5P-04

A Dataset of Dashboard Camera Videos for Driving Danger Detection

Yingying Di[†]   Yu Wang[†]   Jien Kato[†]   Kenji Mase[†]

Graduate School of Information Science, Nagoya University[†]

1 Introduction

With the popularization of automobiles, people could transit more conveniently. However, the miss caused by drivers may lead to severe traffic accidents, which has been a social problem over the years. Therefore, an ability to detect danger during people's commute calls for a high attention. And the foundation in the preparation for training the ability is the datasets of the happening of driving danger. Despite large quantity of datasets with driving record exist, there is still no qualified open dataset for driving danger detection in fact. As a result, we want to propose a dataset which contains thousands of videos with appropriate annotation. Moreover, it can be seen that the dataset has its originality and essentiality, as it could also be applied on other new research field, e.g. autonomous driving.

Aiming at producing a dataset with publicity, we decide to collect the videos from the Internet, mainly on the video-sharing website YouTube. As reported before, over 300 hours of video are uploaded every minute to their servers. Due to the popularity and variety of the open website, it is reliable to discover the specific videos which we are looking for.

In this paper, we promote a video dataset in which each video contains an entire period of driving danger recorded by dashboard camera set on automobile. We intercept original data into segmentations of danger period and annotate them with labels of different stage.

In the following, we first review some related work, and then introduce the details and characteristics of our video dataset. At last, we conclude the work and show our future plan.

2 Related Work

In this section, we briefly review some datasets related to ours.

Several existing datasets of dashboard camera videos, such as Daimler Urban Segmentation [1] and CamVid [2], have been built to study semantic understanding of urban scenes. KITTI [3] is a vision benchmark dataset to study vision-based self-driving

A Dataset of Dashboard Camera Videos for Driving Danger Detection
† Yingying Di, Yu Wang, Jien Kato, Kenji Mase
Graduate School of Information Science, Nagoya University

task. Though the annotation is well detailed, all videos are captured in the same situation with no accident in normal situation. It has no data of driving danger that we want. The Near-Miss Incident Database [4] gathered multiple types of data during taxi drivers' driving. First of its limitations is lack of publicity. Next, it judges the danger from other sensors, e.g. laser radar. As a result, some incidents may not be directly recognize by visual contents. Also, it recorded the data in specified cities of Japan, which limits its universality. Finally, the main purpose for the database is for education or media use. Obviously, there may be several data that are not suitable to be training data.

As existing datasets have these limitations, we are making effort to build a better one to meet our needs.

3 Building the Dataset

The dataset we are building is aimed at providing a high qualified collection of dashboard camera videos which recorded the period of driving danger. In this section, we explain the key ideas while collecting and annotating the data. Also, several characteristics of our dataset are listed and the current situation is introduced at last.

3.1 Data Annotations in Different Stages

Driving danger such as traffic accidents and near-miss incidents are both included in the videos of the dataset. Each video contains only one danger. Considering that some videos on the Internet are lack of reality, we exclude the videos with the danger which rarely happen. The annotations for the dataset is completed with multiple stages as showed in Figure 1.

In the first stage, we label the URL, the danger period and danger object category(automobile, non-four-wheel vehicle, human, animal, object) in each video. Considering danger period, the data is added with the start and end moment of segmentations which are dangerous. For most videos, the length is about 15 seconds, segmented by 10 seconds before the danger happened and 5 seconds after the danger. Our annotation have been checked for several times in order to keep the consistency of labeling.

To annotate the data with the universal viewpoint, we also employed people with or without driving license. In the second stage, the labels below are added: the location of danger appearance and happening in the

segmentations, the direction where danger object came from, the label if it has frame rate problem. Also, to handle with the problem of inappropriate and repeated content, data check is absolutely necessary.

For more details, in the third stage, we review these segmentations then annotate them with the location where drivers take action to avoid the danger and danger category (pileup, lane change, head-on, dash out, accident of others).
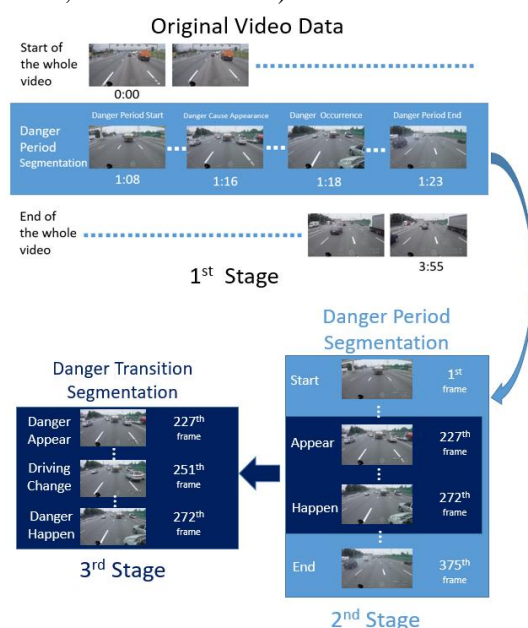

Figure 1: Example of our annotation work on data

3.2 Characteristics of Dataset

Video data we have been collecting by now has the characteristics as below:

Publicity: Most datasets of machine learning are open for academic use. As a dataset which collects videos from the Internet, we take copyright of each video into consideration and plan to publish the dataset only with their URL and correspond annotation respectively, which is similar in Heilbron's work [5].

Variety: The collected videos come from different categories of danger. Among these classifications, we pay much more attention to automobile, since it is majority of accidents in reality. Also, the videos we collected contains various driving environment and they are acquired in many different countries.

Unity: The videos we collected meet the requirements such as the content and the time period. In order to keep the visual field stable, we only choose the dashboard camera videos which are recorded on automobile only. We check the content of the videos to make sure that only one danger event exists in the data. We also keep the period of the danger segmentation to be about 15 seconds.

3.3 Current situation

The number of videos we have collected at present is 1647 totally, with 1307 in the category of automobiles, which is most concerned for the experiments in the future work. The numbers for non-four-wheel, human, animal and object are 186,70,50 and 34 separately.

In summary, we picked up 1200 from automobiles to do more detailed process first. The data is labeled by plural people and checked the result for several times in the following stage. At last, we get 1050 usable data in the second stage and labeled them in the third stage.

4 Conclusion and Future work

In this paper, we introduce our current work of building the dataset of automobile dashboard camera video for driving danger detection. It now consists of over 1600 videos with basic labels, and 1050 of them are available with detailed annotation for danger. We aimed at making our dataset different from others for several characteristics.

In the future work, it is practicable to use the method of Max-Margin Early Event Detectors [6] in the following danger detection experiments. Also, we will increase amount of data continuously.

References

1. Brostow, Gabriel J., et al. "Segmentation and recognition using structure from motion point clouds." European conference on computer vision. Springer Berlin Heidelberg, 2008.
2. Geiger, Andreas, Philip Lenz, and Raquel Urtasun. "Are we ready for autonomous driving? the kitti vision benchmark suite." Computer Vision and Pattern Recognition, 2012.
3. Scharwächter, Timo, et al. "Efficient multi-cue scene segmentation." German Conference on Pattern Recognition. Springer Berlin Heidelberg, 2013.
4. Near-Miss Incident Database. http://web.tuat.ac.jp/~smrc/drcenter.html, 2014.
5. Caba Heilbron, Fabian, et al. "Activitynet: A large-scale video benchmark for human activity understanding." Computer Vision and Pattern Recognition. 2015.
6. Hoai Minh, and Fernando De la Torre. "Max-margin early event detectors." International Journal of Computer Vision, 2014.