Trypucie, 22 March 2026

Marek J. Drużdżel
Politechnika Białostocka
Wydział Informatyki
m.druzdzel@pb.edu.pl

Review of a doctoral dissertation
Of Mr. Mohd Faizan Ansari
Titled:
Using a Camera to Determine Human Gaze Point

Composition of the dissertation

The dissertation consists of seven chapters, of which the first three (1-3) are introductory, the next three (4-6) present three different directions of the work, and the last chapter (7) contains conclusions and plans for future work. The dissertation contains material from two publications (Chapters 4 and 5) and what looks like material for a third publication (Chapter 6). Additional, unnumbered sections of the dissertation are: Abstract (in both English and Polish), Bibliography, List of Figures, and List of Tables. This is a typical composition encountered in doctoral dissertations and is, thus, satisfactory.

Writing

The English language of Mr. Mohd Faizan Ansari's dissertation is very good and does not require many corrections. I have a handful of minor problems and corrections below.

Dissertation's goal

Mr. Mohd Faizan Ansari's dissertation focuses on the problem of gaze estimation. The problem is important and has many applications, such as understanding user intent and attention, thinking, and mood.

The traditional method, developed and used in the field of cognitive psychology, is the use of an eye tracker. The problem with this device is its cost and the typical need for specialized personnel to operate it, although user-operated eye trackers are now available. Their price is still quite high. Mr. Ansari's doctoral dissertation focuses on eye tracking by means of a simple computer camera, a device that is both inexpensive and easy to use. The author devotes particular attention to challenges arising from the technical parameters of standard computer cameras, including their low resolution, position relative to the user's head, and varying lighting conditions.

There are plenty of publications in the area of eye tracking (simple Google Scholar search reports over four million!), gaze tracking (over a million!), gaze tracking by means of a webcam (over 20 thousand!), gaze tracking by means of a webcam based on neural networks (over 20 thousand!), and gaze tracking by means of a webcam based on convolutional neural networks (over four thousand!). This says that it is a high-risk area to get into, as the competition is immense. The author's statement "There are some works that are similar to our work." (page 35) does not seem to do the justice to the immense competition in this field. Still, Mr. Ansari seems to have found a niche in this area. My research is not even remotely close to this field and I have not been following the literature closely enough to make a reliable statement whether Mr. Ansari's contributions are novel. I leave it to his advisor, Prof. Paweł Kasprowski, anonymous reviewers of the two published papers, materials of which has been included in the dissertation, and my colleagues, reviewers of the complete dissertation, to estimate the uniqueness of Mr. Ansari's contributions. In my review, I will focus on constructive critique of the work presented in those areas in which I feel reasonably competent.

Of the three hypotheses proposed in the dissertation, H1 seems to have been tested by other researchers – there seem to be plenty other papers that apply convolutional neural networks to gaze tracking using webcams. I do not know the current literature in this area to make a judgment whether tailoring a CNN model to a specific individual (Hypothesis H2) has been sufficiently tested and whether the current dissertation offers significant new insights into this problem.

Hypothesis H3 (concerning a technique called "transfer learning") is interesting and reminds me of a technique used in Bayesian approaches to learning. Before the actual learning takes place, one might train a model on population data or even on parameters obtained from experts (Bayesian approach to probability allows for obtaining and processing expert judgments). What is called in machine learning "transfer learning" looks like just an instance of Bayesian updating – a general-purpose statistic is adjusted to the problem or to the individual at hand. It is to me not surprising that this adjustment leads to more accurate models.

The goal has been achieved and the hypotheses stated have been tested successfully.

Research methods applied in the dissertation

The author has performed a combination of design, implementation, and testing of his proposed solutions. While I have a few questions, I believe that overall the approach is correct.

It would be interesting to know how the proposed methods perform on high-resolution cameras and better lighting. The trend in technology is to produce better and less pricey devices (for example, in terms of higher resolution) with time, so what precisely is defined to be "low resolution" will change with time. One might be tempted to take a

different approach in testing the hypotheses: Using one of the high-resolution data sets, such as those listed in the dissertation, and pre-processing them to lower the resolution systematically. Starting the study with a high resolution data set would allow for a more meaningful comparison of the results to those obtained by other researchers using the same data sets.

One weakness of the performed experiments is that they were performed on data collected from one subject. Depending on physical characteristics of that subject (e.g., shape and color of the eyes), the results may be specific to that subject. It is precisely variability among subjects that gives us variance, crucial for deriving statistical significance of the results or confidence intervals over the results. I would expect results to be compared for several subjects.

The data set used in "transfer learning" (Chapter 6) includes measurement from diverse subjects (Asians, Africans and Europeans). It would be very interesting to see the accuracy in each of these three groups of subjects in separation. There may be physical characteristics of each of the three groups (such as eye color or shape) that influence accuracy.

I have noticed that the author uses different measures of accuracy in various chapters (angular error, minimal error, mean absolute error). Would it not be better to use either one measure (angular error?) or perhaps two measures simultaneously, as angular error may be too rough with, for example, just 20 predefined screen areas? Sometimes, I did not catch in the text what exactly was measured. For example, Table 4.3 shows results of the experiment. What exactly is the accuracy here? Being correct in predicting exactly one of the 20 regions? It may be a good idea to see the errors. I assume that many of the errors involved neighboring regions. From that perspective, angular error would be very nice to see. I suggest presenting both errors. In Chapter 5, Tables 5.7 and the subsequent tables list minimum error. Is the maximum error not more interesting? Given the confusion that I had over different measures of error, I found the following sentence (page 73): "Based on our findings, it is evident that our results surpassed those of previous research." not really evident.

Discussion of obtained results

The discussion of the obtained results seems correct and I have not identified any major issues with internal and external validity of the experimental work outside of the comments included in the previous section.

Practical applications of the work

The results of the work are promising. In particular, models well-adapted to a specific user perform better than models trained from data derived from interactions with multiple users. The author demonstrates that models based on transfer learning are more accurate in situations with limited data than models with weights learned from the

data. The results obtained with a simple, low-resolution webcam are on par with those obtained by means of more expensive and complicated equipment. From this perspective, one can expect that there will be practical applications of this work in several areas, such as games, advertising, or interviews and hearings.

Irregularities found the work

I have found no irregularities in this work. My comments and suggestions above and below should improve the work but it is by itself worthy of a doctorate.

Rating of the dissertation's originality

As I explained above, I believe that the dissertation solves an original scientific problem. While I am not an expert in this very area and cannot make a confident judgment that the solutions proposed are novel, there are strong indications that they are. Two of the three major contributions have been published and reviewed by independent anonymous reviewers. The candidate was guided by Prof. Paweł Kasprowski, who is an expert in the area of this work. I assume that the other two reviewers of the dissertation will provide an authoritative statement on the originality of the dissertation.

Overall rating

Overall, my impression of the doctoral dissertation of Mr. Mohd Faizan Ansari is positive. The candidate has shown that he is capable of performing quality research independently and in my opinion deserves a doctorate. I recommend allowing the candidate to proceed to the next steps of the formal process of earning a doctorate.

Less important comments and questions

The paragraph on the bottom of page 86 mentions several statistical significance tests performed. One has to be aware of dangers related to multiple hypothesis tests. How many tests were performed there?

I did not understand why incorrectly classified images were manually removed (page 60). Dos this not directly influence accuracy of the method?

Why studying the results for the left eye and the right eye separately? I would think that, except for some clinically interesting cases, both eyes have the same focus.

The reason for worse results in Table 5.8 is explained as "the training images (1880) are fewer in number compared to the test images (2360)". Is this not precisely the problem of overfitting?

It might be better to merge some of the introductory material from the Chapters 4 (the last paragraph on page 41 and the first two paragraphs on page 42), 5 (the three top paragraphs on page 58), and 6 to Chapters 1-3. I realize that the introductions to the

papers contained general introductory material but when this is in a dissertation, it is better to have Chapters 4, 5, and 6 free of introductory material, unless it is directly relevant to the experiments conducted.

I did not catch the details of the setup for data acquisition but I assumed that the users were asked to click on a region that emanated a visual signal and the assumption was that they were gazing at the point or region that they were clicking? It may be a good idea to show a screen shot of the actual user interaction when collecting data, e.g., in Section 4.1.1.

Minor comments and corrections

The English language of the dissertation is very good and does not require many corrections. I will just point out a handful of problems below.

In English technical writing, it is better to avoid passive voice. There are not too many instances of passive voice but I did notice several, especially in the text taken from the two published papers.

It is customary to avoid short forms (like “let’s”, “it’s”, “there’s”, etc.) in papers and dissertations. There is just a handful of these in the dissertation, which is good. I suggest that these be replaced by their full equivalents, i.e., “let us”, “it is”, “there is”, etc.

Mathematical formulae should generally be treated as parts of sentences and, hence, typically be equipped with a comma or a period right after them, depending on the surrounding sentence. For example, equation (2.6) should end with a period and the next sentence should start with a capital letter (i.e., “Minimize ...”). Equation (2.8) should also end with a period, but equation (2.9) should end with a comma, as there is text continuing its description (there should be no dash before “Where” and “where” should not start with a capital letter). Similarly with all other equations.

When we cite several papers in a row, e.g., page 42, [101,66], it is customary to sort the citations, i.e., put [66,101] in this case.

Figure 3.1 is somewhat counterintuitive, as its intention is to show a timeline but it presents what I would call “time wrapping” with the year 2024 being followed by year 1908. I think a linear plot, possibly wrapped over several lines would be more appropriate.

Page 19: The word “data” in English is plural and so, one should write “data are” rather than “data is”.

Page 25, Section 3.2.1: Model based” => “Model-based”

Page 31: Space is missing before the citation [53].

Page 31: “X Zhang et al.” => Zhang et al.” First name initial is not needed.

Page 32: Citation of Dubay et al. is wrong – it should be [36]. I would delete the first name (“Neeru”) and also place the citation right after, i.e., “et al. [36]”.

Page 35: “advance in term of complexity” => “advanced in terms of complexity”

Page 35: “maybe be” => “may be”

Page 35: “a wide range of subjects and exhibiting various environmental conditions” => “a wide range of subjects under various environmental conditions”

Numerals from zero to ten should be written in words, higher numerals in digits. And so, on page 36, for example, one should write “Specifically, dataset has five head poses for each subject, with 21 gaze directions per head pose.” instead of “Specifically, dataset have 5 head poses for each subject, with 21 gaze directions per head pose.”

Page 36: “This dataset aims to tackle the challenge” => “This dataset aims at tackling the challenge”

Page 36: “It include” => “It includes”

Page 36: “boasts” => “includes”

Page 39: Unnecessary spaces before citations [175] and [35].

Page 40: Formatting error citation “citezhang2021eye”.

Page 41: It may be a good idea to include a full citation to the paper on which the chapter is based.

Page 43: Expression “in the wild” does not seem to be appropriate here. Perhaps just remove it? The surrounding expressions explain how the data were collected.

When writing large numbers, it is a good idea to place a comma as a separator of thousands. For example, on page 47, I would write “6,000 images”.

Page 52: “allocated for the test set of Dataset” => “allocated for the test set”

Page 95: “from the scratch” => “from scratch”

Page 95: “Impact of Dataset Amount on Convergence” => “Impact of Dataset Size on Convergence”

Page 95: “This behavior demonstrates the efficiency of transfer learning in adapting diverse data sizes.” => “This behavior demonstrates the efficiency of transfer learning in adapting to diverse data sizes.”

Page 95: “particularly in scenarios where datasets can be limited, noisy, or challenging to obtain” => “particularly in scenarios where data can be limited, noisy, or hard to obtain”

Page 99: “In this thesis, multiple approaches were developed” => “In this thesis, three approaches were developed”