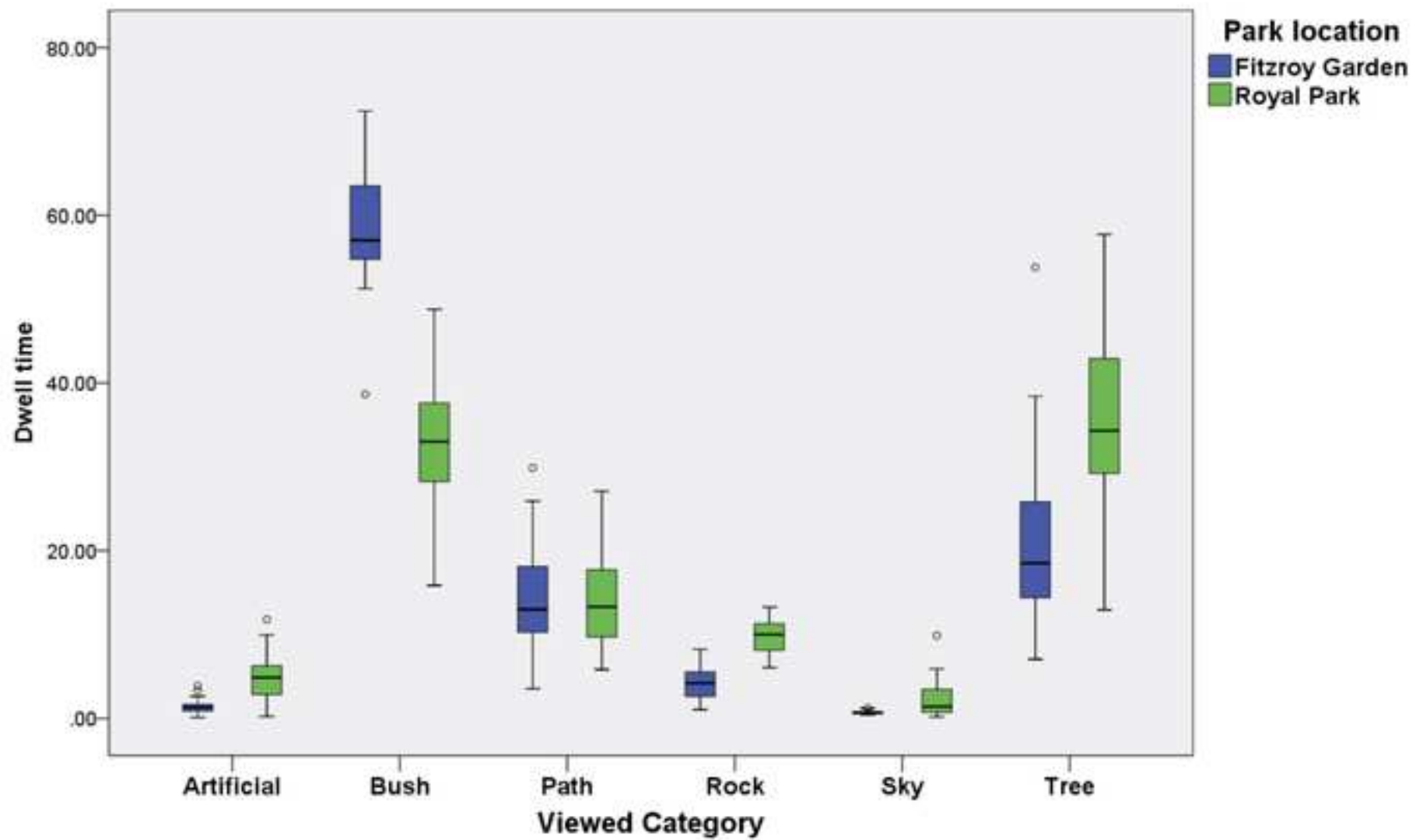
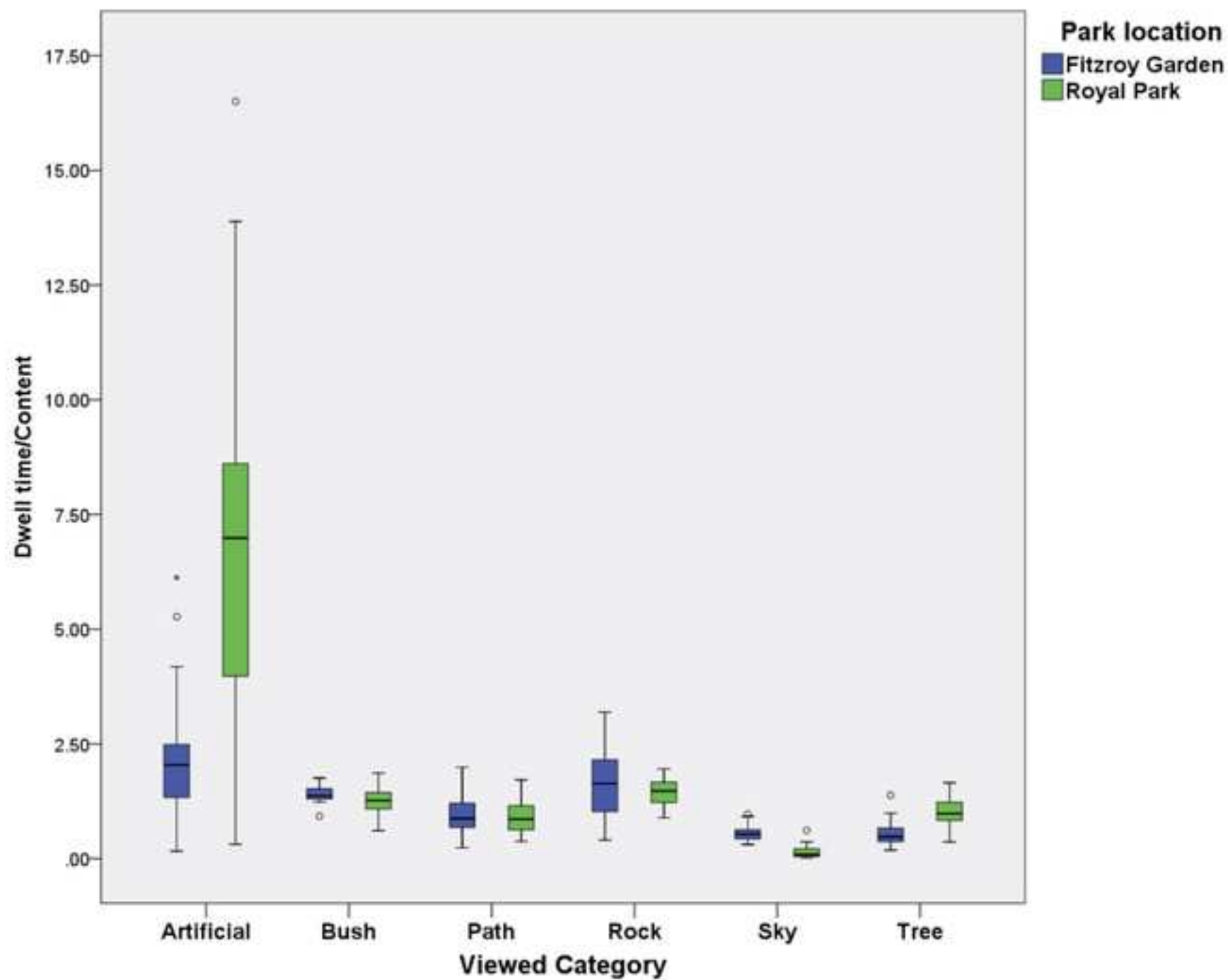
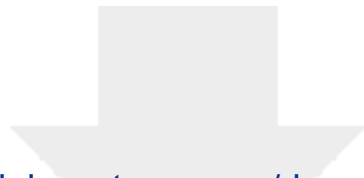


Figure 8

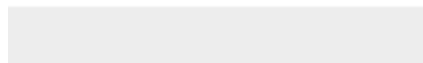




[Click here to access/download](#)

Video or Animated Figure

[Eye_tracking_code_updated.py](#)



Name of Material/ Equipment	Company	Catalog Number	Comments/Description
12 mm lens	Olympus		Lens
Panasonic GH4	Panasonic		Video Camera
Tobii Studio version (2.1.14)	Tobii		Software
Tobii x120 desktop eye-tracker	Tobii		Eye-tracker



1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

Title of Article:

Author(s):

Item 1 (check one box): The Author elects to have the Materials be made available (as described at <http://www.jove.com/author>) via: ☒ Standard Access ☐ Open Access

Item 2 (check one box):

- ☒ The Author is NOT a United States government employee.
- ☐ The Author is a United States government employee and the Materials were prepared in the course of his or her duties as a United States government employee.
- ☐ The Author is a United States government employee but the Materials were NOT prepared in the course of his or her duties as a United States government employee.

ARTICLE AND VIDEO LICENSE AGREEMENT

1. Defined Terms. As used in this Article and Video License Agreement, the following terms shall have the following meanings: “**Agreement**” means this Article and Video License Agreement; “**Article**” means the article specified on the last page of this Agreement, including any associated materials such as texts, figures, tables, artwork, abstracts, or summaries contained therein; “**Author**” means the author who is a signatory to this Agreement; “**Collective Work**” means a work, such as a periodical issue, anthology or encyclopedia, in which the Materials in their entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole; “**CRC License**” means the Creative Commons Attribution-Non Commercial-No Derivs 3.0 Unported Agreement, the terms and conditions of which can be found at: <http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode>; “**Derivative Work**” means a work based upon the Materials or upon the Materials and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Materials may be recast, transformed, or adapted; “**Institution**” means the institution, listed on the last page of this Agreement, by which the Author was employed at the time of the creation of the Materials; “**JoVE**” means MyJoVE Corporation, a Massachusetts corporation and the publisher of *The Journal of Visualized Experiments*; “**Materials**” means the Article and / or the Video; “**Parties**” means the Author and JoVE; “**Video**” means any video(s) made by the Author, alone or in conjunction with any other parties, or by JoVE or its affiliates or agents, individually or in collaboration with the Author or any other parties, incorporating all or any portion of the Article, and in which the Author may or may not appear.

2. Background. The Author, who is the author of the Article, in order to ensure the dissemination and protection of the Article, desires to have the JoVE publish the Article and create and transmit videos based on the Article. In furtherance of such goals, the Parties desire to memorialize in this Agreement the respective rights of each Party in and to the Article and the Video.

3. Grant of Rights in Article. In consideration of JoVE agreeing to publish the Article, the Author hereby grants to JoVE, subject to **Sections 4** and **7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Article in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Article into other languages, create adaptations, summaries or extracts of the Article or other Derivative Works (including, without limitation, the Video) or Collective Works based on all or any portion of the Article and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. If the “Open Access” box has been checked in **Item 1** above, JoVE and the Author hereby grant to the public all such rights in the Article as provided in, but subject to all limitations and requirements set forth in, the CRC License.

ARTICLE AND VIDEO LICENSE AGREEMENT

4. Retention of Rights in Article. Notwithstanding the exclusive license granted to JoVE in **Section 3** above, the Author shall, with respect to the Article, retain the non-exclusive right to use all or part of the Article for the non-commercial purpose of giving lectures, presentations or teaching classes, and to post a copy of the Article on the Institution's website or the Author's personal website, in each case provided that a link to the Article on the JoVE website is provided and notice of JoVE's copyright in the Article is included. All non-copyright intellectual property rights in and to the Article, such as patent rights, shall remain with the Author.

5. Grant of Rights in Video – Standard Access. This **Section 5** applies if the "Standard Access" box has been checked in **Item 1** above or if no box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby acknowledges and agrees that, Subject to **Section 7** below, JoVE is and shall be the sole and exclusive owner of all rights of any nature, including, without limitation, all copyrights, in and to the Video. To the extent that, by law, the Author is deemed, now or at any time in the future, to have any rights of any nature in or to the Video, the Author hereby disclaims all such rights and transfers all such rights to JoVE.

6. Grant of Rights in Video – Open Access. This **Section 6** applies only if the "Open Access" box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby grants to JoVE, subject to **Section 7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Video in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Video into other languages, create adaptations, summaries or extracts of the Video or other Derivative Works or Collective Works based on all or any portion of the Video and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. For any Video to which this Section 6 is applicable, JoVE and the Author hereby grant to the public all such rights in the Video as provided in, but subject to all limitations and requirements set forth in, the CRC License.

7. Government Employees. If the Author is a United States government employee and the Article was prepared in the course of his or her duties as a United States government employee, as indicated in **Item 2** above, and any of the licenses or grants granted by the Author hereunder exceed the scope of the 17 U.S.C. 403, then the rights granted hereunder shall be limited to the maximum rights permitted under such

statute. In such case, all provisions contained herein that are not in conflict with such statute shall remain in full force and effect, and all provisions contained herein that do so conflict shall be deemed to be amended so as to provide to JoVE the maximum rights permissible within such statute.

8. Likeness, Privacy, Personality. The Author hereby grants JoVE the right to use the Author's name, voice, likeness, picture, photograph, image, biography and performance in any way, commercial or otherwise, in connection with the Materials and the sale, promotion and distribution thereof. The Author hereby waives any and all rights he or she may have, relating to his or her appearance in the Video or otherwise relating to the Materials, under all applicable privacy, likeness, personality or similar laws.

9. Author Warranties. The Author represents and warrants that the Article is original, that it has not been published, that the copyright interest is owned by the Author (or, if more than one author is listed at the beginning of this Agreement, by such authors collectively) and has not been assigned, licensed, or otherwise transferred to any other party. The Author represents and warrants that the author(s) listed at the top of this Agreement are the only authors of the Materials. If more than one author is listed at the top of this Agreement and if any such author has not entered into a separate Article and Video License Agreement with JoVE relating to the Materials, the Author represents and warrants that the Author has been authorized by each of the other such authors to execute this Agreement on his or her behalf and to bind him or her with respect to the terms of this Agreement as if each of them had been a party hereto as an Author. The Author warrants that the use, reproduction, distribution, public or private performance or display, and/or modification of all or any portion of the Materials does not and will not violate, infringe and/or misappropriate the patent, trademark, intellectual property or other rights of any third party. The Author represents and warrants that it has and will continue to comply with all government, institutional and other regulations, including, without limitation all institutional, laboratory, hospital, ethical, human and animal treatment, privacy, and all other rules, regulations, laws, procedures or guidelines, applicable to the Materials, and that all research involving human and animal subjects has been approved by the Author's relevant institutional review board.

10. JoVE Discretion. If the Author requests the assistance of JoVE in producing the Video in the Author's facility, the Author shall ensure that the presence of JoVE employees, agents or independent contractors is in accordance with the relevant regulations of the Author's institution. If more than one author is listed at the beginning of this Agreement, JoVE may, in its sole discretion, elect not take any action with respect to the Article until such time as it has received complete, executed Article and Video License Agreements from each such author. JoVE reserves the right, in its absolute and sole discretion and without giving any reason therefore, to accept or decline any work submitted to JoVE. JoVE and its employees, agents and independent contractors shall have

ARTICLE AND VIDEO LICENSE AGREEMENT

full, unfettered access to the facilities of the Author or of the Author's institution as necessary to make the Video, whether actually published or not. JoVE has sole discretion as to the method of making and publishing the Materials, including, without limitation, to all decisions regarding editing, lighting, filming, timing of publication, if any, length, quality, content and the like.

11. **Indemnification.** The Author agrees to indemnify JoVE and/or its successors and assigns from and against any and all claims, costs, and expenses, including attorney's fees, arising out of any breach of any warranty or other representations contained herein. The Author further agrees to indemnify and hold harmless JoVE from and against any and all claims, costs, and expenses, including attorney's fees, resulting from the breach by the Author of any representation or warranty contained herein or from allegations or instances of violation of intellectual property rights, damage to the Author's or the Author's institution's facilities, fraud, libel, defamation, research, equipment, experiments, property damage, personal injury, violations of institutional, laboratory, hospital, ethical, human and animal treatment, privacy or other rules, regulations, laws, procedures or guidelines, liabilities and other losses or damages related in any way to the submission of work to JoVE, making of videos by JoVE, or publication in JoVE or elsewhere by JoVE. The Author shall be responsible for, and shall hold JoVE harmless from, damages caused by lack of sterilization, lack of cleanliness or by contamination due to the making of a video by JoVE its employees, agents or independent contractors. All sterilization, cleanliness or decontamination procedures shall be solely the responsibility of the Author and shall be undertaken at the Author's

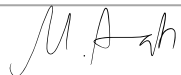
expense. All indemnifications provided herein shall include JoVE's attorney's fees and costs related to said losses or damages. Such indemnification and holding harmless shall include such losses or damages incurred by, or in connection with, acts or omissions of JoVE, its employees, agents or independent contractors.

12. **Fees.** To cover the cost incurred for publication, JoVE must receive payment before production and publication the Materials. Payment is due in 21 days of invoice. Should the Materials not be published due to an editorial or production decision, these funds will be returned to the Author. Withdrawal by the Author of any submitted Materials after final peer review approval will result in a US\$1,200 fee to cover pre-production expenses incurred by JoVE. If payment is not received by the completion of filming, production and publication of the Materials will be suspended until payment is received.

13. **Transfer, Governing Law.** This Agreement may be assigned by JoVE and shall inure to the benefits of any of JoVE's successors and assignees. This Agreement shall be governed and construed by the internal laws of the Commonwealth of Massachusetts without giving effect to any conflict of law provision thereunder. This Agreement may be executed in counterparts, each of which shall be deemed an original, but all of which together shall be deemed to be one and the same agreement. A signed copy of this Agreement delivered by facsimile, e-mail or other means of electronic transmission shall be deemed to have the same legal effect as delivery of an original signed copy of this Agreement.

A signed copy of this document must be sent with all new submissions. Only one Agreement required per submission.

CORRESPONDING AUTHOR:

Name:	Marco Amati	
Department:	GUSS	
Institution:	RMIT	
Article Title:	Combining eye-tracking data with an analysis of video content: a pilot study of a free-viewing field-based study in an urban park environment	
Signature:		Date: 11 May 2018

Please submit a signed and dated copy of this license by one of the following three methods:

- 1) Upload a scanned copy of the document as a pdf on the JoVE submission site;
- 2) Fax the document to +1.866.381.2236;
- 3) Mail the document to JoVE / Attn: JoVE Editorial / 1 Alewife Center #200 / Cambridge, MA 02139

For questions, please email submissions@jove.com or call +1.617.945.9051

RMIT University
Marco.amati@rmit.edu.au

07 October 2018

Dear Phillip,

Please find attached our revised submission which has been re-titled: **Combining eye-tracking data with an analysis of video content from free-viewing a video-ed walk in an urban park environment**

In the submission pack you will find an excel table with a list of comments responded to. Thank you again for your patience and support.

Kind regards,
Marco

Editorial comments:

Changes to be made by the Author(s) regarding the written manuscript:

1. Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammar issues.
2. Please obtain explicit copyright permission to reuse any figures from a previous publication. Explicit permission can be expressed in the form of a letter from the editor or a link to the editorial policy that allows re-prints. Please upload this information as a .doc or .docx file to your Editorial Manager account. The Figure must be cited appropriately in the Figure Legend, i.e. "This figure has been modified from [citation]."
3. Please revise lines 135-136, 170-171, 173-175, 182-185, 192-194, 204-209, and 388-390 to avoid previously published text.
4. Please revise the title to avoid punctuation.

5. JoVE cannot publish manuscripts containing commercial language. This includes trademark symbols (™), registered symbols (®), and company names before an instrument or reagent. Please remove all commercial language from your manuscript and use generic terms instead. All commercial products should be sufficiently referenced in the Table of Materials and Reagents. You may use the generic term followed by "(see table of materials)" to draw the readers' attention to specific commercial names. Examples of commercial sounding language in your manuscript are: Panasonic GH4, Olympus, Nebula Lite, <https://www.bakewood.com.au>, Microsoft Excel, Tobii Technology, Tobii Studio, etc.

6. Please include an ethics statement before the numbered protocol steps, indicating that the protocol follows the guidelines of your institution's human research ethics committee.
7. Please revise the protocol text to avoid the use of any personal pronouns (e.g., "we", "you", "our" etc.).
8. Please revise the protocol to contain only action items that direct the reader to do something (e.g., "Do this," "Ensure that," etc.). The actions should be described in the imperative tense in complete sentences wherever possible. Avoid usage of phrases such as "could be," "should be," and "would be" throughout the Protocol. Any text that cannot be written in the imperative tense may be added as a "Note." Please include all safety procedures and use of hoods, etc. However, notes should be used sparingly and actions should be described in the imperative tense wherever possible.
9. The Protocol should contain only action items that direct the reader to do something. Please move the discussion about the protocol to the Discussion.
10. Please add more details to your protocol steps. There should be enough detail in each step to supplement the actions seen in the video so that viewers can easily replicate the protocol. Please ensure you answer the "how" question, i.e., how is the step performed? Alternatively, add references to published material specifying how to perform the protocol action. See examples below:
 - 1.4: Please specify the standard length and predefined routes. It is unclear what is filmed here.
 - 1.5: How many films are taken?
 - 1.6: Where are the generic urban sounds recorded, and for how long?
 - 1.8: It is unclear how the sitting and walking scenes are recorded. Please describe.
 - 2.1: Are the participants watching the filmed sequences individually or as a group? What filmed sequences are showing in this step?
 - 2.2: Please describe the calibration step in the imperative tense.
 - 2.3: How to situate a white cross? What is a park video?
 - 4.5: Where is the desktop system set up (distance from the participant, etc.)?
 - 4.6: Where is the eye-tracking unit set up?

- 4.9: Please describe how this is done. Software steps must be more explicitly explained ('click', 'select', etc.).
11. 6.2, 8 and their sub-steps: To be filmed, software must have a GUI (graphical user interface) and software steps must be more explicitly explained ('click', 'select', etc.). Please add more specific details (e.g. button clicks for software actions, numerical values for settings, etc.) to your protocol steps.
12. 8.2: Can the Python code be provided as a Supplemental Coding File?
13. Can the protocol describe how to produce a heat map as the results show such data?
14. Please include single-line spaces between all paragraphs, headings, steps, etc.
15. After you have made all the recommended changes to your protocol (listed above), please highlight 2.75 pages or less of the Protocol (including headings and spacing) that identifies the essential steps of the protocol for the video, i.e., the steps that should be visualized to tell the most cohesive story of the Protocol.
16. Please highlight complete sentences (not parts of sentences). Please ensure that the highlighted part of the step includes at least one action that is written in imperative tense.
17. Please include all relevant details that are required to perform the step in the highlighting. For example: If step 2.5 is highlighted for filming and the details of how to perform the step are given in steps 2.5.1 and 2.5.2, then the sub-steps where the details are provided must be highlighted.
18. Figures 3-6: Please change "sec" to "s".
19. Discussion: As we are a methods journal, please also discuss critical steps within the protocol, any modifications and troubleshooting of the technique, and any limitations of the technique.
20. Please revise the table of the essential supplies, reagents, and equipment to include the name, company, and catalog number of all relevant materials in separate columns in an xls/xlsx file.

Reviewers' comments:

Reviewer #1:

Manuscript Summary:

This paper presents a valuable method for analysing eye-tracking data that has potential to improve the evidential output from research exploring gaze behaviour. The method attempts to address a significant issue or limitation with eye-tracking research, particularly research that takes place in real-world settings - the issue of variable exposure to stimuli. I recommend publishing this article, although have a number of suggestions and queries that may want addressing to improve the article's quality and readability.

Major Concerns:

[General comments]

-The Introduction would benefit from additional references to support some of the statements made, for example in relation to the link between well being and green spaces (lines 58-60), the need for additional eye-tracking studies outside the laboratory (lines 68-70); and background about previous eye-tracking studies and illustrations of the problems such previous studies have, e.g. in relation to exposure to different stimuli.

- In the Results section, it would be useful to know the extent of any data loss, e.g. where ET signal is lost or is suspect, and how this is dealt with. E.g. if 15% of video frames did not have ET data, were these frames excluded from the calculations about video content, i.e. amount of time/area each of the categories were visible for? More generally, it would be useful to discuss loss of ET data and how this impacts on / should be dealt with for the methods proposed in the paper. Loss of ET data is a significant issue particularly for real-world, outdoor ET research, and it is exactly this type of research that the method proposed in this paper will be extremely useful for, as it helps account for variations in the visual stimuli experienced by each individual participant.
- The plots presented in the paper should be in higher resolution - some axes titles and labels are unreadable.
- Use consistent names for the different categories of environmental feature - e.g. 'path' and 'road' are both used to refer to the same feature.
- It is not clear exactly what the authors are trying to show with Figures 7 and 8. Plots showing summary values across all 39 participants, for the two different parks, would potentially be more useful, and more illustrative in demonstrating the value of accounting for video content. For example, a plot showing the average ratio of fixation frequency or dwell time to total video content for each feature, for the two different parks, would be useful.

Minor Concerns:

[Specific comments]

- Lines 115-117 - would be useful to elaborate more about these two points related to limitations with AOI labelling - it is not clear why the two issues described (missing finer edges, and calculation of the size of objects in the field of view) are indeed limitations/
- Line 126 - mentions "17-dimensional filter bank" - not clear what this is, to a reader who is not familiar with machine learning methods.
- Line 151 - refers to sitting and walking scenes but I do not see mention of this previously and it is not clear what this refers to. Were the videos recorded from two positions, sitting and standing, and why was this done if it happened?
- Line 155-156 - suggests a screen as large as possible should be used. If a screen is particularly large, it may result in head movements as well as eye-movements - does this complicate the analytical method proposed in this paper or does it not matter?
- Was wearing spectacles acceptable for taking part in the study, i.e. could the eye-tracker cope with spectacles?
- Lines 181-185 - I am dubious about whether asking someone to imagine a certain scenario can lead to realistic eye movements that reflect this imaginary situation. Was there any way of checking the validity of using this method of an imaginary scenario?
- Line 192 - states that a 120 Hz eye-tracker is used - would be useful to confirm whether such a high frequency sampling rate is required as part of the method described in this paper, or if a lower frequency would be ok. Also on this line, refers to an eye-tracker accuracy of 0.50 - should state what units this is in (presumably degrees?)
- Line 214 - 'Automatically', should this be 'Automatic'?
- Lines 218-220 - states that the elements forming the different categories of item within the environment should be 'visually distinguishable' from and have a 'consistently different position' from each other. This implies the categories are chosen based largely on the analytical method being used, rather than based on theory. This seems a limitation.
- Line 225 - states that there is no fixed number of frames to use for training that is appropriate. Is it possible to give an indication of the range of proportion of frames that might be required for training, e.g. what proportion of frames were used in the example analysis provided in the paper?

- Line 278-279 - This may be a comment from one of authors that has accidentally been left in the submitted manuscript
- Line 295 - state that the sampling rate of the eye-tracker is 60 Hz, but on line 192 the sampling rate is stated as 120 Hz
- Line 296-297 - 'It is not possible to label completely accurately but the labelling should change every 2.5 times'. It is not very clear what this sentence means. I think it refers to the fact that there will be approximately 2.5 eye-tracking samples per frame, but it would be helpful to make this clearer.
- Line 301-302 - It is not clear why 3 minutes of eye-tracking will produce a million rows of data. Even assuming a sampling rate of 120 Hz, there would be $120 \times 60 \times 3 = 21,600$ samples of eye-tracking.
- Lines 305 - refers to Python code - is it possible to make this Python code openly available?
- Lines 315-317 - I am not sure I follow this interpretation of the two heat maps. Reference to 'horizontal elements' makes me think the authors are talking about the extent of scanning left and right (i.e. x coordinate of the gaze position), but it is not apparent to me that Figure 2 has more of this than Figure 1. Further explanation would be useful.
- Line 323 - 'in the more the scenes' - does not seem to make sense
- Lines 330-342 - This paragraph seems to suggest the measures shown in figures 7 and 8 are the absolute value of frequency or dwell time (e.g. 50 fixations, or 6 seconds of dwell time), divided by the % of content for that category. However, it may be more informative (and it seems that this is what the two plots show) if the numerator is actually the proportion of fixations or total dwell time that are towards the particular category. Therefore, a value of 1 would indicate that the proportion of time spent looking at a category (or proportion of fixations towards a category) was equal to the proportional size of that category in the video. This would be a useful threshold, as it would indicate looking behaviour was in line with exposure to the particular category.
- Line 375 - 'controlled' seems like the wrong word - participants' exposure to different stimuli is not being controlled (unless this is done through an experimental manipulation). Perhaps 'accounted for' would be more appropriate?
- Line 377-378 - Not clear what this finale sentence of the paragraph means, or where the 30% of eye movements can be explained" statement comes from - further explanation would be helpful.
- Lines 388-392 - I do not follow the point being made in this paragraph. Were participants asked to rate the parks after they viewed their own ET data? I am also dubious about the usefulness of asking participants, after the fact, to explain why they looked where they did. My suspicion is they will often post-rationalise, and objective information that may explain their eye movements may be difficult to obtain. For example, there is evidence to show that people are not good at knowing where they have looked (e.g. Clarke, A. D., Mahon, A., Irvine, A., & Hunt, A. R. (2017). People are unable to recognize or report on their own eye movements. *The Quarterly Journal of Experimental Psychology*, 70(11), 2251-2270.). If they are less aware of where they looked, it also follows that they may not have an objective justification for looking.

Reviewer #2:

Inclusion of references

Minor Concerns:

1. All the result visualization is a raster graphics. Please use vector graphics (EPS) for better visibility.
2. Protocol 6.2.4.1.5: Please remove the texts inside brackets.
3. Protocol 8: "programming language is better used for 7 and for ..." I believe the authors meant "I

Response	Status
MA	OK
MA	OK
MA	All but 388-390 has been changed. Couldn't find what was copied here.
MA	OK
MA	OK
JS	OK
MA	OK
MA	OK
MA	OK
MA	OK
JS	OK - this is in the text
MA	OK
MA	OK
MA	OK
JS	OK. It is noted that participants are seated 60-65cm from the viewing screen
JS	OK

CMcC	OK
------	----

CMcC	OK
EGP	OK
JS	OK
MA	OK

MA	OK
----	----

MA	OK
----	----

MA	OK
MA	OK

MA	OK
----	----

MA	OK
----	----

MA	OK
----	----

JS OK

MA OK

MA This has been made consistent throughout

MA The alternative I have obtained rights to reuse the figure from the UFUG article

MA OK

CMcC This has been deleted

MA OK This has been dealt with through editorial comments

JS OK Comment inserted 2.4

JS OK

JS OK some small adjustment to the text has been made

JS OK

JS OK

CMcC OK This wording of this statement suggests a more stringent limitation than

CMcC OK We have now included text providing, as an example, how many traits

MA OK

JS **** need EGP to confirm if the RAW data was in 60 or 120HZ- it can record in BO1

EGP OK

MA This has beenI guess this is right
EGP Provided

MA OK This has been provided. The heat maps are different
MA OK

MA OK We have included a figure from a previous publication with rights.

MA OK

CmcC OK - deleted

MA OK - we agree with this point and thank the reviewers for the text

MA OK

MA OK
MA OK
MA OK

han is actually the case for this application. The only limitation imposed by the technique is that ,

ining frames were needed to process our video sequences

ΓH rates. He will know from the RAW data and we need to go with the HZ he used for the next pa

Walk 7 & 8 & 9 – Royal Park West

Landscape features – native and indigenous vegetation, unkempt grass habitat area, wetland environs.





Walk 5 – Fitzroy Gardens

Walk landscape features –large shade trees, lawn, presence of
bodies/ponds.



AOIs be visually distinguishable to the human eye, with the assumption that sufficient training exa

rt of the data analysis.

assland

of water

mples exist to allow the software to discern these differences. Optimal performance with minim

ial training requirements is achieved when the AOIs are very obviously visually different. More su

Subtle differences requires more training examples. The number of training examples required is r

not directly calculable, however, our experience has shown that if the object classes are obvious

ly different in appearance, and generally occupy distinctly different parts of the image, then the

number of training examples depicting those types of objects may be as low as 40 examples. W

/e further note that in the context of eye-tracking coding, there is generally the implicit assumpt

ion that AOIs selected are visually and/or spatially distinguishable in or

ELSEVIER LICENSE TERMS AND CONDITIONS

Sep 15, 2018

This Agreement between RMIT University -- Marco Amati ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4430550146552
License date	Sep 15, 2018
Licensed Content Publisher	Elsevier
Licensed Content Publication	Urban Forestry & Urban Greening
Licensed Content Title	How eye-catching are natural features when walking through a park? Eye-tracking responses to videos of walks
Licensed Content Author	Marco Amati,Ebadat Ghanbari Parmehr,Chris McCarthy,Jodi Sita
Licensed Content Date	Apr 1, 2018
Licensed Content Volume	31
Licensed Content Issue	n/a
Licensed Content Pages	12
Start Page	67
End Page	78
Type of Use	reuse in a journal/magazine
Requestor type	author of new work
Intended publisher of new work	AIP Publishing
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	2
Format	electronic
Are you the author of this Elsevier article?	Yes
Will you be translating?	No
Original figure numbers	Figure 5 and 6
Title of the article	Combining eye-tracking data with an analysis of video content: a pilot study of a free-viewing field-based study in an urban park environment
Publication new article is in	JoVE
Publisher of the new article	AIP Publishing
Author of new article	Marco Amati, Chris McCarthy" cdmccarthy@swin.edu.au, "Ebadat Ghanbari Parmehr" egparmehr@gmail.com, "Jodi Sita" jodi.sita@acu.edu.au
Expected publication date	Jan 2019

Estimated size of new article 10
(number of pages)

Requestor Location RMIT University
124 Latrobe St

Melbourne, Victoria 3000
Australia
Attn: RMIT University

Publisher Tax ID GB 494 6272 12

Total 0.00 AUD

Terms and Conditions

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at <http://myaccount.copyright.com>).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com). No modifications can be made to any Lancet figures/tables and they must be reproduced in full.

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either

by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. **Warranties:** Publisher makes no representations or warranties with respect to the licensed material.

10. **Indemnity:** You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. **No Transfer of License:** This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. **No Amendment Except in Writing:** This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. **Objection to Contrary Terms:** Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. **Revocation:** Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation:** This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.

16. **Posting licensed content on any Website:** The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at <http://www.sciencedirect.com/science/journal/xxxxx> or the Elsevier homepage for books at <http://www.elsevier.com>; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at <http://www.elsevier.com> . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.

17. For journal authors: the following clauses are applicable in addition to the above:

Preprints:

A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).

Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).

If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.

Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.

Authors can share their accepted author manuscript:

- immediately
 - via their non-commercial person homepage or blog
 - by updating a preprint in arXiv or RePEc with the accepted manuscript
 - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
 - directly by providing copies to their students or to research collaborators for their personal use
 - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- After the embargo period
 - via non-commercial hosting platforms such as their institutional repository
 - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all

value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.

Policies for sharing publishing journal articles differ for subscription and gold open access articles:

Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.

Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a [CrossMark logo](#), the end user license, and a DOI link to the formal publication on ScienceDirect.

Please refer to Elsevier's [posting policy](#) for further information.

18. **For book authors** the following clauses are applicable in addition to the above:

Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.

19. **Thesis/Dissertation:** If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

Elsevier Open Access Terms and Conditions

You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our [open access license policy](#) for more information.

Terms & Conditions applicable to all Open Access articles published with Elsevier:

Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated.

The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.

If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

Additional Terms & Conditions applicable to each Creative Commons user license:

CC BY: The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the

user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. The full details of the license are available at <http://creativecommons.org/licenses/by/4.0>.

CC BY NC SA: The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at <http://creativecommons.org/licenses/by-nc-sa/4.0>.

CC BY NC ND: The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at <http://creativecommons.org/licenses/by-nc-nd/4.0>. Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee.

Commercial reuse includes:

- Associating advertising with the full text of the Article
- Charging fees for document delivery or access
- Article aggregation
- Systematic distribution via e-mail lists or share buttons

Posting or linking by commercial companies for use by customers of those companies.

20. Other Conditions:

v1.9

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.