

Automatic Detection and Understanding of Roadworks

Data Collection

What data will you collect or create?

Visual data from cameras will be collected of public streets. The purpose of the data collection is to analyze the data with computer vision and machine learning to compute analytics that describe road work zones. Data will be collected with custom written software over the duration of the award period (1 year). A dataset will be created by manually labeling and segmenting objects of interest. Data has long-term value to the computer vision community for developing new computational algorithms. DMP will be reviewed frequently by the PI and co-PIs to ensure adherence.

How will the data be collected or created?

Input visual data will be collected from cameras that were previously installed in outside, mounted on a research vehicle, mounted on a passenger bus, and accessible through online data streams. Calculated analytics will be created from the input visual data using custom developed algorithms based on computer vision and machine learning. Analytics are aggregate statistics.

Documentation and Metadata

What documentation and metadata will accompany the data?

Documentation will consist of descriptions of how the data was collected and created. Common data formats will be used. Methods developed based on the data will be published in peer-reviewed conferences. Data will be collected in a standard compression video format such as H.264. Still frames will be captured and compressed in the JPG format. Data will stored according to capture data and accompanying documentation will record the circumstances of the collection. Standard software (usually default viewers for major operating systems) with the proper codecs will be capable of viewing the data

Ethics and Legal Compliance

How will you manage any ethical issues?

The university's internal review board will be contacted regarding collection of data. Identity of individuals captured in visual data will be blurred out. All data will be stored on secure servers in a lab located on CMU's main campus.

How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?

Visual data will be anonymized and shared for research purposes. Visual data will not be available for commercial use. No IP protection will be sought after for the data collected.

Storage and Backup

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

The data will be stored on a redundant data server located in a research lab on campus. The redundancy allows for multiple drive failure without loss of data. Data will be backed up by School of Computer Science's computing facilities.

How will you manage access and security?

Access to the data will be strictly restricted to only active researchers for the project. Access will be controlled by having one person with control authority. Data collected in the field will be securely transferred to our datasever using established secure data transfer protocols.

Selection and Preservation

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

Visual data collected for the purpose of testing algorithm performance will be destroyed once the algorithm is validated. Visual data collected for the purpose of training algorithms will be stored indefinitely as they will be part of the novel dataset being developed. Dataset can be used in future research related to understanding roadworks. Such research spans multiple fields from computer vision to autonomous driving to smart cities.

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

Upon conclusion of the project, data will be moved to an archival repository in the lab. The long-term data will be backed up by SCS computing facilities. Effort for data preservation and sharing is part of normal lab function.

Data Sharing

How will you share the data?

Information about the data will be disseminated through publications, presentations, and a project website. Data will be shared only for educational purposes under the condition that they will not be used for purposes related to commercialization or recovering the identity of people.

Are any restrictions on data sharing required?

Prior to sharing data with anyone, they will be vetted to ensure they meet the established criteria.

Responsibilities and Resources

Who will be responsible for data management?

Robert Tamburo (co-PI) will be responsible for implementing and enforcing the data management plan.

What resources will you require to deliver your plan?

No additional specialist expertise or exceptional hardware/software will be required.

Planned Research Outputs

Dataset - "Roadworks Dataset"

There are not any know datasets that provide the needed information about roadworks. Therefore, a novel, comprehensive dataset will be created. The dataset will include exemplar images, descriptive tags for various conditions (e.g., weather, time, location, etc.) and manually segmented objects of interest. Images will be acquired from cameras in public spaces. Any identifiable information such as faces and license plates will be anonymized (blurred).

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Roadworks Dataset	Dataset	2023-06-30	Restricted	None specified	10 GB	Creative Commons Attribution Share Alike 4.0 International	None specified	No	No