

Mode-of-disparities error correction of eye-tracking data

Yunfeng Zhang · Anthony J. Hornof

Published online: 13 April 2011
© Psychonomic Society, Inc. 2011

Abstract In eye-tracking research, there is almost always a disparity between a person's actual gaze location and the location recorded by the eye tracker. Disparities that are constant over time are systematic error. In this article, we propose an error correction method that can reliably reduce systematic error and restore fixations to their true locations. We show that the method is reliable when the visual objects in the experiment are arranged in an irregular manner—for example, when they are not on a grid in which all fixations can be shifted to adjacent locations using the same directional adjustment. The method first calculates the disparities between fixations and their nearest objects. It then uses the annealed mean shift algorithm to find the mode of the disparities. The mode is demonstrated to correctly capture the magnitude and direction of the systematic error so that it can be removed. This article presents the method, an extended demonstration, and a validation of the method's efficacy.

Keywords Eye tracking · Error correction · Data analysis

Systematic instrumentation error occurs quite often in eye-tracking studies. Systematic error is the relatively constant deviation between the locations recorded by the eye tracker and the locations where people were actually looking. The error may be caused by imperfect calibration, head movement, astigmatism, and other sources (LC Technologies, 2000).

There are only a few methods available for dealing with systematic error, and yet the error can have a negative impact on eye movement data analysis.

Figure 1 illustrates what systematic error may look like. The data are from a test of the Tobii T60 eye tracker, which is widely used in usability studies and has a reported accuracy of 0.5° of visual angle. In the test, the participant was asked to look at the four corners of the rectangle. Unlike in a typical experiment, the participant was asked to adjust her head position to test the sensitivity of the tracking accuracy to head movements. As can be seen in Fig. 1, the four fixations are all somewhat above the corners by a similar amount of disparity. The systematic error in the figure is large—roughly 1.1° , on average—well over the manufacturer's stated accuracy. The figure shows a pattern that is typical with systematic error, in that all recorded fixations are shifted by a somewhat similar vector.

Systematic error has a negative impact on eye-tracking studies. The error may not be a problem in studies in which the areas of interests are large (e.g., 5°) and are separated by large distances (e.g., 5°), but the error may lead to incorrect conclusions in studies with smaller targets separated by smaller distances. Systematic error in one study (Mello-Thoms, Nodine, & Kundel, 2002) led to a rather absurd observation: Radiologists, who were searching for breast cancers on the X-rays, were reported to be studying blank areas of the display rather than the lesions. Similarly, if the fixations in Fig. 1 were simply assigned to their nearest objects, the systematic error would lead to a misinterpretation that the lower two fixations were on the top two corners of the rectangle, whereas in fact they were on the bottom two corners.

To reduce systematic error, researchers sometimes recalibrate the eye tracker during the course of an experiment (Abrams & Jonides, 1988; Juhasz, Liversedge, White, & Rayner, 2006), but frequent recalibration cannot be used in

Y. Zhang (✉) · A. J. Hornof
Department of Computer and Information Science,
University of Oregon,
1202 University of Oregon,
Eugene, Oregon 97403–1202, USA
e-mail: zywind@cs.uoregon.edu

A. J. Hornof
e-mail: hornof@cs.uoregon.edu

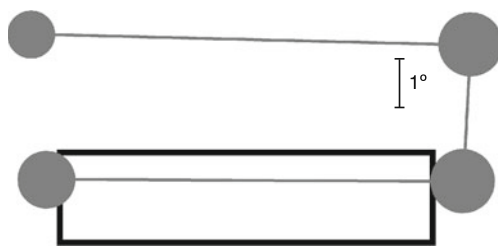


Fig. 1 Eye movement data recorded with a participant looking at the four corners of the rectangle. The circles represent fixations. Due to systematic error, the eye-tracking data are all above the corners

experiments in which interruptions would interfere with task performance. For example, when executing an air traffic control task, participants maintain contextual information in their short-term memory that would be lost during a calibration. In such cases, recalibration interferes with performance.

The required fixation location (RFL) technique, which was introduced by Hornof and Halverson (2002), can reduce systematic error for a variety of eye-tracking experiments. The key step in this technique is to identify *required fixation locations*, which are points on the screen that the analyst can be relatively certain that a participant fixated at a specific point in time, provided that the participant completed the trial accurately. For example, in Hornof and Halverson's visual search experiment, the search targets that were selected by participants (using a mouse cursor) were selected as RFLs because, on the basis of task design (no time pressure between trials and monetary rewards for fast responses), it was reasonable to assume that participants were looking at the targets when they clicked on them.

Once identified, the disparities between the RFLs and their associated fixations reveal the systematic error. For example, if the recorded fixations are 1° above their corresponding RFLs, the systematic error is 1° up. The eye movement data can thus be corrected by shifting the recorded locations 1° down.

The RFL technique is useful, but it has a limitation: The technique for identifying RFLs will vary from experiment to experiment, and some experiments may not have any RFLs. It would be useful to have a more general technique that responds to trends in the recorded data without relying on the eyes being at specific locations at specific times.

This article presents a general method for correcting systematic error in eye-tracking data. The method estimates the magnitude and direction of systematic error by exploiting a pattern in the scatterplots of the disparities between fixations and their nearest objects. These disparity scatterplots are drawn by repositioning every fixation from a session to (0, 0) and then plotting a point on the graph at the relative position of the object that was closest to that fixation. When scatterplots from fixations on stationary (or

slowly moving) objects are plotted, a cluster of disparities often appears near, but not directly at, (0, 0). This cluster is likely formed by the disparity between the location reported by the eye tracker, and where the person was truly looking. The cluster of disparities typically represents the systematic error, and the mode of the disparities best captures the magnitude and direction of the error. This article describes an algorithm for correcting eye movement data using the mode of the disparities, applies it to an experiment with moving visual stimuli, and presents a validation study to measure the effectiveness of the technique.

The mode-of-disparities error correction method

The mode-of-disparities (MoD) error correction method estimates systematic error by using the mode of the cluster of the disparities between fixations and their nearest objects. This section first demonstrates that this mode can correctly capture the size and direction of systematic error and then introduces an algorithm for finding the mode of two-dimensional data.

The pattern of the disparities

Figure 2 shows a scatterplot of the disparities between fixations and their nearest objects, again with all fixations repositioned to (0, 0). The data in Fig. 2 are from one session of a single participant. (The experiment will be introduced in the next section.) In the scatterplot, the disparity between (0, 0) and each data point corresponds to

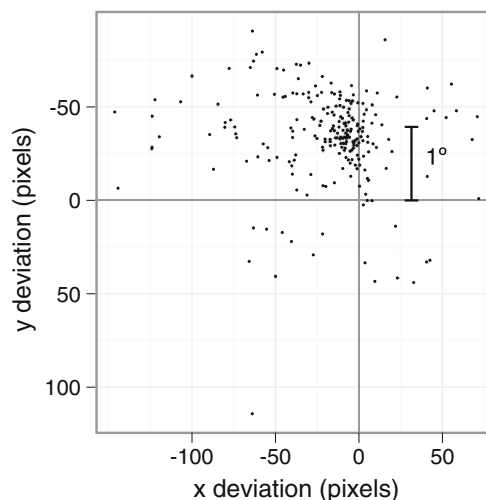


Fig. 2 A scatterplot of the disparities for every uncorrected fixation recorded in an eye-tracking session. All fixations are repositioned to (0, 0). Each data point shows the relative position of the object that was closest to a fixation. The dense cluster around (-10, -35) shows that many objects were about 10 pixels to the left and 35 pixels above their fixations

the disparity between each fixation and its nearest object. As a convention, disparity vectors will be discussed with positive x -coordinates to the right of a fixation and positive y -coordinates below a fixation. For example, a disparity at (10, -20) means that the object is 10 pixels to the right and 20 pixels above a fixation.

In Fig. 2, there is a cluster of disparities around (-10, -35). This graph suggests that a large portion of objects are roughly 10 pixels to the left and 35 pixels above fixations. Although the cluster has no clear boundary, a single cluster emerges. A similar clustering pattern occurred in all the other sessions.

Although a fixation's nearest object is not necessarily the object that was truly fixated, especially when the recorded gaze locations are shifted by systematic error, there are reasons to believe that the dense clusters indicate not just the nearest objects, but also the truly fixated objects. First, if the fixation's nearest object is the truly fixated object (a correct assignment), the disparity should correspond to the systematic error. The disparities from all correct assignments would be similar, and a cluster would form. Second, if the fixation's nearest object is not the truly fixated object (an incorrect assignment), and if the experimental layout is irregular—for example, not on a grid—the disparity would be different from fixation to fixation, and the data points would scatter. The first factor condenses the disparities from correct assignments, and the second factor disperses the disparities from incorrect assignments. The center of the cluster thus becomes a good estimate of the systematic error.

Because there is only one cluster in each graph, the mode of all data points is the best estimate of the cluster center. The *mode* refers to the *peak of the hill*, consistent with Comaniciu and Meer (2002). Figure 3 shows the same

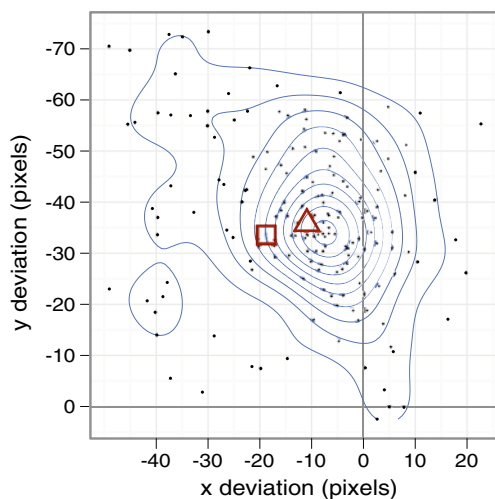


Fig. 3 The disparity scatterplot for the cluster in Fig. 2, superimposed with contours that show the relative densities of different regions. The contours connect equal density points. The cluster center has the greatest density and is at the mode of disparities. The triangle and the square show the median and mean of all disparities, respectively

data points in the dense cluster of Fig. 2, with contours connecting points that have equal densities. As the space between adjacent contours gets smaller, the density of the data points becomes higher. As can be seen in the square and triangular plot symbols in the graph, the mean and the median of all data points are somewhat to the left of the cluster center, because they are overly influenced by points outside of the cluster. But the highest-density point, the mode, is at the center of the cluster. This point is the MoD.

The systematic error can thus be reduced by shifting the eye movement data on the basis of the error vector from (0, 0) to the MoD. This error correction method should work if the following four assumptions are met.

- Objects on the display must be arranged somewhat randomly—for example, not on a grid. Otherwise, the error correction might systematically shift fixations to the wrong objects.
- Objects must be fixated individually, not in groups. This error correction technique might not work in experiments in which participants tend to foveate multiple objects with a single fixation (as discussed in Findlay, 1982).
- Each object must have a predictable fixation location, such as its center. The clustering described above may not occur if participants can obtain task-relevant visual information from any part of a large object.
- The distance between adjacent objects must exceed the size of the clusters on the disparity graph. Otherwise, distinct clusters will not form. Dense displays may thus be problematic.

Given that the MoD method correctly captures the systematic error, the remaining question is how to find the mode in a two-dimensional space. Although the location of the cluster center can be roughly guessed by just visually examining the disparity scatterplot, it would be much more efficient for analyzing large data sets if the error correction procedure could be executed automatically. The next section describes an approach to finding the mode using the annealed mean shift algorithm (Shen, Brooks, & van den Hengel, 2007). This algorithm was originally designed for solving computer vision problems and is adapted here to extract systematic error from eye-tracking data.

The annealed mean shift algorithm

The mean shift algorithm, which was developed for solving computer vision problems (Comaniciu & Meer, 2002), can be adapted to solve the problem of identifying the systematic error in eye movement data. The mean shift algorithm is derived from a nonparametric density estimation method, the kernel density estimate method. Because nonparametric statistics can work with any distribution and

because the disparities between fixations and their nearest objects do not follow any particular distribution, the algorithm is suitable for finding the MoD.

The algorithm is adapted to work with the disparities, and the procedure can be summarized as follows. (1) Randomly pick a starting point x which is a two-dimensional point on the disparity scatterplot. (2) Calculate the weighted average of all disparities. The weights here are determined by a two-dimensional Gaussian distribution whose center is set to x . The standard deviation of the Gaussian distribution is controlled by a parameter h , which is also called the *bandwidth* parameter. (3) Set x to the result of Step 2 and repeat Steps 2 and 3 until x does not change.

The procedure above can be formally described by the following formula,

$$x \leftarrow \frac{\sum_{i=1}^n x_i g\left(\left\|\frac{x-x_i}{h}\right\|\right)^2}{\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|\right)^2}$$

where each x_i is a data point on the scatterplot and g is the Gaussian function for setting the weights. In the standard mean shift algorithm, g can be other functions as well, but the annealed mean shift algorithm works best with the Gaussian function (Shen et al., 2007).

To illustrate how the mean shift algorithm works, consider the point x as being pulled by the x_i points. The force exerted by one x_i on x is inversely correlated with the distance between the two points. Therefore, x would be pulled in directions in which there are more x_i , and also in directions in which the x_i are closer to x . As the location of x changes, so does the force exerted by the x_i on x . Eventually, x will move to a location at which the forces from all directions are canceled out. However, along the way, x might be captured by a local cluster and if this happens, the output of the mean shift algorithm is only a local mode instead of the global mode of all x_i points. To ensure that the algorithm finds a global mode rather than a local mode we use the *annealed* mean shift algorithm (Shen et al., 2007), which reliably finds the global mode.

The annealed mean shift algorithm finds the global mode by applying multiple passes of the standard mean shift process with a sequence of decreasing bandwidths. To allow the standard deviation of the Gaussian distribution g to cover all of the data points on the first pass, the bandwidth h is set initially to a very large value. The standard mean shift algorithm is then applied in multiple passes, with each iteration starting from the final stable point obtained from the previous iteration, and each iteration using a smaller bandwidth than did the previous iteration. The large bandwidths in the initial passes of the procedure allow the algorithm to find the correct cluster.

The small bandwidths in the final passes allow the algorithm to pinpoint the global mode in the cluster. The algorithm has been demonstrated, in the context of computer vision problems (by Shen et al., 2007), to reliably find the global mode even when the starting position of the mean shift process is far from the mode. The process is formally defined in Fig. 2 in Shen et al.

To apply the annealed mean shift algorithm to identify systematic error, the parameters of the algorithm need to be set appropriately for this domain. The first and the largest bandwidth should be set to the distance between the two most distant data points such that a circle with a radius of this bandwidth can cover all of the data points regardless of the position of the circle. This way, the initial pass of the mean shift procedure should stop somewhere near the mean of all data points, close to the global mode. The smallest bandwidth should be set to a value such that a circle with a radius of this bandwidth would include only a small number of disparities near the cluster center, which would provide high accuracy for determining the location of the cluster center. The smallest bandwidth should also be large enough to be sure to include some data points. We have found that setting the smallest bandwidth to 1° allows an accurate and stable estimation of the systematic error.

In summary, the mode of disparities (MoD) is demonstrated to be a good estimate of the systematic error under many experimental circumstances, and the annealed mean shift algorithm can be adapted to find this mode. The eye movement data can be shifted toward their true locations on the basis of the error vector from (0, 0) to the mode.

The MoD error correction method was used to correct the eye movement data collected in an experiment (Hornof, Zhang, & Halverson, 2010). The next section presents the experimental setup that was used to both illustrate and validate the technique.

Validation of the MoD error correction method

The experiment

A dual-task experiment was used to illustrate and validate the MoD error correction method. The experiment consisted of two subtasks performed in parallel: a radar task and a tracking task. Each task occupied half of the screen. In the radar task, participants examined blips (small colored icons) that moved slowly down the screen and classified them as hostile or neutral. Immediately after the blip changed from black to green, red, or yellow, indicating that it was active and ready to be classified, the participant had to key-in the number that was on the blip and a classification of hostile or neutral. After classification, the blip changed to white. In each session, 57 blips were presented and were grouped into 16 waves, in

which 1, 2, 4, 6, or 8 blips were visible at the same time. In the tracking task, on the other half of the screen, the participant simply used a joystick to keep a circle on a moving target.

Twelve participants from the University of Oregon and surrounding communities successfully completed the experiment. They completed four sessions of the experiment on each of three consecutive days. Participants were financially motivated to perform as quickly as possible while maintaining very high accuracy. Given the practice and motivation, the participants' performance by the third day likely approached that of an expert.

To reduce systematic error, the instrumentation settings were kept as consistent as possible. A chinrest maintained a constant eye position 610 mm from the display, such that 1° of visual angle covered 40 pixels on the display. At this distance, the radar display was 16° wide and 13° high. Blips maintained a 2° separation. The size of blip icons was 32×32 pixels. The screen resolution was set to $1,280 \times 1,024$. Eye movements were recorded using an LC Technologies dual-camera eye tracker, which has a sampling rate of 120 Hz and a manufacturer-reported accuracy of 0.45° . Each session of the experiment took about 9 min to complete. Because the task was continuous across these 9 min, the eye tracker could not be recalibrated during a session. Even though the equipment was adjusted to reduce systematic error, when eye movement data for such a long duration were collected without recalibration, systematic error was still likely to occur.

Eye movement data were parsed using the dispersion-based fixation detection algorithm. Minimum fixation duration was set to 100 ms, as suggested by Karsh and Breitenbach (1983). Minimum dispersion threshold was set to 0.7° after testing a range of values. This dispersion threshold is small enough to characterize a smooth pursuit in the tracking task as a sequence of short fixations rather than as one long fixation. The dispersion threshold is large enough to correctly identify a fixation in the radar task as a single fixation instead of breaking it up into several small fixations.

Applying the MoD error correction method

The MoD error correction method was applied to the eye movement data on the radar display from the third day. Since there were 12 participants and each participant had four sessions on the third day, the error correction was applied 48 times. For each session, first a disparity scatterplot for all fixations on the radar display was created. Then the annealed mean shift algorithm was applied to find the global mode of the disparities.¹ Finally, the eye

movement data were shifted on the basis of the error vector between the origin of the scatterplot and the global mode.

Validation results

This section demonstrates the effectiveness of the error correction method. The validation takes two approaches. First, the improvement is qualitatively illustrated with visualizations. Second, the improvement is quantitatively assessed with objective measures. The results show that the MoD error correction method successfully reduced systematic error for the eye movement data of the dual-task experiment.

Visualizations of the eye movement data

Figure 4 shows a summary visualization for one wave of blips with superimposed eye movement data before (white circles) and after (gray circles) error correction. The corrected fixations in Fig. 4 are at more plausible locations because they are close to green and yellow blip icons. In the dual-task experiment, participants were motivated to look at a blip to get its number immediately after it changed from black to red, green, or yellow. Participants then keyed-in the classification, and the blip turned white. Some of the uncorrected fixations in Fig. 4 are close to white blips, but there was no motivation to look at white blips. The corrected fixations are more plausible than the uncorrected.

Examining visualizations of other sessions showed a similar effect: All fixations were shifted to more plausible locations after applying the error correction. However, in 1 out of 48 sessions, the error correction reduced the systematic error only for the later waves, and not for the

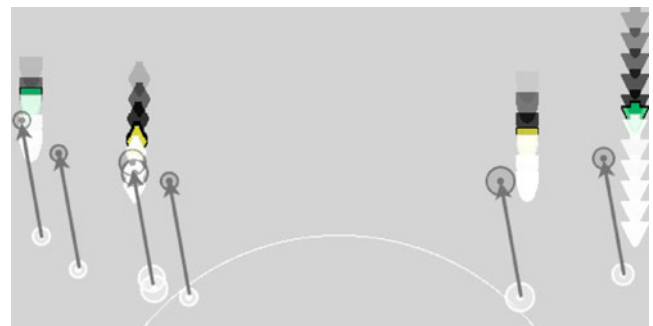


Fig. 4 The summary visualization for one wave in the dual-task experiment. The trail of each moving blip is shown as a series of icons. Each blip started as black, changed to yellow or green as it moved down, and then changed to white after it had been classified. The uncorrected fixation data (white circles) and the MoD-corrected fixation data (gray circles) are both shown. The arrows show the error vector ($-0.38^\circ, -2.25^\circ$). On the basis of task constraints, the corrected fixations are at more plausible locations because they are close to the yellow and green blip icons

¹ The algorithm, implemented in the Python programming language, required 14 s to run through all 48 sessions (9,730 data points in total) on a 2.4-GHz Intel Core 2 Duo Macintosh.

initial waves. The solution for this session is discussed below in the Possible Extensions section.

It would have been difficult to apply the RFL technique (Hornof & Halverson, 2002) to this experiment. Although participants were motivated to look at red, green, or yellow blips, it would be hard to determine the RFL for a fixation if multiple blips of these colors were near the fixation. For example, without knowing the error vector, the second-to-left uncorrected (white) fixation in Fig. 4 could be misassigned to the yellow blip above and to the right of that fixation. To avoid assigning fixations to ambiguous target locations, complicated constraints would need to be established to identify the RFLs. These constraints would be hard to implement because they involve many experimental details—for example, when a blip changed color or how many active blips were on the radar display at a specific moment. Experimental analysis (Hornof et al., 2010) also demonstrated that when participants keyed-in a response, they were not even looking at the blips but, rather, were back on the tracking task, further illustrating the challenge in applying the RFL technique to the experiment. The MoD error correction method, however, needs to know only the locations and times of the fixations and the stimuli. This information is generally easy to acquire in experiments, making the method generalizable.

Objective validation

The visualization in Fig. 4 helps to qualitatively show the effect of the error correction, but quantitative validation is also needed. The effect of the error correction is evaluated with two measures: (1) the deviations between fixations and the truly fixated objects and (2) the percentage of correct assignments when fixations are assigned to their nearest objects. Correct error correction will reduce the deviations and increase the percentage of correct assignments.

Ground truth mappings. To acquire the two validation measurements—the deviations between fixations and their intended objects, and the percent of correct assignments—it is necessary to know which fixation–object mappings are truly correct. In other words, we need the *ground truth* of where people were looking. These mappings were established by identifying trials on which experimental data, independently of eye tracking, could confirm that a participant viewed an object during a span of time. These mappings were established using the RFL technique with constraints such as: (1) During the time span of the fixation, there had been only one active blip (red, green, or yellow blip) on the radar display, and (2) the blip was correctly classified. The constraints were implemented in a computer program to automatically identify the ground truth mappings.

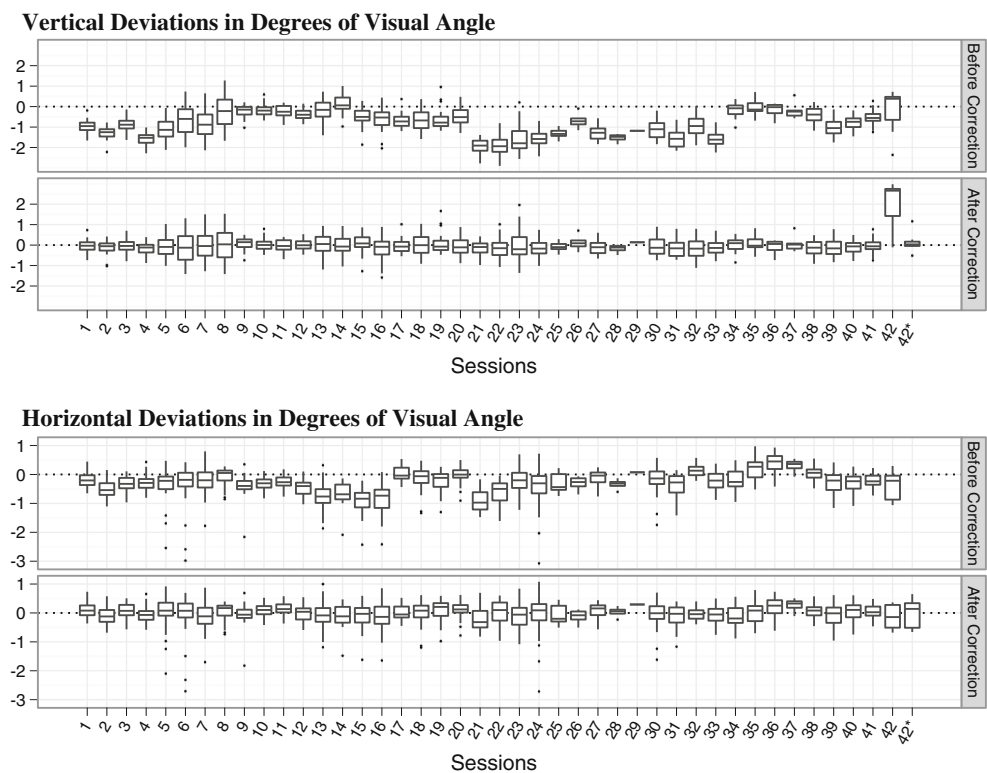
Twenty-three percent of the fixations on the radar display were found to satisfy the constraints and thus produced ground truth mappings. Specifically, 1,237 ground truth mappings were identified across 42 of the 48 sessions. The other 6 sessions produced no ground truth mappings. Because the ground truth mappings could be generated from any part of the display, they can be regarded as a random sample from all of the fixations. Thus, evaluating the error correction against the ground truth mappings should accurately estimate the efficacy of the method.

Deviations in ground truth mappings. Figure 5 shows the effect of the MoD error correction on the 42 sessions for which ground truth mappings were available. The top two panels show the vertical deviations of the uncorrected and MoD-corrected data, and the bottom two panels show the horizontal deviations. As can be seen in the top two panels, the vertical deviations vary dramatically in the uncorrected data, with many of the median deviations reaching -1° to -2° , but the median deviations in the corrected data all align at 0° , except for Session 42, in which the systematic error changed over time. This was subsequently corrected by applying a time-based error correction (discussed in the next section), and these results are shown in Fig. 5 as Session 42*. The median vertical deviations of over two thirds of the sessions in the corrected data are within 0.1° , and all (except for Session 42) are within 0.2° . The horizontal deviations, shown in the bottom two panels, were small in the uncorrected data but still improved in the corrected data. The mean absolute horizontal deviation was reduced from 0.3° to 0.1° . These results show that the MoD error correction can reduce systematic error down to nearly zero.

The percentage of correct assignments. Overall, the MoD error correction increased the correct assignments by 2.4%. In the uncorrected data, 97% of the fixations were correctly assigned to their intended targets by simply assigning them to their nearest objects. In the data set that was corrected by the MoD error correction, the percentage increased to 99.4%. The accuracy of the uncorrected data was already high due to an experimental design that reduced systematic error on the front end (such as by using a chinrest and maintaining a 2° blip separation). But the technique still improved the accuracy.

The improvement in fixation assignments for some participants was substantially higher than the average of 2.4%. For one participant, 15% of the assignments were wrong in the uncorrected data, and only 2% of the assignments were wrong in the corrected data. Such differences in accuracy across different participants could create illusory individual differences. With the MoD correction, this participant's data can be included in a reliable experimental analysis.

Fig. 5 The vertical (top two panels) and horizontal (bottom two panels) error in the uncorrected and corrected data. Each boxplot shows the median and quartiles of the deviations in each session. Before error correction, many deviations were -1° to -2° . After error correction, all deviations were around 0° , except the vertical deviation in Session 42, which was subsequently corrected with a time-based correction and is shown in the graph as Session 42*



Both the eye movement data visualizations and the quantitative measures corroborate the effectiveness of the MoD error correction method. When the systematic error stayed constant within a session, as was the case for nearly all the participants, the method worked effectively to reduce the systematic error down to nearly 0° . In the following section, we demonstrate how to extend the method to incorporate dynamic systematic errors.

Possible extensions

This section presents two ways to extend the error correction method to handle situations in which systematic error changes over time or across regions. Each extension is applied as follows: (1) Run the core MoD error correction method for all fixations to remove the systematic error for the majority of the fixations, (2) group fixations on the basis of time or region, and (3) apply the error correction again for each group.

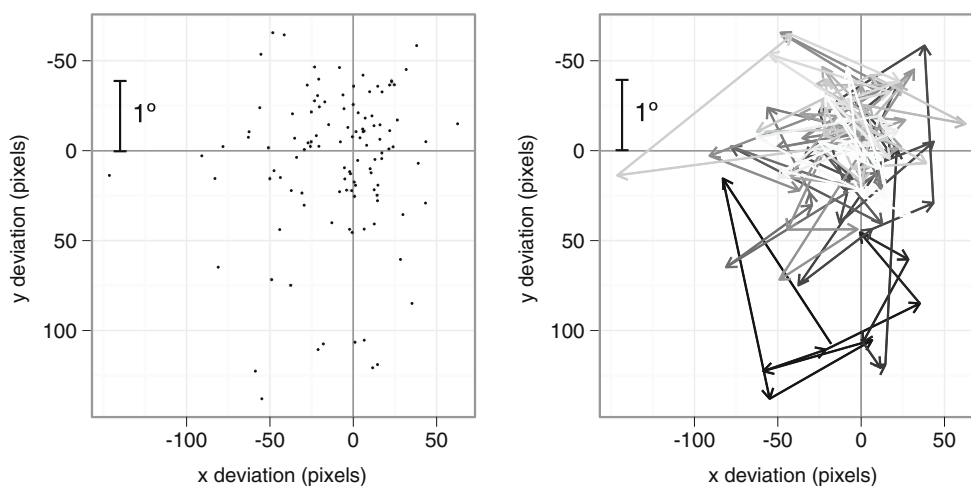
Systematic error that changes over time

For systematic error that changes over time, it is possible to identify the approximate moment in time at which the shifts occur with the aid of a modified disparity scatterplot that incorporates the temporal ordering of the disparities.

Figure 6 shows two disparity scatterplots for Session 42, the session with a systematic error that changed over time. The left panel shows the original disparity scatterplot, and the right panel shows a *time-based* disparity scatterplot. For each graph, the MoD technique has already been applied once, and so the cluster is centered at the origin. But some disparities, such as those below 100 on the y -axis, are relatively far from the cluster. The original disparity scatterplot (left panel) does not help to determine whether the disparities below 100 are caused by a different systematic error, incorrect assignments, or random error. However, the time-based disparity scatterplot (right panel) reveals that initial disparities, shown in black, are near each other in both space and time, forming a separate cluster. They are likely caused by a different systematic error that existed for a span of time.

After identifying the time span during which the systematic error changed, the fixations can be separated into two groups, and the error correction can be applied again for each group individually. For Session 42, we divided the fixations into two groups at the 90-second mark. The MoD before 90 seconds was $(+1, +108)$, and the MoD after 90 seconds was $(-1, -6)$. After applying each vector to each group, the mean absolute deviation in Session 42 was reduced from 2.2° to 0.6° . The new boxplot is shown in Fig. 5 as Session 42*. The error correction now works for all 42 sessions.

Fig. 6 The disparity scatterplots for Session 42, in which the systematic error changed over time. The right panel adds arrows to show temporal orderings of the disparities, with time moving from black to white. The black arrows, from the start of the session, form a second cluster that is below the primary cluster



The time-based disparity scatterplot is useful for detecting systematic error that changes over time and may be an appropriate visualization technique to use whenever the MoD error correction method is used.

Systematic error that changes across regions

Evidence of systematic error that changes across the visual display can be revealed by visualizations that show the fixations and the nearest objects in their original positions on the display. We studied such visualizations for the dual-task experimental data and found that there were small differences between regions in terms of the disparities between fixations and their nearest objects.

To remove these small differences, the display was first divided into three regions—top-left, top-right, and bottom—such that each region occupied one third of the display and such that the number of fixations in the three regions was about the same. Then an additional error correction was applied for each of the three regions. The additional error correction provided very little additional improvement. The mean absolute deviations of the region-based corrected data were almost identical to those of the original corrected data. The other measure, the percentage of correct fixation assignments, also showed that the region-based error correction provided no accuracy improvement: Only one new correct fixation–object assignment was identified, which is less than a 0.1% improvement. For this experiment, it was better to just use the initial error correction with the extension of the time-based correction. In other experiments, however, the region-based correction might be required to reduce error. Again, as in any data analysis, exploratory analyses must direct the application of the data analysis technique.

In summary, the MoD error correction method can reliably estimate systematic error for experiments with or without moving objects, and it can also be extended to identify the change of systematic error across different

windows of time. The core error correction method, in which a single error vector is applied to all data, works sufficiently well. To clean up systematic error that changes over time, researchers can visually examine the time-based disparity scatterplot to look for the shifting points of the error. For the dual-task experimental data, there seemed to be little benefit in doing an additional pass of error correction for subregions, but it might be needed in other circumstances, such as for larger displays.

Conclusion

Instrumentation error needs to be considered when doing scientific research, but this important practice is often overlooked in eye-tracking research. Many eye-tracking studies overestimate the accuracy of the eye tracker used in the study. This article discusses the adverse influence of systematic error in eye tracking and presents a general and robust method for improving the accuracy of eye movement data.

The MoD error correction method estimates the systematic error on the basis of the mode of disparities between fixations and their nearest objects. When each fixation is paired with its nearest object and all fixations are repositioned to a common location, the data points tend to form a cluster that reflects the size and direction of the systematic error, whereas the disparities of incorrect assignments tend to scatter. By using the annealed mean shift algorithm, the MoD method is able to find the center of the cluster. Other algorithms could also be used to find the center of the cluster, as long as they can reliably find the mode in a two-dimensional space.

The key strength of the MoD error correction method, as compared with the RFL technique, is that the MoD method does not need to be adapted to different experimental designs. The only parameters that need to be set are a series

of bandwidths used by the annealed mean shift algorithm. The only inputs required are the fixation locations, which are directly available from the eye-tracking data, and the visual stimulus locations, which are directly available from the task. The RFL technique, however, requires a task analysis to find the *required* fixation locations, making it more difficult to implement. The difficulty was demonstrated here in that the RFL technique was applied to identify the ground truth mappings, and it worked with only 23% of the fixations and 87% of the sessions.

The MoD error correction method has a limitation. It can work reliably only for experiments in which the stimuli appear at somewhat random locations. The method requires the objects to be arranged so that if the entire set of objects is shifted with the same vector, the new locations will not precisely overlap with the original locations of other objects. When the visual objects are arranged regularly (such as on a grid), the systematic error could shift all fixations to adjacent objects, the mode of disparities could show the disparity between fixations and adjacent objects, and the technique could fail. When the visual objects are arranged on a grid, it is better to use the RFL technique.

One possible direction in which the MoD error correction method could be improved would be to develop a technique to automatically find the periods of time or the regions in which the systematic error changes. The current process requires judgments that are derived, in part, on the basis of data visualizations. This could be improved by having the mean shift algorithm identify clusters in a five-dimensional search space, in which time and horizontal and vertical fixation location are the additional dimensions. Appropriate bandwidths would need to be selected for each dimension, with some consideration of task constraints. The boundaries of the resulting clusters will reveal when and where the systematic error changed.

Another way to improve the MoD technique would be to incorporate methods to more accurately correct systematic error that changes across the display. The current method applies one constant error vector to each region of the display; that is, the error vectors change discretely across the regions. However, the systematic error may change in a continuous way. It is possible that the continuous change of systematic error across the display could be approximated by a biquadratic function. This function has been shown to describe the mapping between the raw eye tracker data (eyeball rotations) and the screen locations (Sheena & Borah, 1981). If the biquadratic function can also correctly describe the change of systematic error, each fixation could be corrected on the basis of an error vector that is specifically calculated with the biquadratic function for that location. But the manner in which the systematic error changes across the display would still need to be examined, such as with the MoD technique presented here.

Applying an error correction method to eye movement data requires a certain dedication to the science and art of eye tracking, especially if it is applied with the level of rigor described here. For the dual-task experiment, numerous parameter studies were conducted to determine, for example, how many different error vectors should be calculated for the temporal periods of a 9-min task and for different spatial regions of the display. It is much easier to simply report the eye tracker accuracy given by the manufacturer and, from then on, to ignore any possible error in the eye-tracking data or, if error happens to be noticed on some trials, to just discard those trials. However, we believe that a bold, daring, and honest look at eye movement data and a commitment to attacking error is critical for the advancement of eye-tracking research and application.

Author Note This material is based upon work supported by the National Science Foundation under Grants IIS-1017593 and IIS-0713688.

References

- Abrams, R. A., & Jonides, J. (1988). Programming saccadic eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *14*, 428–443.
- Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*, 603–619.
- Findlay, J. M. (1982). Global visual processing for saccadic eye movements. *Vision Research*, *22*, 1033–1045.
- Hornof, A. J., & Halverson, T. (2002). Cleaning up systematic error in eye-tracking data by using required fixation locations. *Behavior Research Methods, Instruments, & Computers*, *34*, 592–604.
- Hornof, A. J., Zhang, Y., & Halverson, T. (2010). Knowing where and when to look in a time-critical multimodal dual task. In *Proceedings of ACM CHI '10: Conference on Human Factors in Computing Systems* (pp. 2103–2112). New York: ACM.
- Juhász, B. J., Liversedge, S. P., White, S. J., & Rayner, K. (2006). Binocular coordination of the eyes during reading: Word frequency and case alternation affect fixation duration but not fixation disparity. *Quarterly Journal of Experimental Psychology*, *59*, 1614–1625.
- Karsh, R., & Breitenbach, F. W. (1983). Looking at looking: The amorphous fixation measure. In R. Groner, C. Menz, D. F. Fisher, & R. A. Monty (Eds.), *Eye movements and psychological functions: International views* (pp. 53–64). Hillsdale, NJ: Erlbaum.
- Technologies, L. C. (2000). *The Eyegaze Development System: A tool for eyetracking applications*. Fairfax, VA.
- Mello-Thoms, C., Nodine, C. F., & Kundel, H. L. (2002). What attracts the eye to the location of missed and reported breast cancers? In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications* (pp. 111–117). New Orleans, LA.
- Sheena, D., & Borah, B. (1981). Compensation for some second order effects to improve eye position measurements. In D. F. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye movements: Cognition and visual perception* (pp. 257–268). Hillsdale, NJ: Erlbaum.
- Shen, C., Brooks, M. J., & van den Hengel, A. (2007). Fast global kernel density mode seeking: Applications to localization and tracking. *IEEE Transactions on Image Processing*, *16*, 1457–1469.