Stereo Processing of Images Set from a Car Video Recorder

Aleksei V. Smirnov[1,2]

[1] Moscow State University of Geodesy and Cartography
[2] LLC Racurs

KEY WORDS: Car DVR, highway monitoring, photogrammetry, PHOTOMOD, cylindrical stereopair, 3D measurements.

ABSTRACT:

Three-dimensional geospatial data are actively used for monitoring highways and controlling traffic accidents. In accident management, this data ensures an objective assessment of situations. There is potential to use car video recorders (DVRs) to enhance the productivity of three-dimensional measurements of road situations. The experience of developing and using algorithms within a domestic digital photogrammetric system is presented below. These improvements are based on a cylindrical stereo pair of DVR images and are designed for three-dimensional measurements and the creation of orthophotos of road surfaces. The article describes the step-by-step photogrammetric processing of DVR images from start to finish using the developed stereo pair. The proposed method of three-dimensional measurement is accurate, allowing for the measurement of road infrastructure objects, the measurement of distances between them, and the assessment of traffic accident monitoring.

1. Introduction

Highways are a major component of communication between cities. Changeable weather conditions and high vehicle speeds increase the number of traffic accidents. For effective monitoring of highway emergencies, it is necessary to obtain information in the shortest possible time. One approach to rapid data collection is the ability to measure XYZ coordinates and 3D vectors of a road situation using stereo models with DVR images (Figure 1).

Figure 1. Car DVR

A digital model of a traffic accident site or vehicle is not just an illustration but a basis for conclusions. It can be attached either to the final expert's conclusion for verification purposes or to the materials of a criminal case if the measurements were taken during the incident site inspection (Nedobitkov, 2022).

Research also exists in the field of three-dimensional measurements of urban infrastructure using ground-based survey materials captured by a multi-camera system including GPS/INS (Mordohai et al., 2007).

Analysis of video recordings from DVR cameras can give important information useful in reconstructing a traffic accident. These include qualitative and quantitative data such as vehicle speeds, their relative positions, and angular orientations (Abramowski et al., 2022).

SLAM technology, known as Simultaneous Localization and Mapping of the Environment, is used on mobile platforms. Research in this area with ground-based surveys has proven to be complex and requires further investigation (Krzyzanowski, 2024).

The accuracy of DVR recordings at traffic accident scenes was investigated using ground-based instrumental methods. This accuracy can be confirmed by the correspondence of linear and angular values between stationary objects in the video recordings and their actual locations at the scene of the incident (Kolotushkin, 2019).

Images from a car DVR can be used not only for monitoring highways but also for 3D modelling of buildings near roads (Hui En Pang et al., 2022).

A common problem is that most automatic methods of measuring or constructing 3D models of objects depend on external factors that negatively affect the results.

Algorithms and approaches to solving photogrammetric processing problems in different software are designed for standard survey technology, where the camera is set directly towards the object. The main feature of DVR data is its direction parallel to the object—the highway. Therefore, the initial task is to analyse the difference between standard and DVR data and check the feasibility of photogrammetric processing to obtain a stereo model.

1.1 The difference between classic aerial surveying and video recording with DVR

Standard professional aerial photography for obtaining documentation about the area is characterized by the following:

Publisher's note: Copernicus Publications has not received any payments from Russian or Belarusian institutions for this paper.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-2-W9-2025-263-2025 |

1. The object is located directly in front of the camera.
2. The object is shot from different points in space strictly along a straight line (Figure 2).
3. The specified image overlaps are strictly maintained.

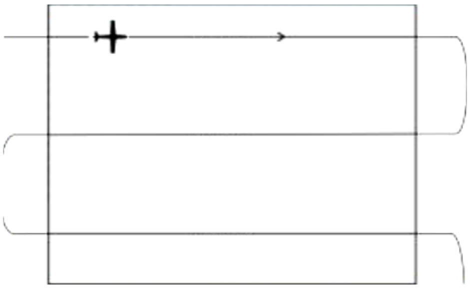


Figure 2. Straight surveying strip

Shooting from a DVR is characterized by the following:

1. The object is always on the periphery of the image.
2. Surveying is conducted parallel to the object with a constantly receding horizon.
3. Shooting is linear only on straight sections of the road.
4. Image overlaps depend on changes in vehicle speed. Each subsequent image completely overlaps the previous one (Figure 3).

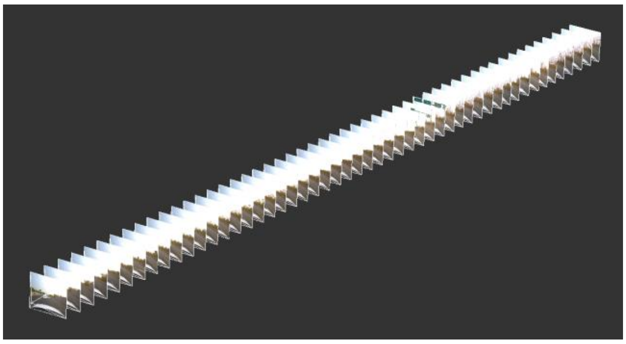


Figure 3. Surveying geometry with DVR

Modern photogrammetry and the field of computer vision offer a variety of approaches for solving the problem of three-dimensional object modeling based on their images. Each method has its own advantages, limitations, and optimal areas of application.

The task of modeling road infrastructure objects using footage from a car DVR presents several technological challenges. In this context, it is essential to analyze existing technologies and compare them with the proposed cylindrical stereo pair method.

1.2 The difference between classic and cylindrical stereomodels

Classical stereophotogrammetry relies on the use of a pair of images captured with a known baseline to calculate the three-dimensional coordinates of objects. This method is grounded in a robust theoretical framework and is widely applied in geodesy, cartography, and other practical fields.

The fundamental distinction between the cylindrical stereopair method and classical stereophotogrammetry lies in the approach to forming the stereo model. The classical method utilizes a pair of images with either parallel optical axes or a known mutual orientation, captured from a fixed baseline. In contrast, the cylindrical stereopair method employs images with optical axes aligned along the same line as the survey baseline, which are subsequently transformed into a specialized cylindrical projection.

Classical stereophotogrammetry imposes strict requirements on the mutual orientation of images and camera calibration. For precise measurements, adherence to recommended imaging protocols or the use of specialized equipment—such as stereo cameras with a fixed baseline—is essential. The cylindrical stereopair method, however, offers greater flexibility, enabling the use of conventional video recorders while compensating for their limitations through advanced processing algorithms. A significant advantage of the cylindrical stereopair method is its ability to effectively handle objects located along the direction of vehicle movement. In classical stereophotogrammetry, objects aligned parallel to the survey baseline are measured with reduced accuracy due to the small intersection angle. Furthermore, creating a stereo model using the classical approach with footage from a car DVR may face challenges due to the presence of a mono-area in the central part of the frame. The cylindrical projection method overcomes these limitations, making it particularly effective for measuring road infrastructure elements positioned along the roadway.

1.3 The difference between cylindrical stereomodels and SLAM Visual (Simultaneous Localization and Mapping)

SLAM (Simultaneous Localization and Mapping) is a set of algorithms designed to simultaneously construct a map of the surrounding environment and determine the position of a sensor within that space. In the context of computer vision, Visual SLAM is employed, which relies on analyzing a sequence of images captured by a monocular or stereo camera.

The fundamental distinction between the cylindrical stereopair method and SLAM lies in their primary objectives and approaches to forming a three-dimensional model. SLAM is primarily aimed at determining the camera's trajectory and constructing a sparse map of the surrounding environment. In contrast, the cylindrical stereopair method focuses on generating metrically accurate models of specific road infrastructure objects and enabling precise measurements.

SLAM algorithms offer significant advantages in solving navigation and spatial orientation tasks, operating in real time and efficiently handling large volumes of data. However, for applications requiring high precision, SLAM may lack the necessary level of detail and metric accuracy, particularly under conditions involving high vehicle speeds or when observing distant objects. Unlike SLAM, the cylindrical stereopair method is specifically adapted for use with car video recorders and takes into account the unique characteristics of capturing images from a moving vehicle. Notably, the cylindrical stereopair method excels in accurately measuring objects located along the direction of the vehicle's motion (along the road), a challenging task for SLAM algorithms due to the small intersection angle when the camera moves parallel to the observed objects.

1.4 The difference between cylindrical stereomodels and SfM (Structure from Motion)

Structure from Motion (SfM) is a set of algorithms designed to reconstruct the three-dimensional structure of a scene and determine camera positions from a collection of images taken from different angles. This approach is widely utilized in modern digital photogrammetry and serves as the foundation for many three-dimensional modeling systems.

Unlike SfM, which requires substantial overlap between images captured from varying angles, the cylindrical stereo pair method operates with a linear sequence of frames obtained during vehicle motion. This makes it particularly effective for processing data from car DVRs, without the need for specialized imaging equipment.

SfM typically requires a large number of images to generate a high-quality three-dimensional model, resulting in significant computational costs during processing. In contrast, the cylindrical stereo pair method is more computationally efficient, as it utilizes an ordered sequence of frames with a known trajectory of motion.

A notable limitation of SfM when applied to data captured from a moving vehicle is the inconsistency in modeling: objects near the road are reconstructed with high detail, whereas distant objects are often represented with low accuracy, which can degrade the overall quality of the model. The cylindrical projection method addresses this issue by providing a uniform representation of the entire road scene while excluding unnecessary distant objects. This uniformity is critical for subsequent tasks such as recognition and measurement of road infrastructure elements.

An important advantage of the cylindrical stereo pair method is its specialization for road scenes, allowing it to account for the unique characteristics of car DVR footage and the specific objects of interest. SfM, being a general-purpose method, is not inherently optimized for this task and may require additional adaptations to effectively process such data. Nevertheless, certain algorithmic solutions from the SfM domain can be adapted for the cylindrical stereo pair method, particularly in optimizing camera parameters and automating the search for corresponding points.

2. Materials and methods

2.1 Initial data set

The source materials consist of 51 images from the DVR video (Figure 4).

Figure 4. The first image from the car DVR

The initial camera parameters are shown in Table 1.

Parameter	Value
Frame size, px	2448 x 2048
Phys. pixel size, µm	3.45
Focal length, mm	4.914

Table 1. Initial DVR camera parameters

Additionally, there is a file containing preliminary exterior orientation data in the WGS 84 longitude/latitude coordinate system, as well as camera calibration information. All development and processing were conducted using the Digital Photogrammetric Workstation PHOTOMOD 8. In this project, there are no reference or control geodetic points on the object.

PHOTOMOD is a digital photogrammetric workstation developed by the Moscow-based company Racurs. Initially released in 1993, this software has been continuously updated and improved. Historically, PHOTOMOD has employed fundamental classical approaches for processing aerospace imagery to generate surface and terrain models, digital elevation models (DEMs), orthophoto maps, three-dimensional models, and other geospatial documents.

In the context of processing images captured by a car DVR, the workflow initially failed at the first stage of phototriangulation. It is evident that data from a car DVR differs significantly from traditional aerial data. To process the specific characteristics of such data, modifications and enhancements to the algorithms within PHOTOMOD software were implemented.
However, it remains crucial to analyze the differences between data obtained from a car DVR and classical aerial photography to better understand the challenges and requirements for processing these types of imagery.

2.2 Photogrammetric processing

The initial data from the car DVR is a video file. For photogrammetric processing, it is necessary to make a storyboard of this file. It is necessary considering that the car speed can be significant. It means that the storyboard interval should be small to reduce the basis between frames. This is required so that the necessary objects along the road for phototriangulation fall into the overlap of frames. The frames after the storyboard do not have any information about the camera in the metadata. But as written above, within the current project, information about the camera was known in advance.

To obtain a stereo pair from such images, it is necessary to perform three stages of photogrammetric processing:

1. Create the block layout, taking into account the viewing direction.
2. Conduct phototriangulation, considering the sky and the hood of the car, which appear in all frames, as well as the overlap of all images.
3. Develop a cylindrical stereo pair for measuring the traffic situation.

To correctly orient the DVR images, a change was made to the shooting direction (angle χ "kappa"). For this case, the angle χ was rotated "forward" by 90°. After changing the shooting direction, the throw-on mount is assembled correctly (Figure 5). In this procedure it is important to correctly set the rotation of the camera angle so that the X axis coincides with the direction of movement.

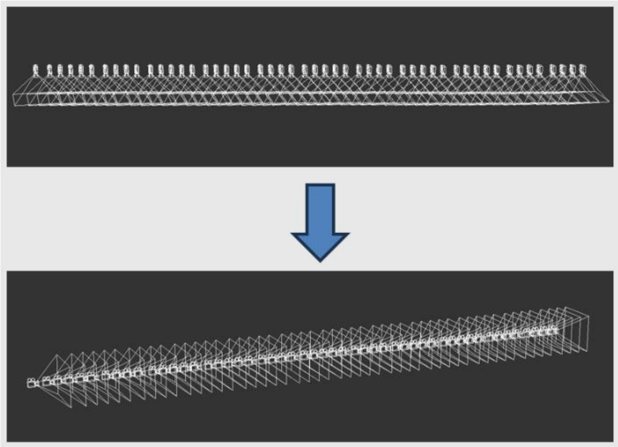


Figure 5. The block layout after modifications

Finding, identifying, and measuring tie points in PHOTOMOD uses an object-oriented algorithm (Lowe, 1999). In most digital photogrammetric systems, the default processing parameters are selected for standard aerial surveying technology and the most common image properties. In the case of dashcam images, several factors contribute to measurement errors. The first factor is the sky, which occupies the upper half of each image. The second factor is the hood of the car. Research has shown that these areas of the image should be excluded from processing to avoid the accumulation of erroneous tie point measurements. With standard correlator parameters, the percentage of incorrect points for this survey would be up to 40%. The location of the tie points without excluding problem areas is shown in the figure 6.

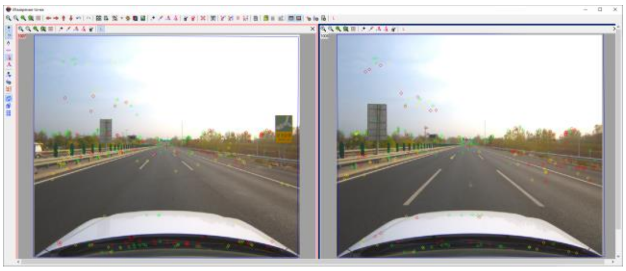


Figure 6. Tie points locations (multi-colored points)

The problem was addressed by modifying the automatic tie point measurement algorithm. A limit on the number of overlapping images used, as well as the ability to cut off parts of images in percentage terms, were added to the program.

2.3 Assessment of phototriangulation accuracy

Based on the results of phototriangulation with self-calibration, the parameters of the camera's internal orientation were refined (Table 2).

Calibration parameters	Initial calibration	PHOTOMOD self-calibration
F, mm	4,9143	4,9143
Principal point, mm	0,0266 / 0,0030	-0,0422 / 0,0059
K1	0,001573630	0,07374183
K2	-0,0001381035	-0,006500438
K3	0,000005299457	0,0003269113
K4	-6,674720e-08	-0,00000656633
P1	-0,0003378957	0,001638411
P2	0,002078974	-0,0007148821

Table 2. Interior orientation results after camera self-calibration

The interior orientation elements of the camera in the table differ slightly between the original and adjusted values. The average measurement error in the project images is 0.8 pixels. There are no excluded images in the adjusted block.

2.4 Creating cylindrical stereo pair

Constructing epipolars for a stereo pair of DVR images in the direction of the car's movement is impossible due to the infinitely receding horizon. Each image has a central area of the frame, which effectively forms a mono pair with any other image. The intersection angle in this area tends to 0.

The working area in such a stereo pair will be on the edges of the road, where there is a sufficient intersection angle for stereo measurements. The upper and lower parts of the stereo pair will be non-working, as they include the sky and the hood of the car (Figures 7 and 8).

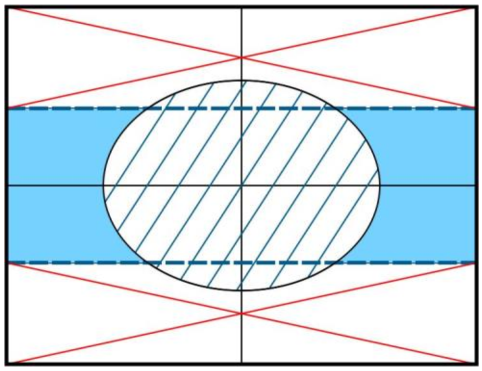
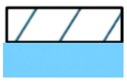


Figure 7. Scheme of mono and stereo zones



- Area of the image with no stereo

- Working area with sufficient intersection angle for stereo

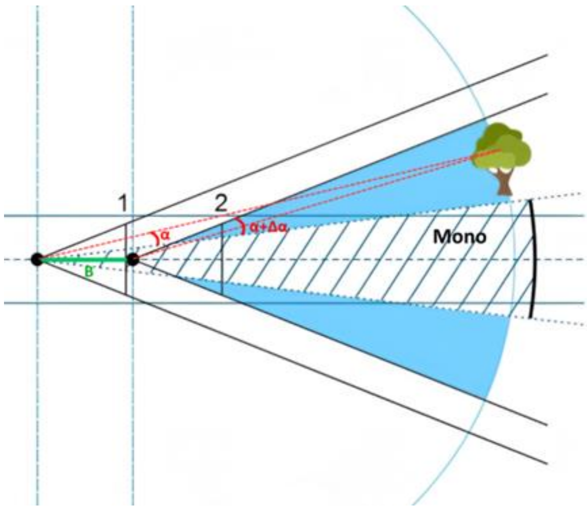


Figure 8. Scheme of a stereo pair

1 - is the first image, 2 - is the second image, B - is the basis between the images, α - is the view angle in the first image, (α +

Δ α) - is the view angle in the second image, Δα - is the intersection angle between the images.

3. Measurement with cylindrical stereo pair and creating orthophoto

To display such a stereo pair on a flat screen, the cylinder needs to be “unfolded” (Figure 9).

Figure 9. The cylindrical stereo pair (anaglyph mode)

To assess the accuracy of the internal geometry of the stereo model, a functionality was developed for manually measuring object coordinates with a parallel visual assessment of the position of the measuring mark on the nearest images. The measuring mark is synchronized on all open images. With the correct geometry of the stereo model, the measuring mark will be located at the same position on the object in all open images. In the current project, all stereo pairs were examined using this tool, and the accuracy of the internal geometry was determined to be 1 pixel (Figure 10).

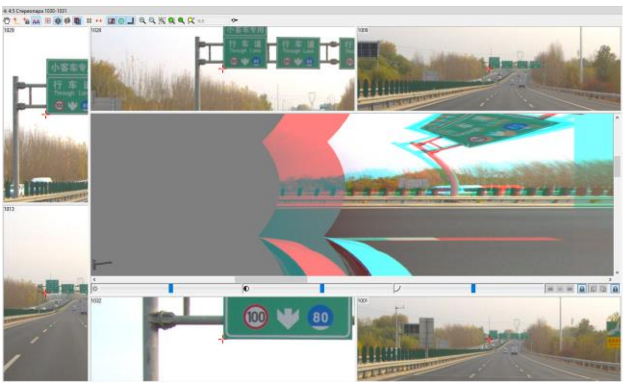


Figure 10. Control of internal geometry of a stereo pair

Using stereo and 3D vector creation tools, it was possible to reconstruct part of the roadway surface. 3D points on the edges and along the axis of the roadway were used. These points were then transformed into a primary surface model consisting of irregular triangles (TIN). Subsequently, a digital elevation model (DEM) was built on an irregular grid (Figure 11).

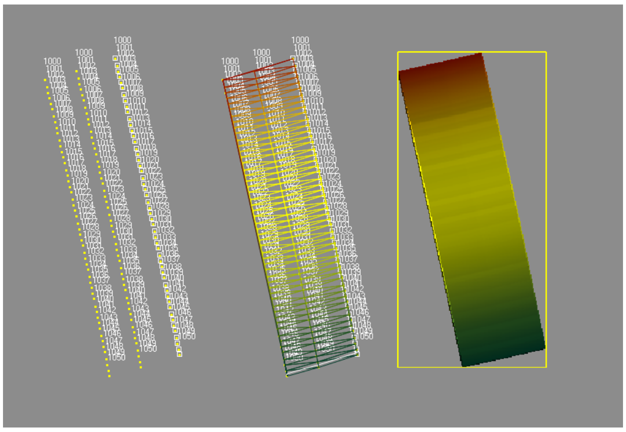


Figure 11. Manual 3D points, TIN and DEM from cylindrical stereo pair

Using the DEM, an orthophoto was built. The process of orthorectification each image was also modified. In the standard process, a photo rotated by 90 degrees would be transformed into an infinitely receding triangle. Technically, such an orthophoto cannot be created due to limited computer resources. Therefore, restrictions on transformation zones were included in the process (Figure 12).

Figure 12. The fragment of orthophoto

4. CONCLUSIONS

As a result of the research, photogrammetric processing of images from car video recorders has been improved to measure coordinates and assess the road situation in the digital photogrammetric system PHOTOMOD. Automatic algorithms for creating block diagrams, phototriangulation, and creating and unfolding a special cylindrical stereo pair have been enhanced. A special window has been developed to control the internal geometry of the stereo pair.

It is important to emphasize that the methods considered in this comparison are not mutually exclusive. A promising direction for future development lies in the integration of ideas and algorithms from different approaches to create a comprehensive solution that leverages their respective strengths.

For instance, SLAM algorithms for tracking key points and determining camera positions can be utilized to automate the process of constructing three-dimensional object models. Similarly, methods for optimizing and processing large datasets from SfM can enhance the accuracy and reliability of object modeling. Furthermore, the rigorous mathematical models of classical photogrammetry can provide a metrological foundation and enable precise accuracy assessments for measurements.

This development will allow for prompt measurement of road situations in the event of an accident. It also opens up the prospect of automatically obtaining a model of the road infrastructure and surface.

ACKNOWLEDGEMENTS

I would like to express my gratitude to the Deputy Scientific Director of Racurs, Anatoly Zubarev, and the Deputy Head of the Development Department, Alexander Golubev, for their scientific and technical consultation.

REFERENCES

Alexander I. Nedobitkov, (2022). Digital transport traseology based on Agisoft Metashape and unmanned aerial vehicle. *Transport*, part 2, Doi: 10.26518/2071-7296-2022-19-6-890-899

P. Mordohai, J.-M. Frahm, A. Akbarzadeh, B. Clipp, C. Engels, D. Gallup, P. Merrell, C. Salmi, S.Sinha, B. Talton, L. Wang, Q. Yang, H. Stewenius, H. Towles, G. Welch, R. Yang, M. Pollefeys, D. Nistér (2007). A Computer Vision Approach for 3D Reconstruction of Urban Environments.
URL:
https://people.inf.ethz.ch/pomarc/pubs/MordohaiIJAC07subm.pdf

Michał Abramowski, Andrzej Reński (2022). Analysis of Video Recording in Accident Reconstruction. Research Square. Doi 10.21203/rs.3.rs-1318556/v1

Kamil Krzyzanowski (2024). Off-board feature detection on live video feeds for VSLAM on micro-UAVs. University of Oslo, Master's thesis.

Sergey M. Kolotushkin (2019). The assessment of the correct record of a traffic accident in the video materials of the car DVR. The online scientific journal "Transport law and security"
2019. № 1 (29).

Hui En Pang, Filip Biljecki (2022). 3D building reconstruction from single street view images using deep learning. International Journal of Applied Earth Observations and Geoinformation 112. Doi.org/10.1016/j.jag.2022.102859.

David G. Lowe (1999). Object Recognition from Local Scale-Invariant Features. Proc. of the International Conference on Computer Vision, Corfu (Sept. 1999). URL: https://home.cis.rit.edu/~cnspci/references/dip/feature_extraction/lowe1999.pdf