Driving Road Safety Forward: Video Data Privacy Task at MediaEval 2021

Alex Liu[1], Andrew Boka[1], Asal Baragchizadeh[2], Chandini Muthukumar[2], Victoria Huang[2], Arjun Sarup[1], Regina Ferrell[3], Gerald Friedland[1], Thomas P. Karnowski[3], Meredith M. Lee[1], Alice J. O'Toole[2]
[1]Division of Computing, Data Science, and Society, University of California, Berkeley, USA
[2]School of Behavioral and Brain Sciences, The University of Texas at Dallas, USA
[3]Oak Ridge National Laboratory, USA

ABSTRACT

This paper gives an overview of the Driving Road Safety Forward: Video Data Privacy Task organized as part of the Benchmarking Initiative for Multimedia Evaluation (MediaEval) 2021. The goal of this video data task is to explore methods for obscuring driver identity in driver-facing video recordings while preserving human behavioral information.

1 INTRODUCTION

The lifetime odds for dying in a car crash are 1 in 107 [7]. Each year, vehicle crashes cost hundreds of billions of dollars [5]. Research shows that driver behavior is a primary factor in $\frac{2}{3}$ of crashes and a contributing factor in 90% of crashes [6].

Video footage from driver-facing cameras presents a unique opportunity to study driver behavior. Indeed, in the United States, the Second Strategic Highway Research Program (SHRP2) worked with drivers across the country to collect more than 1 million hours of driver video [1, 2]. Moreover, the growth of both sensor technologies and computational capacity provides new avenues for exploration. However, video data analysis and interpretation related to identifiable human subjects bring forward a variety of multifaceted questions and concerns, spanning privacy, security, bias, and additional implications [9]. The goal of the Task will be to develop identity masking methods that effectively conceal the identity of the driver, while simultaneously preserving facial actions that can be informative for understanding driver behaviors that contribute to accidents and other driving actions that pose potential safety hazards. This Task aims to advance the state-of-the-art in video de-identification, encouraging participants from all sectors to develop and demonstrate techniques to mask facial identity and preserve facial action using the provided data. Successful methods balancing driver privacy with fidelity of relevant information have the potential to not only broaden researcher access to existing data, but also inform the trajectory of transportation safety research, policy, and education initiatives [3].

2 DATA

The dataset consists of both high- and low-resolution driver video data prepared by Oak Ridge National Laboratory (ORNL) for this Driver Video Privacy Task. The videos were captured using the same data acquisition system as the larger SHRP2 dataset mentioned above, which currently has limited access in a secure enclave. For the data in this Task, there are drivers in choreographed situations designed to emulate different naturalistic driving environments. Actions include talking, coughing, singing, dancing, waving, eating, and various other actions that are typical among drivers [8]. Through this unique partnership, annotated data from ORNL will be available to registered participants, alongside experts from the data collection and processing team who will be available for mentoring and any questions.

Figure 1: A sample frame and masking methods

3 EVALUATION OVERVIEW

To assess the de-identification of faces and measure the consistency in preserving driver actions and emotions, there will be a preliminary automated evaluation as well as a human evaluation. The scores for each of the automated and human evaluations will be combined for an overall assessment, prioritizing the human assessment of de-identification and action preservation. This Task is heavily reliant on human evaluation, and we encourage participants to include in their submission any ideas, methods, and results from their own evaluation approaches. This includes any available data from participants, descriptions of methodology, assumptions, and results. This information will be shared with reviewers and the project organizers for additional discussion and opportunities for seed funding for further research.

Although we encourage all Task participants to think creatively and holistically about how the expectations of privacy, the risk from potential attackers, and various threat models may evolve, our starting assumptions are that: (1) The drivers are not known to the potential attacker; there is no relationship between the attacker and the driver; the driver is not a public figure. (2) Information from the driver's surroundings will not influence the attacker's ability to identify the driver. (3) Access to the data is limited to registered users who have signed a Data Use Agreement specifying they will not attempt to learn the identity of individuals in the videos. (4) Attackers have access to basic computational resources. (5) There is a low probability of attackers launching an effective

MediaEval'21, December 13-15 2021, Online

crowd-sourcing strategy to re-identify the drivers, in part due to the Data Use Agreement and context in which the data were collected.

4 DE-IDENTIFICATION TESTING

Human evaluation is adapted from the methodology as described by Baragchizadeh et al. in Evaluation of Automated Identity Masking Method (AIM) in Naturalistic Driving Study (NDS) [4].

4.1 Human participants

Undergraduate student volunteers will be recruited from the University of Texas at Dallas (UTD) to participate in the study in exchange for research credit. All procedures will be approved by the Institutional Review Board (IRB) of UTD. In all cases, a minimum of 10 students will evaluate each video for identity masking success.

4.2 Procedure

Evaluations will be conducted using the masked videos submitted by the Task participants. For each submission[1], a subset of at least 116 masked videos will be evaluated by human participants using an identity-matching procedure. The selected video subset will be identical for all Task participants and will be chosen by the organizers of the evaluation. The particular set of videos to be used in the evaluation will not be revealed until all submissions have been processed.

On each trial, the participant will be asked to match the identity of the person shown in a masked video to one of 5 high-resolution static facial images presented simultaneously at the top of the screen. The participants will be offered responses to indicate which of the static images shows the person pictured in the video, or to indicate that the person pictured in the video does not appear in the set of the static images. The participant will have the option of replaying the video as many times as they want before entering a response.

The static face images will be matched demographically to the person in the video so that gender, race, and age cannot provide cues to the identity of the person in the video. Each static face image will be cropped to show only the internal face, so that identification cannot be based on peripheral face cues such as hair style.

4.3 Results Analysis

Identification of the original (unmasked) videos from this dataset was assessed in a previous study at the Univ. of Texas at Dallas, using the methods described here for the Task evaluation. The identification accuracy results for these unmasked videos will be used to assess the success of Task participants in masking the face identities. It is important to note that matching the identity of the faces between the unmasked videos and the face images is not perfect. This is due to differences in the image/appearance conditions between the static face images (high resolution, cropped to show only the internal face) and the videos (inside the car, high- and low-resolution, variable expression, etc.). Therefore, masking success will be measured for each trial as the difference between the identification of the unmasked video (from the previous evaluation at UTD) and the identification of the masked video (from the human evaluation to be conducted on the masked videos submitted by Task participants).

[1]subject to the constraint that the algorithm is submitted by the deadline published on the MediaEval website

Task participants will be given summary data on the overall accuracy of their submission, as well as complete data on their performance for the individual videos. These data should be helpful for troubleshooting and improving the performance of the masking algorithm. We will also make available summary data on the performance of the other participants, so that individual Task participants can determine how their algorithm performed relative to other submissions.

The computational face recognition evaluation consists of face recognition and face detection steps. In the face recognition step, masked faces from selected frames are compared with the gallery face of the driver. We log the number of instances where the matching metric between the gallery face and the unmasked face indicate a match. In the face detection step, we attempt to detect faces in the masked video, and compute the intersection of union (IoU) score. We compute the cumulative score for IoU across tested frames.

5 ACTION PRESERVATION TESTING

The human evaluation procedures and results analysis for action preservation will be similar to those described for de-identification testing, with the following changes. On each trial, a masked video will be presented alongside a list of possible actions. The participant will be asked to select the "most obvious" action they detect in the video. Again, the results will be compiled as the difference between the accuracy of identifying the action in the unmasked video (from the previous UTD evaluation) and the masked video (from the Task participant) submission.

The automated approach to measuring action preservation will use a deep-learning based gaze estimator [10]. The action preservation is estimated by extracting the predicted gaze-vectors from both the original un-filtered video and de-identified video and measuring the Euclidean difference between the two unit vectors. Scoring closer to 0 implies quality of action preservation since the gaze estimation is relatively unchanged.

6 DISCUSSION AND OUTLOOK

With the increased availability, prominence, and applicability of data in our daily lives, multidisciplinary connections and engagement are critical to harnessing societal benefit from advances in technology and methodology. This focused video de-identification task serves as a key example of how data science collaborations designed to bridge research and practice can simultaneously help address a pragmatic need, while sparking new lines of inquiry and research trajectories. The Driving Road Safety Forward: Video Data Privacy Task strives to raise awareness about transportation fatalities and how data might enable thoughtful discussion, analysis, and actions for the betterment of our community safety.

ACKNOWLEDGMENTS

Special thanks to our collaborators and advisors, including David Kuehn, Charles Fay, Daniel Morgan, Natalie Evans Harris, Lauren Smith, René Bastón, Mae Tanner, David E. Culler, the U.S. Department of Transportation (USDOT), and the National Science Foundation (NSF) Big Data Hubs network. This effort is made possible through community volunteers, NSF Grants 1916573, 1916481, and 1915774, and an inter-agency agreement between NSF and USDOT.

REFERENCES

[1] 2021. About safety Data: Strategic Highway research Program 2. (2021). http://www.trb.org/StrategicHighwayResearchProgram2SHRP2/SHRP2DataSafetyAbout.aspx

[2] 2021. A Brief Look at the History of SHRP2. (2021). http://shrp2.transportation.org/pages/History-of-SHRP2.aspx

[3] 2021. Exploratory Advanced Research Program Video Analytics Research Projects. (2021). https://www.fhwa.dot.gov/publications/research/ear/15025/15025.pdf

[4] Asal Baragchizadeh, Thomas P Karnowski, David S Bolme, and Alice J O'Toole. 2017. Evaluation of Automated Identity Masking Method (AIM) in Naturalistic Driving Study (NDS). In *12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*. IEEE, 378–385.

[5] Lawrence Blincoe, Ted R Miller, Eduard Zaloshnja, and Bruce A Lawrence. 2015. *The economic and societal impact of motor vehicle crashes, 2010 (Revised)*. Technical Report.

[6] Thomas A Dingus, Feng Guo, Suzie Lee, Jonathan F Antin, Miguel Perez, Mindy Buchanan-King, and Jonathan Hankey. 2016. Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences* 113, 10 (2016), 2636–2641.

[7] NSC Safety Facts. 2021. Odds of dying. (2021). https://injuryfacts.nsc.org/all-injuries/preventable-death-overview/odds-of-dying/

[8] R. Ferrell, D. Aykac, Thomas Karnowski, and N. Srinivas. 2021. A Publicly Available, Annotated Data Set for Naturalistic Driving Study and Computer Vision Algorithm Development. (2021). https://info.ornl.gov/sites/publications/Files/Pub122418.pdf

[9] K. Finch. 2016. A visual guide to practical data de-identification. (2016). https://fpf.org/blog/a-visual-guide-to-practical-data-de-identification

[10] Xucong Zhang, Yusuke Sugano, Mario Fritz, and Andreas Bulling. 2017. It's Written All Over Your Face: Full-Face Appearance-Based Gaze Estimation. (2017). arXiv:cs.CV/1611.08860