Supplementary Material

Yini Fang*[1], Jingling Yu*[1], Haozheng Zhang[2], Ralf van der Lans[1], and Bertram Shi[1]

[1] Hong Kong University of Science and Technology, Clear Water Bay, HK
[2] University of Durham, Durham, UK
{yfangba, jyubj}@connect.ust.hk, haozheng.zhang@durham.ac.uk, rlans@ust.hk, eebert@ust.hk

We provide more details about our collected dataset in this document.

1 Preparation

Prior to conducting the experiment, a selection of 22 wine products from 21 distinct brands was made from the Bivino website[3], along with 27 yogurt products from 13 different brands sourced from the Walmart website[4]. For each type of product, 20 items were designated as target packages, and a total of 100 shelf images were created by arranging these packages in permutations on a two-dimensional grid shelf. The arrangement for the wine products was made into 2 rows and 11 columns, whereas the yogurt products were organized into 3 rows and 9 columns. The images were generated with a resolution of 1680×1050 pixels.

The data collection involved 39 subjects aged between 18 and 30 years old, with 47% male and 53% female participants. Eye-tracking data was collected using the Tobii T60XL eye tracker, with a sampling rate of 60Hz.

2 Collection Process

The experimental procedure involved instructing the subjects to search for a specific target product within the given shelf image. A screenshot of the instruction provided to the subjects is shown in Figure 3.

To ensure variability, we randomly assigned 30 unique combinations of target products and shelf images to each subject. In each trial, the target product was initially displayed for a duration of five seconds, followed by the presentation of the corresponding shelf image. The participants were instructed to locate the target product and click on it within a time limit of three seconds. Their compensation was contingent upon both the speed and accuracy of finding the target product.

A total of 1170 scanpaths were collected during the experiment. We utilized the Binocular-Individual Threshold algorithm [2] to cluster the raw gaze samples into fixations. This algorithm takes into account binocular viewing and leverages

[3] https://bidvino.com/
[4] https://www.walmart.com/

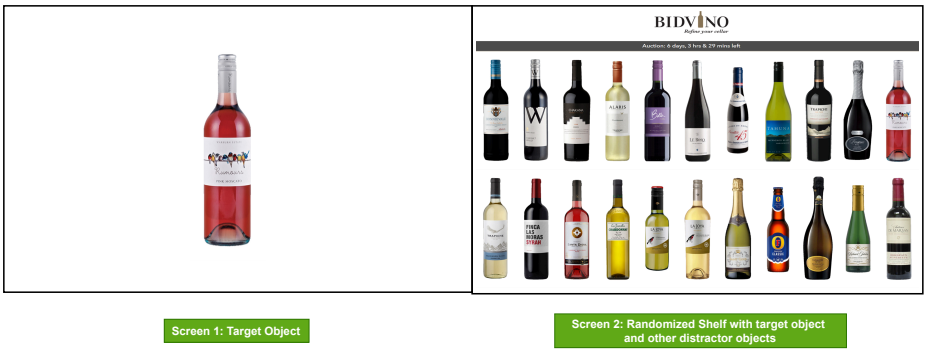

Fig. 1: Example of wine stimuli.

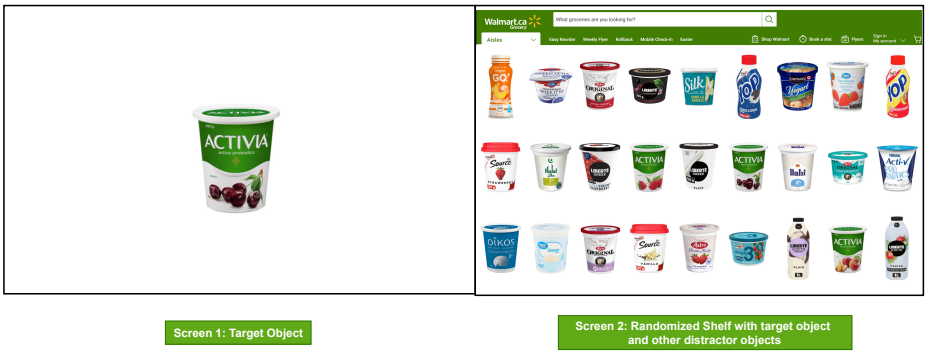

Fig. 2: Example of yogurt stimuli.

the covariation between the two eyes. Subsequently, we performed data cleaning procedures, which involved removing 312 scanpaths with less than 80% record integrity [1]. Additionally, redundant fixations were eliminated from the scanpaths after subjects clicked the mouse. Following these postprocessing steps, our dataset consists of 858 scanpaths with an average length of 7.7 fixations.

3 Consent From Subjects

Prior to conducting the study, ethical approval was obtained from an Institutional Review Board (IRB) to ensure the appropriateness and adherence to ethical guidelines. Informed consent was obtained from all subjects, who were required to sign a consent form that explicitly outlined the following provisions:

- The collected data would be used solely for research purposes.
- All participant responses would remain confidential, and no personally identifiable information (such as names or email addresses) would be collected or stored in a manner that could link the responses or eye-tracking data to individual identities.

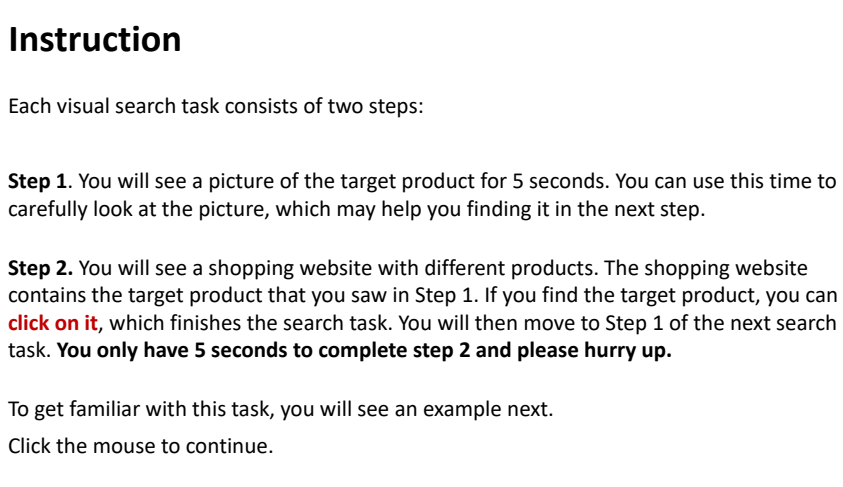

Fig. 3: Visual search study instruction.

– Only aggregated data would be utilized in any publication that incorporates this dataset.

The consent form is included in the appendix. Names and institutions were anonymized for the initial paper review.

References

1. Van der Lans, R., Wedel, M.: Eye movements during search and choice. Handbook of marketing decision models pp. 331–359 (2017)
2. Van der Lans, R., Wedel, M., Pieters, R.: Defining eye-fixation sequences across individuals and tasks: the binocular-individual threshold (bit) algorithm. Behavior research methods **43**, 239–257 (2011)