

Video Article

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

Marco Amati<sup>1,2</sup>, Chris McCarthy<sup>3</sup>, Ebadat Ghanbari Parmehr<sup>4</sup>, Jodi Sita<sup>5</sup>

<sup>1</sup>Visiting Professor, Dipartimento di Scienze Agro-Ambientali e Territoriali, Università degli Studi di Bari

<sup>2</sup>Centre for Urban Research, Royal Melbourne Institute of Technology (RMIT University)

<sup>3</sup>School of Software and Electrical Engineering, Swinburne University of Technology

<sup>4</sup>Faculty of Civil Engineering, Babol Noshirvani University of Technology

<sup>5</sup>School of Science, Australian Catholic University

Correspondence to: Marco Amati at [Marco.Amati@rmit.edu.au](mailto:Marco.Amati@rmit.edu.au)

URL: <https://www.jove.com/video/58459>

DOI: [doi:10.3791/58459](https://doi.org/10.3791/58459)

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

Date Published: 10/26/2018

Citation: Amati, M., McCarthy, C., Parmehr, E.G., Sita, J. Combining Eye-tracking Data with an Analysis of Video Content from Free-viewing a Video of a Walk in an Urban Park Environment. *J. Vis. Exp.* (), e58459, doi:10.3791/58459 (2018).

## 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 experience<sup>1,2,3</sup>, 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 setting<sup>4</sup>. 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 wayfinding<sup>5</sup>. 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 stress<sup>6,7</sup>. There has also been a rapid increase in the sophistication of eye-tracking equipment to measure gaze<sup>8</sup>. 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 elsewhere<sup>9</sup> 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 toolbox<sup>10</sup> 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. Use a high-quality digital video camera attached to a gimbal to create a simulated 'walking' or 'sitting' video.
2. Record the film with smooth fluid movements in 4k resolution at 25 fps and save them at 1920 x 1080 HD.
3. Use a professional filming company to produce digital films if unfamiliar with this or lacking access to equipment.
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.
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.
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.
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

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. 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.
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.
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

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.

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. Conduct near vision reading tests and a pen torch eye movement excursion test to demonstrate that they have full eye movement excursions.
4. Survey participants about their age, gender and frequency of park use.

## 4. Participant Experimental Setup

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'.
2. Sit participants comfortably in front of the computer screen at a distance of 60–65 cm to view the films.
3. Using a spreadsheet program order the films randomly before playing in front of participants.
4. Use speakers to play the same audio for all films.
5. Record the gaze trajectory of the participants using a desktop eye-tracking system. Mount the desktop eye-tracking device as per the manufacturer's instructions. This could be just below the screen on a table, or it could clip to the edge of the screen, on top for example.
6. Use an eye-tracking unit that has a sampling rate of 120 Hz and an accuracy of 0.50°, which allows large freedom of head movements enabling recording during natural head movements. A lower frequency eye-tracker is also acceptable. At 57.3 cm, 1° of visual angle equates to ~1 cm on the screen.
7. Allow participants to watch the footage on a monitor in high definition.
8. Record eye-movements using the eye-tracking software
9. Employ a fixation filter to convert the raw eye sample data for analysis. From the main menu, click on **Tools > Settings**. Select the **Fixation** filters tab. Select the fixation filter according to the manufacturer's specifications on the best type of filter to aggregate the raw eye-tracking data into fixations.
10. If recording eye-tracking data using multiple films, give breaks between recording sessions whenever participants request it.

## 5. Connecting the Viewed Parks to Impressions of the Videos

1. Compare the extent to which people viewed objects and their opinions of the videos by asking participants 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 they liked the park (1, not very much, to 10, very much).
2. Ask the participants whether they would use the park to relax or relieve stress (Y/N) and whether they recognized the park they were looking at (Y/N) to control for the potential impact of a prior association with the park.
3. Record short answer responses from the participants to explain their impressions using a voice recorder and then transcribe these.

## 6. Automatic Video Analysis for Area of Interest Extraction

### 1. Selection of AOIs

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

### 2. Training classifiers for AOI extraction

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

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.
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.
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.

#### 2. Manual pixel labelling for training frames

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

1. Pixel labelling of training frames
  1. In turn, open each training frame from the video in the image editing software.
  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).
  3. Ensure that the color palette and the mapping of color to object classes is the same throughout the labelling process.
  4. Select the color for the sample AOI.
  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. 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".
7. Ensure all labeled images are stored in a single folder.

## 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.

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.
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.
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.

## 3. Automatic pixel labelling of the video sequence

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.

2. Launch the Darwin GUI
3. In the Darwin GUI, click on **Load Training Labels**.
4. Using the file explorer dialogue box that appears, select the folder containing the labeled training images for the relevant video sequence.
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. 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.

## 4. Validating the trained classifier

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

## 5. Video pixel labelling

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.
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

1. In the eye-tracking software click on **File > Export...** Export the eye-tracking file as a CSV file.
2. Open the file in a spreadsheet program.
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.
4. Add a column to the eye-tracking data.
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.

1. Using the included Python code, calculate the amount of overlap between the objects of interest and the eye-tracking fixation time and frequency.
2. Sum this data to understand the amount of time that participants spent looking at different objects.
3. Employ a histogram to show the total amount of time that the objects occupied the screen.
4. Compare this with the total amount of time that participants looked at different objects.
5. To produce a heatmap in the eye-tracking software click on one of the park walk videos using the tick box.
6. Click the **Heat Map** tab.
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. Export the image or video that includes the heat map as a JPEG file or video file.

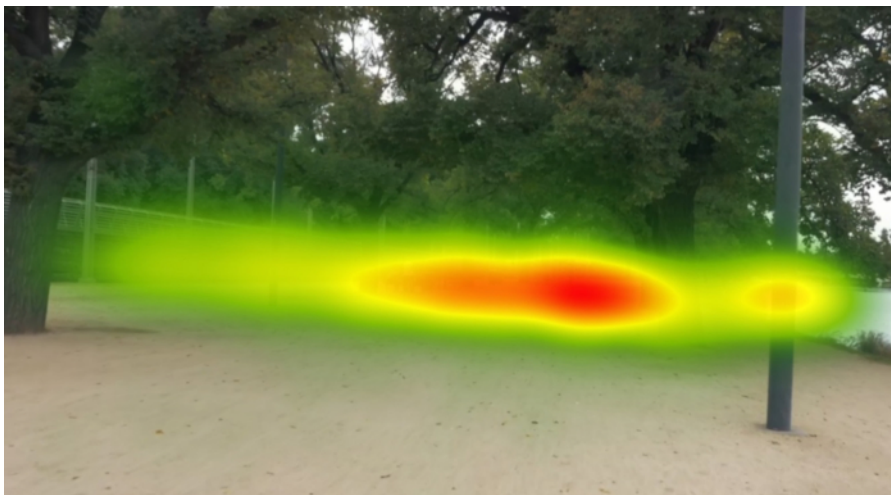
### Representative Results

**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.

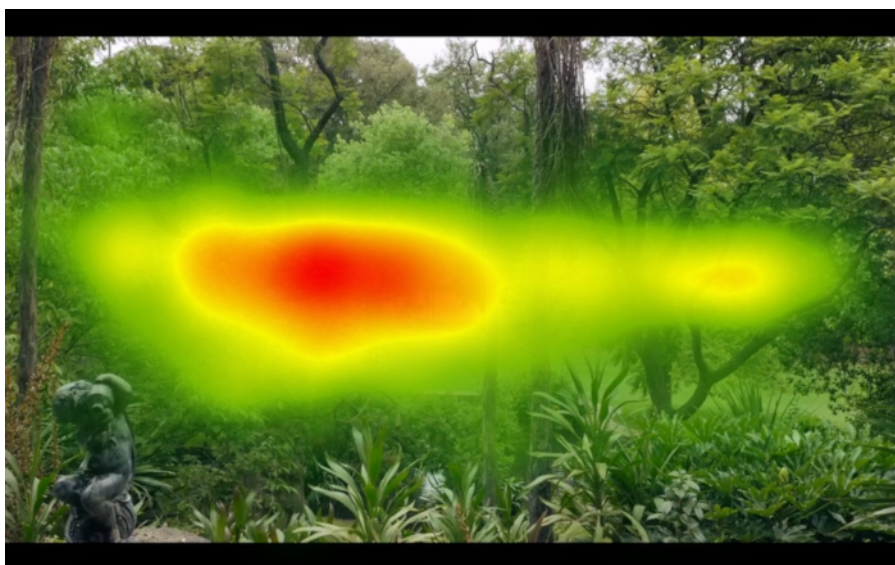
**Figure 3** and **Figure 4** show the result of using the protocol and use of machine learning to analyze the content of the videos. **Figure 3** shows the dense vegetation of Fitzroy Gardens compared to the relatively sparse vegetation of Royal Park (**Figure 4**). In the latter, more sky is visible, 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).

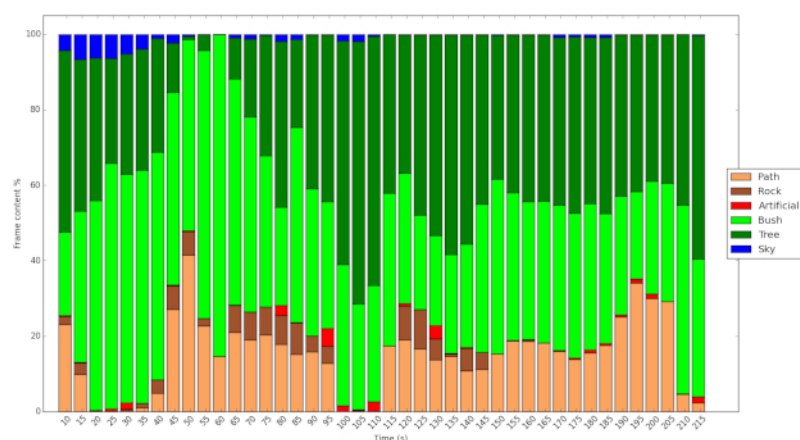


**Figure 1: Example heat map that is produced for the whole video of one of the parks.** This shows where the majority of the eye tracks were located. Note the horizontal shape of the heat map because of the dominance of horizontal elements in the video. [Please click here to view a larger version of this figure.](#)

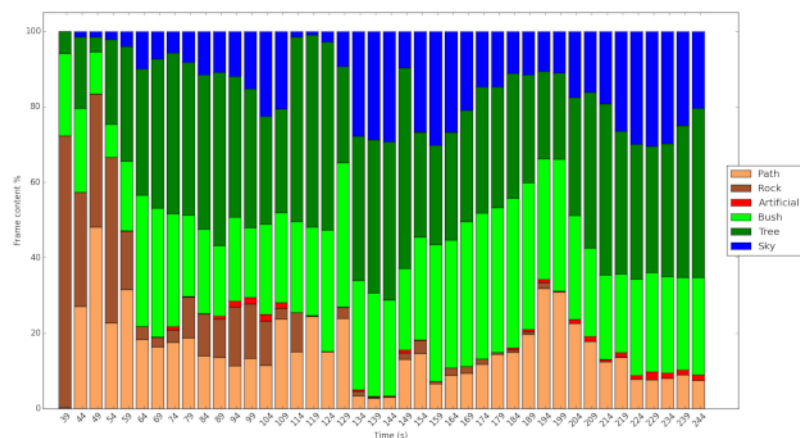




**Figure 2: Example heat map that is produced for the whole video of another one of the parks.** Note the more rounded shape because of the presence of vertical and horizontal elements in the video. [Please click here to view a larger version of this figure.](#)



**Figure 3: Histogram of content in the video of Fitzroy Gardens analyzed using the machine learning technique.** [Please click here to view a larger version of this figure.](#)



**Figure 4: Histogram of content in the video of Royal Park analyzed using the machine learning technique.** [Please click here to view a larger version of this figure.](#)

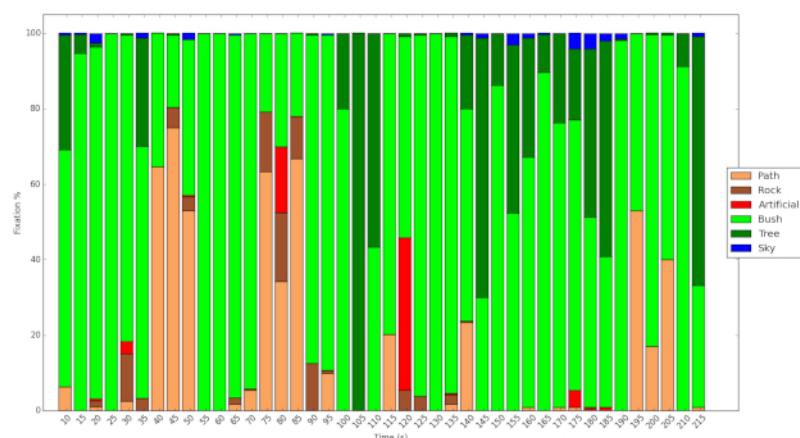


Figure 5: Eye tracking fixation time spent looking at objects in Fitzroy Gardens. [Please click here to view a larger version of this figure.](#)

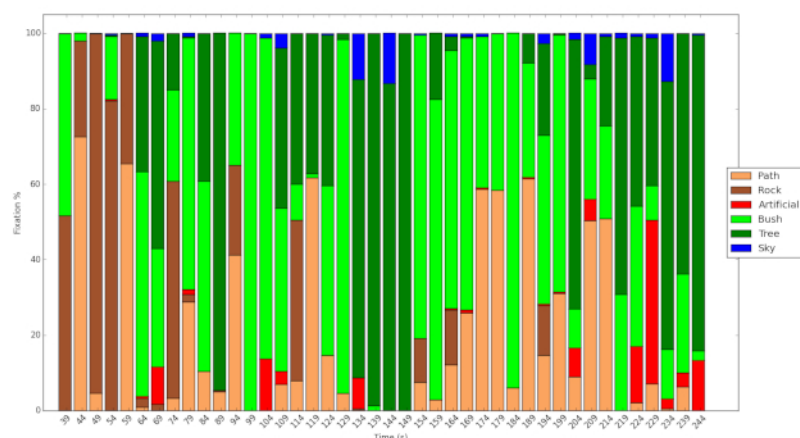


Figure 6: Eye tracking fixation time spent looking at objects in Royal Park. [Please click here to view a larger version of this figure.](#)

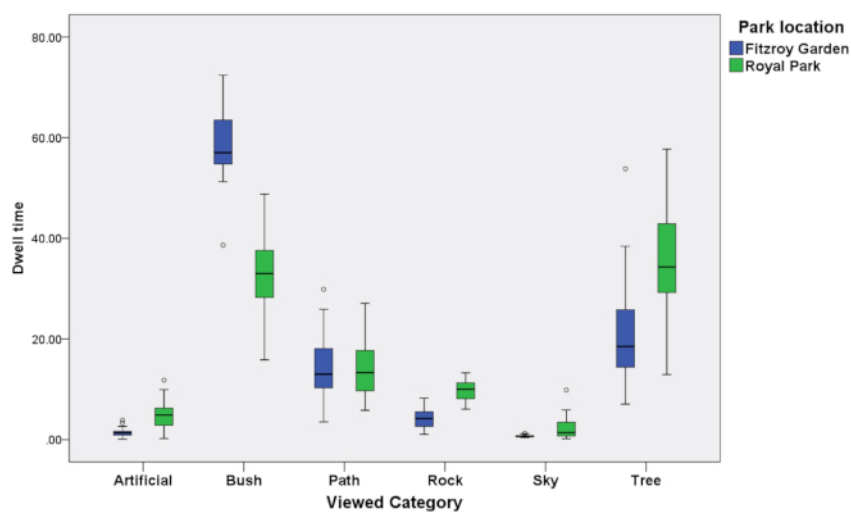
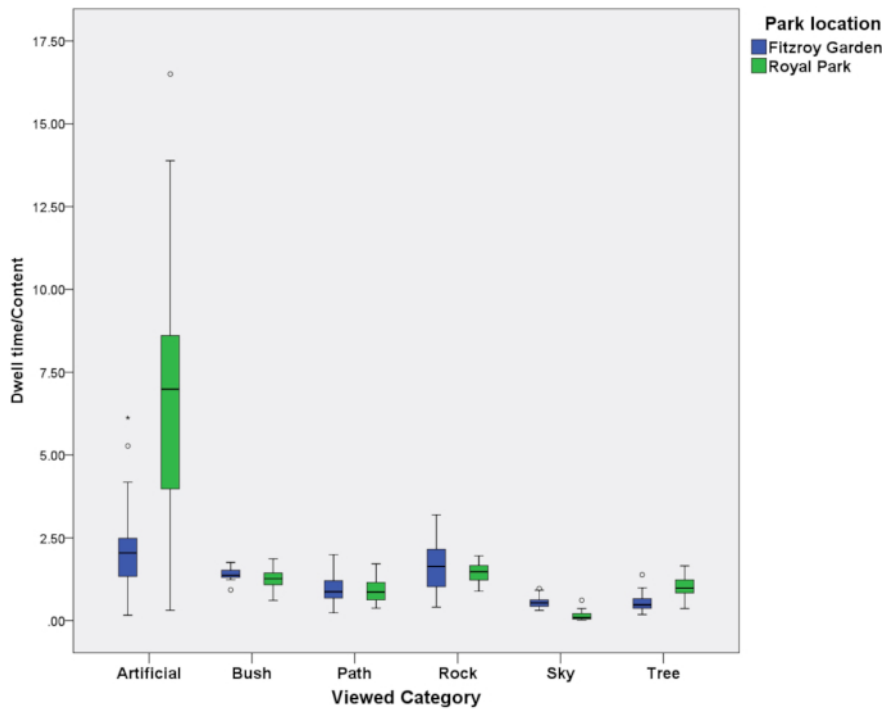


Figure 7: Aggregated eye tracking dwell times for all participants and objects for both parks. [Please click here to view a larger version of this figure.](#)



**Figure 8: Aggregated eye tracking dwell times relative to content for all participants and objects for both parks.** [Please click here to view a larger version of this figure.](#)

## 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 movements<sup>11</sup>.

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 objects<sup>12</sup>. 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).

## Disclosures

None



## 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.

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