

# Computational Imaging for Improving Vehicle Safety

## Data Management Plan

### Data Collection

#### What data will you collect or create?

- (1) **Dataset:** We will create two kinds of dataset: realistic simulation of sensor measurements in fog/rain, and actual measurements of sensor measurements in bad weather. The specific sensors we will simulate are RGB image sensors, Single-photon avalanche diode (SPAD) sensors, and LIDAR-based depth sensors.
- (2) **Software:** There will be two categories of software. One is for creating synthetic images and image sequences. The other is the code for algorithms for estimating scene depth.
- (3) **Model:** Deep learning models for depth estimation.
- (4) **Publications & Reports:** We plan to publish our research findings in top vision/graphics conferences and journal venues. We will freely distribute these publications in PDF format, via the research group's web page.

#### How will the data be collected or created?

The simulation dataset will be created using standard rendering tools like Blender, PBRT, and other in-house implementation of transient rendering tools commonly used for SPADs and LIDAR. These simulators allow for obtain ground truth depth maps as well as sensor measurements.

For the real-world dataset, we will use a rig consisting of projector, RGB and SPAD cameras. We will first use this dataset indoors with smoke generators to create loss of visibility of various extents. Outdoor datasets will be captured opportunistically by mounting the sensor rig on a portable cart.

### Documentation and Metadata

#### What documentation and metadata will accompany the data?

The documentation will include details on the methodology for dataset creation, as well as associated codebase for benchmarked algorithms including steps to replicate the presented results. For the dataset, we will utilize universally accepted data formats. Videos from all cameras will be collected in a compressed format, accompanied by their camera intrinsic and extrinsic transformation parameters. The data will be organized based on various criteria such as the nature of visibility, the nature of the road, and prevailing weather conditions. Details of the data-creation process will be carefully documented.

## **Ethics and Legal Compliance**

In the initial phase, we anticipate no ethical concerns with our research data since all the data we will be creating and using will be simulated. Before we transition to using real-world datasets, we will engage with the university's internal review board (IRB) to obtain appropriate approvals. Any personally identifiable information in such real datasets will be anonymized to ensure privacy.

## **Storage and backup**

### **How will the data be stored and backed up during the research?**

The dataset will be stored on a server in the research group's laboratory. Additionally, the training dataset will be backed up on Google Drive for redundancy and to minimize the risk of data loss.

### **How will you manage access and security?**

Both the research group lab server and Google Drive allow secure access. Only active project researchers will have permission to access the data. Any data gathered in the field will be safely relayed to our data server using recognized secure data transfer methods.

## **Selection and Preservation**

### **Which data are of long-term value and should be retained, shared, and/or preserved?**

Data created for training, validation, and testing of algorithms and the developed code and results will be retained for at least 10 years as they provide a baseline for subsequent research. The dataset can be used in future research on imaging and inference in bad weather, which will benefit researcher areas in traffic safety, autonomous vehicle, and smart cities.

### **What is the long-term preservation plan for the dataset?**

The dataset will be archived and stored in the research group lab server. The code will be stored in standard version control tools –*GitHub* with a link to the pre-trained model. The technical report and research papers will be stored on *arXiv*. Published papers will also be made available in line with the specific guidelines of the publishers.

## **Data Sharing**

### **What resources will you require to deliver your plan?**

The dataset, dataset generation code, algorithm implementation code, trained model, and technical reports will be posted on code and paper repositories.

### **Are any restrictions on data sharing required?**

We do not expect to require any significant restrictions on data sharing. We might delay posting some technical reports or simulation codes until the corresponding technical papers are published.

## **Responsibility and Resources**

### **Who will be responsible for data management?**

The PI, Aswin Sankaranarayanan, will have overall responsibility for data management.

### **What resources will you require to deliver your plan?**

We don't expect to need any specialized expertise or extraordinary hardware/software. We plan to use the data storage resources available at CMU.