A Novel Approach to Measurement of the Transverse Velocity of the Large-Scale Objects

Ivan Goncharov¹, Alexey Mikhaylichenko¹(✉), and Anatoly Kleschenkov²,³

¹ Southern Federal University, Institute of Mathematics, Mechanics and Computer Science, Rostov-on-Don, Russia
ivan.goncharov6@gmail.com, alexey.a.mikh@gmail.com
² Southern Federal University, Physics Faculty, Rostov-on-Don, Russia
aktech@inbox.ru
³ ISOSCAN, Ltd., Rostov-on-Don, Russia

Abstract. This paper presents a novel approach for measuring the transverse velocity of large-scale objects based on stereo vision. The suggested approach uses a high-speed stereo camera located perpendicular to the traffic lane, and consists in matching frames from the left and right cameras. Compared to methods based on monocular vision, this approach solves the problem of dependence of object speed on distance. The proposed algorithm is part of a system for non-contact measurement of large-sized objects and has a calculated measurement error that does not exceed 1.5% of the measured speed up to 30 km/h while having low computational complexity.

Keywords: Computer vision · Speed measurement · Stereo vision

1 Introduction

Systems for vehicle detection and speed measurement have become an integral part of our daily life. This is evidenced by the growing interest of researchers in this area as well as the density of implementation of these systems in the environment. As digital cameras are becoming cheaper and able to produce images with higher quality, video-based systems are becoming increasingly popular for speed measurement tasks. Most of the published works are devoted to the problem of traffic flow mean speed estimation, which is necessary when building traffic management or smart city systems. The current paper considers the speed estimation approach for use in a different application problem.

A suggested approach was developed as part of the system for non-contact measurement of large-sized objects using laser triangulation [7]. The specificity of this task is the need to get speed values with the highest possible frequency. This is necessary to minimize the error in measuring the length of the object. To ensure the required frequency of obtaining the speed values, the proposed approach uses a stereo vision system running at 250 frames per second.
In this paper, we introduce the pipeline for the transverse velocity measurement of large-scale objects with a measurement error not exceeding 2%. In Sect. 3 we describe computer vision techniques and stereo vision algorithms used in the suggested approach. In Sect. 4 we show the measurement results obtained from applying the described method to real-world data. Finally, in Sect. 5, we discuss results and potential improvements.

2 Related Works

Currently, many approaches to speed measurement have been proposed using both classical computer vision algorithms and machine learning approaches. However, most of them have a similar sequence of actions required to calculate the speed.

The first step is to detect a moving object in the frame. In the second step, the object’s motion vectors are calculated. And, in general, algorithms differ in methods for solving the problems of the above stages. For example, the authors of [4] suggest a deep convolutional neural network to detect a moving object. This approach shows stable results and is less dependent on the equipment used. Also, simpler algorithms for detecting a moving object are presented—background subtraction [8] and neighboring frames difference histogram filtering [1,2,6]. These are conventional algorithms with low computational complexity. But these methods are sensitive to conditions such as shadows and illumination variations.

Keypoints tracking [5,8] or optical flow computation [2,4] are often used for building motion vectors. Approaches based on optical flow computation demonstrate more accurate results. This is due to sensitivity of keypoint localization algorithms to lighting conditions—feature point position may differ in neighboring frames due to changing lighting.

As already mentioned in Sect. 1, a special feature of our task is the need to obtain the instantaneous speed of a moving object during the entire passage through the measuring system. In addition, the moving object is located in close proximity to the measuring system. These features make it impossible to use existing methods of speed measurement with the required accuracy.

3 Materials and Methods

In this section, we introduce our method for measurement the transverse velocity of the large-scale objects. The proposed approach is based on classical computer vision algorithms and consists of several important steps. We will discuss, in turn, our strategy for improving the results obtained and reducing the computational complexity.

In speed measurement approaches based on monocular vision, there is a problem with evaluating a perspective conversion. To avoid this, we suggest using the following property of the stereo camera.

Consider $X$—some point in space, $(x, y)$—coordinates of the corresponding point on the left camera image plane. By shifting the point $X$ in a direction...
strictly perpendicular to the optical axes of the cameras, we get the point $X'$ with coordinates $(x, y)$ of the corresponding point on the right camera image plane. In this case, the distance between the points $X$ and $X'$ will be equal to the stereo camera baseline (Fig. 1). It is important to note that this fact is true for all points and does not depend on the coordinates of their projections on the image plane, nor the distance to the stereo camera.

![Fig. 1. Theoretical basis](image)

If optical axes of the cameras are not parallel, the distance between points will be equal to the length of baseline projection on the plane of points $X, X'$. This fact allows us to determine the speed of a moving object using the conventional formula: $v = \frac{S}{t}$, by defining the unknown values as follows:

1. find a pair of frames with the same pixel coordinates of the object (according to property described above, the distance will match camera baseline);
2. extract the timestamps of the receiving frames, thus obtaining time.

Thus, speed estimation task is reduced to searching for a pair of frames $(frame_l, frame_r)$, where the object has the same pixel coordinates. Or, rephrasing—find frame $frame_l$ from the first camera $C_l$ on moving direction for frame $frame_r$ from the second camera $C_r$ on moving direction, where the pixel coordinates of the object will match up to some infinitesimal $\theta$—measurement error.

An important requirement is that the object’s motion vector is perpendicular to the optical axes of the cameras. In some applications, this cannot be guaranteed but in our case, this is due to the operating conditions. The proposed solution is used to measure the speed of vehicles, so the stereo camera module is located on the side of the roadway. This ensures that the above requirement is met.
3.1 Input Data

The suggested approach uses a stereo camera consisting of two high-speed cameras with external shutters synchronization. The cameras are placed exactly parallel at a distance of 750 mm. The stereo pair is located perpendicular to the lane at a height of about 3 m. The working frame rate is 250 frames per second. Figure 2 shows a sample of input data.

![Input data sample](image)

Fig. 2. Input data sample (before rectification procedure)

3.2 Matching Function

Let's consider several features of the input data:

1. the camera angles of the subject are completely the same for left and right cameras, due to their parallel arrangement in stereo pair;
2. position of the object can only be changed along the coordinate axis X of the image plane due to perpendicular arrangement of stereo pair relative to the lane

Because of this, the image search function with matching pixel coordinates of the object has been replaced with the search function for the most similar image. To determine the similarity of images, the difference module function is used:

$$S(F_r) = \min \left\{ \sum |F_r - F_l|, i \in Q \right\},$$

where $F_r$—image from the right camera; $F_l$—image from the left camera; $Q$—set of previous frames from the left camera. The minimum of this function will correspond to the most similar image in search neighborhood.

3.3 Camera Calibration

Due to the extreme naivety, the image comparison function is sensitive to any outliers in the data. Such outliers, for example, are lens distortion and mismatch of image lines in the left and right stereo image frames. Camera calibration [9],
stereo calibration, and application of rectification transform [3] significantly minimize the number of false positives of the image comparison function.

An important addition to the conventional camera calibration algorithm is filtering input patterns by distribution on the frame. The main motivation is to exclude patterns with a potentially high re-projection error from the calibration procedure. It is enough to leave the minimum possible number of normally distributed patterns on the frame. This approach has improved the quality of camera calibration by 20% according to the RMSE metric.

### 3.4 Increasing Resolution

The accuracy of speed measurement is affected not only by the quality of the target search function but also by the resolution of the proposed approach.

In the proposed method (Sect. 3), the speed is determined by the time the pixel coordinates of the object match on the left and right frames of the stereo camera. Since the frame rate is not infinite, the time obtained during stereo frame matching will be discrete. This directly affects the resolution of the method. Obviously, the resolution is inversely proportional to the discreteness. Figure 3 demonstrates a dependence graph of discreteness/FPS in the range of the measured speed.

![Graph showing the dependence of speed measurement discreteness on frame rate](image)

**Fig. 3.** Dependence of speed measurement discreteness on frame rate

You can see that as the speed increases, the discreteness of measurements increases, and therefore the resolution decreases. This is clearly noticeable in Fig. 4 – as the speed increases, the decrease in resolution becomes more and more obvious. This fact clearly demonstrates the need to increase the frame rate.

Decimation and binning operations make it possible to increase the FPS in most machine vision cameras. However, when they are applied, the frame size changes, and therefore requires updating the matrices of the camera model [3]. Calibrating the camera on a smaller frame means a deliberate increase of the
re-projection error, and therefore a decrease in the quality of the rectification transform. Therefore, we propose to change the matrices as follows:

\[ K'_i = S K_i \]
\[ S = \begin{bmatrix} k_h & 1 & k_h \\ 1 & k_v & k_v \\ 1 & 1 & 1 \end{bmatrix}, \]
where \( K_i \) — camera calibration matrix, \( k_h \) — horizontal scaling factor; \( k_v \) — vertical scaling factor.

\[ H'_i = S H_i S^{-1} \]
\[ S = \begin{bmatrix} k_h & 0 & 0 \\ 0 & k_v & 0 \\ 0 & 0 & 1 \end{bmatrix}, \]
where \( H_i \) — homography matrix of \( i \)-th camera in stereo pair.

### 3.5 Reducing the Search Space Dimension

Target search function of the proposed approach searches for a match in some search buffer. In general, the search buffer is a set of all frames received after the previous one found. However, as FPS increases, the buffer size also increases and the calculation of the search function for buffer can’t be completed in real time.

We propose to use a heuristic which physical value can be represented as follows: given that the speed values are obtained with an interval of \( \frac{1}{FPS} \) s, it can be assumed that the speed cannot change vastly between two neighboring measurements. Then the index of the correct solution in the buffer will not change or will change slightly. Therefore, we consider only a certain part (5–10 frames nearby with the index of the previous one found) of the entire buffer.

### 4 Results

In this section, we analyze the results obtained by applying proposed approach to real-world data. We first describe accuracy and computational complexity and then discuss the results of practical implementation.

Figure 4 shows several speed graphs samples of various vehicles. Each individual graph contains measurements of the speed during the entire passage through the measuring system. Such low measurement discreteness is provided by high frame rate—250 frames per second.

We did not use the proposed method as a separate solution. As mentioned above, the proposed method is part of the system for non-contact measurement of large-sized objects using laser triangulation. Therefore, the accuracy of the proposed method was checked by comparing the actual lengths of objects and the results obtained. Of course, this does not directly indicate the accuracy of
speed measurement, but it provides indirect confirmation of the validity of the results, since many experiments were performed with different length and nature of movement.

It is important to note that the proposed approach is a real-time solution. Which means that the algorithm’s running time is less than $\frac{1}{250}$ s. This is provided by the simplicity of the search function and the efficiency of the heuristic.

5 Conclusion

In this paper, we proposed a novel approach for real-time speed estimation using side view stereo video images. Due to its simplicity, the resulting method was able to perform with a frequency of 250 frames per second which reduces the estimated error rate to 1.5% of the measured value. The approach was tested for almost half a year on real-world data. This solution is not suitable for standard traffic monitoring systems due to focusing on the presence of a single vehicle in the frame. However, it can be extended for this by adding the segmentation of vehicles in the frame.

We continue to work on improving the accuracy and stability of the proposed approach. As future work, we intend to reduce the effect of the image background on the target search function. A separate issue is to reduce the influence of shadows of scene objects on the search function.

References

1. Dailey, D., Cathey, F., Pumrin, S.: An algorithm to estimate mean traffic speed using uncalibrated cameras. IEEE Trans. Intell. Transp. Syst. 1, 98–107 (2000). https://doi.org/10.1109/6979.880967
2. Dogan, S., Temiz, M., Küllür, S.: Real time speed estimation of moving vehicles from side view images from an uncalibrated video camera. Sensors (Basel, Switz.) 10, 4805–4824 (2010). https://doi.org/10.3390/s100504805
3. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision (2003)
4. Hua, S., Kapoor, M., Anastasiu, D.C.: Vehicle tracking and speed estimation from traffic videos. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 153–1537 (2018)
5. Kumar, T., Kushwaha, D.: An efficient approach for detection and speed estimation of moving vehicles. Procedia Comput. Sci. 89, 726–731 (2016). https://doi.org/10.1016/j.procs.2016.06.045
6. Madasu, V.K., Hanmandlu, M.: Estimation of vehicle speed by motion tracking on image sequences. In: 2010 IEEE Intelligent Vehicles Symposium, pp. 185–190 (2010)
7. Mikhaylichenko, A.A., Kleshchenkov, A.B.: Approach to non-contact measurement of geometric parameters of large-sized objects. Program. Comput. Softw. 44(4), 271–277 (2018). https://doi.org/10.1134/S0361768818040096
8. Yabo, A., Arroyo, S., Safar, F., Oliva, D.: Vehicle classification and speed estimation using computer vision techniques (2016)
9. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. 22(11), 1330–1334 (2000)