Using Hampel Identifier to Eliminate Profile-Isolated Outliers in Laser Vision Measurement

Zhibin Yao, Jiaquan Xie, Yaqin Tian, and Qingxue Huang

1Collaborative Innovation Center of Taiyuan Heavy Machinery Equipment, Taiyuan 030024, China
2Shanxi Provincial Key Laboratory of Metallurgical Device Design Theory and Technology, Taiyuan, 030024 Shanxi, China
3College of Mechanical Engineering, Taiyuan University of Science and Technology, Shanxi, Taiyuan 030024, China
4College of Mechanical and Vehicle Engineering, Taiyuan University of Technology, Taiyuan 030024, China

Correspondence should be addressed to Zhibin Yao; yzb@tyust.edu.cn

Received 17 December 2018; Accepted 3 March 2019; Published 21 July 2019

Academic Editor: Antonio Lazaro

Copyright © 2019 Zhibin Yao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, the profile of the bar is detected by laser vision technology. During the detection process, obvious isolated outliers can be observed in the profile data; dimension parameter and profile-fitting accuracy are seriously affected by these outliers. In order to eliminate these outliers and improve the measurement accuracy, this paper uses Hampel identifier and moving mean identifier to identify isolated outliers. At the same time, the profile data is fitted, and the fitting results and fitting accuracy were analyzed and compared between the original data and the renovated data. The experiment proves that the outliers in the data must be identified and processed in the data measurement process. The Hampel identifier has better recognition effect, its algorithm is simple, efficient, and robust, and it can play an important role in the preprocessing of profile data based on structured light.

1. Introduction

In the metallurgical and metalworking industry, the dimensions of bars and tubes, especially the profile size, are very important, such as diameter, roundness, and straightness, and other product parameters must rely on profile measurement. In general, contact measurement and noncontact measurement [1–5] are used for the profile size of bars. Machine vision is an important method of noncontact measurement. Compared with traditional measurement methods, it has the characteristics of high accuracy, fast speed, and no damage [6]. This project adopts the laser vision system based online structured light to measure the bar profile. The line structured light measurement system uses the principle of laser triangulation to obtain high-precision digital images and then obtain contour morphological data [7, 8].

Outliers are inevitable in the process of structural light measurement. Outliers generally mean that the observed values deviate significantly from most observed values, that is, the outliers do not obey the statistical distribution law of measured data [9]. Outliers can be generated for a variety of reasons, such as sensor noise, channel interference, distortion during processing, and even human factors. When the data is contaminated by outliers, it will lead to data model errors, incorrect estimation of model parameters, and wrong analysis results [10]. In this experiment, outliers in the profile often appear independently, which do not have a necessary relationship with the quality of data before and after. Therefore, they are often referred to as isolated outliers.

The emergence of outliers has a great impact on the accuracy and precision of measurement, which leads to erroneous measurement results. Therefore, it is necessary to identify and eliminate the outliers in the laser measurement process to improve the measurement effect.

In the course of data processing, outliers are generally judged based on location and scatter of data. The usual outlier recognition methods are Nair, $\sigma$, Grubbs, Dixon, GESD, and so on [11]. The above method is limited by the data distribution and amount, so it cannot be directly applied to the profile data processing. However, the moving identification method can judge the isolated outliers without obtaining the complete trend of data, which has a good real-time
performance and recognition effect; therefore, it becomes a better choice for online outlier detection. Among the moving identification methods, Hampel identifier is considered as one of the most robust and effective outlier identification method, which has been applied in different scientific fields and has achieved good results [12–14]. Hampel [15, 16] proposed the concept of the breakdown point to measure the robustness of an estimator against outlier; meanwhile, he pointed out that the identification of outliers that uses the median was used to estimate the position of the data and the median absolute deviation (MAD) was used to estimate the standard deviation of data. Davies and Gather [17] improved the method by using moving windows for identification. The work of Astola and Kuosmanen [18] classifies this method into decision-making filters and describes them in detail. Pearson [19] used the Hampel identification method to process the real measurement data in the control system. Liu et al. [20] conducted an in-depth discussion on Hampel identification method and combined the work of Martin and Thomson [21] to perform on-line detection and cleaning of outliers. Mercorelli and Frick [22] used MAD and wavelet packet to analyze and detect different types of gross error in industry. Allen [23] applied the Hampel method to frequency filtering and obtained good results. Pearson et al. [24, 25] classified the generalized Hampel filter and discussed the influence of window width and adjustment parameters on outlier identification and applied it to nonlinear system filtering.

This paper is organized as follows: Section 2 introduces the laser vision measurement system, as well as, observes and analyzes the outliers during the bar profile scanning process. In Section 3, the moving mean identifier and Hampel identifier are introduced, the typical outliers on the profiles are identified by using two methods, respectively, and the influence of parameter changes on the identification effect is discussed. In Section 4, the profile-fitting effect is compared and analyzed before and after outlier removal. Lastly, a conclusion is drawn in Section 5.

2. Laser Vision Measurement System and Profile Outliers

The laser measurement system consists of a laser vision sensor and scanning platform, which can effectively measure the profile of the bar. The measurement system is shown in Figure 1. The laser vision sensor adopts the line structure light scanning mode, that is, the laser profiler projects a laser line into space, and the light establishes a perspective relationship with the image plane through the camera lens, and the profile data is measured; the model is shown in Figure 2.

A 20 mm diameter bending bar was measured along its length using the above laser vision measurement system. The measurement result of the entire profile of the bar is obtained as shown in Figure 3, i.e., the cloud point data of each profile of the bar.

Observed in the profile data, there are obvious outliers in the profiles, as shown in Figure 4; spiking outliers often appear in the middle of some profiles. The outliers in the section are analyzed one by one, and two typical isolated outliers are summarized: single-point outliers and double-point outliers, as shown in Figures 5 and 6. That is the profiles at \( i = 149 \) and \( i = 159 \).

3. Identification of Profile Outliers

3.1. Moving Mean Identifier Recognizes Isolated Outliers. Moving mean identifier or movmean identifier is the easiest way to identify isolated outliers. The basic definition is as follows for the data sequence \( x_1, x_2, x_3, \ldots, x_n \): Moving window length is \( w_l \), the local mean is \( \mu_i \), and its local standard deviation (LSD) is \( \sigma_i \) within the moving window. Then, the rule of judging the outliers is as shown in equation (1).

\[
y_i = \begin{cases} x_i, & |x_i - \mu_i| \leq 3\sigma_i, \\ \text{outlier}, & |x_i - \mu_i| > 3\sigma_i. \end{cases}
\]

The movmean identifier is to use the 3\( \sigma \) criterion to judge the outliers in the data window when the length is \( w_l \). When the difference between the measured data and the local mean is greater three times than the local standard deviation, it is considered as an outlier.

It is necessary to pay attention to the processing of the starting and ending points of the data sequence, that is, the influence of edge effect, especially when symmetric moving window is adopted, the endpoint data is insufficient to fill the moving window. Gabbouj et al. [26] and Yin et al. [27] introduced the typical endpoint processing method “carry-on appending strategy” in the analysis of moving window filters; data expansion is carried out on the start and end side of input data to meet the operation of moving window. In this paper, the Hampel identifier described below will use the above method, while the movmean identifier adopts the window truncation method. The window size is truncated at the endpoints when there are not enough elements to fill the window, and the average value of the data is taken which is left in the window for the moving mean identifier.

The moving mean identification method is used to identify the single- and double-dot outliers in the middle of the profile of Figures 5 and 6.

As seen in Figure 7, the central outlier can be exactly identified, for the profile of \( i = 149 \), under the condition of 3 times local standard deviation and \( w_l = 11 \). At the same time, after trying different window lengths, it is found that \( w_l \geq 11 \) can correctly identify this outlier. In Figure 7, the blue points are the measured data points, the outliers are represented by black boxes, and the upper and lower fine lines indicate the upper and lower thresholds for outlier detection, respectively. These two fine lines represent the outlier identification range of the profile using the moving mean identifier, and the data falling outside the thresholds are considered as outliers.

Then, the moving mean identifier is used to identify the two-point outliers shown in Figure 6. The results are shown in Figures 8 and 9. As shown in Figure 8, when the window lengths are 17, only one outlier in the middle is identified. When the window lengths reach 28, two outliers can be correctly identified, as shown in Figure 9. As described above,
when the window lengths are $17 \leq w_l < 28$, only one outlier can be identified and the identification omission occurs, and the window lengths $w_l \geq 28$ can correctly identify the double-point outliers.

According to the above identification process, it is found that when the moving mean is used to deal with outliers, the window lengths need to change with different outliers, so the reasonable selection of window lengths becomes the focus of the moving mean identifier. At the same time, in the middle of the profile where the outlier occurs, the upper and lower detection thresholds increase greatly. This is because the existence of the outliers makes the local mean in the middle increase greatly, indicating that the moving mean identifier is less robust. With the increase of window lengths, the threshold ranges of the central data will decrease, the upper and lower threshold curves will gradually smooth, and the identification ability of outliers in the middle part of the profile will also improve. However, as the window width
increases, the upper and lower threshold ranges on both sides of the curved profile increase, that is, the ability to recognize outliers is decreasing on both sides of the profile.

3.2. Hampel Identifier Recognizes Isolated Outliers. Hampel identification method or Hampel identifier uses the median and median absolute deviation as a robust estimate of the location and spread of the outliers, that is, using the median value to estimate the data position, using the median absolute deviation to estimate the standard deviation of the data, thereby effectively identifying the outliers.

Hampel identifier is defined as that, for data sequences \( x_1, x_2, x_3, \ldots, x_n \), the number of neighbors on either side of \( w_l \) is \( k \), then the moving window lengths are \( 2k + 1 \), and the local median is \( m_i \), as shown in equation (2).

\[
m_i = \text{median}(x_{i-k}, x_{i-k+1}, x_{i-k+2}, \ldots, x_i, x_{i+k-2}, x_{i+k-1}, x_{i+k}).
\]  

The scale estimate of median estimated deviation (MADse) is \( S_i \),

\[
S_i = k \cdot \text{median}(|x_{i-k} - m_i|, \ldots, |x_{i+k} - m_i|),
\]  

where \( k = (1/(\sqrt{2} \text{ erfc}^{-1}(1/2))) \approx 1.4826 \) is the unbiased estimation of the Gaussian distribution.

\[
y_i = \begin{cases} x_i, & |x_i - m_i| \leq tS_i, \\ \text{outlier,} & |x_i - m_i| > tS_i. \end{cases}
\]  

It can be seen from equation (4), when the difference between the measured data and the local median is greater than \( t \) times MADse, the measured value is considered to be an outlier, and the value of \( t \) is 3 generally. When \( t = 0 \),
then $y_i = m_i$, the Hampel identifier becomes the standard median filter. In addition, the Hampel identification method uses symmetric moving window, so the strategy mentioned in the above literature [26, 27] is referred to deal with endpoints.

The Hampel identification method was used to identify outliers of the above-mentioned profiles, as shown in Figures 10 and 11. As can be seen from the figure, when $t = 3$ (i.e., 3 times MADse) and $k = 3$ (i.e., $wl$ is 7), the two kinds of isolated outliers can be well identified. Compared
with the moving mean identifier, the upper and lower threshold detection ranges are smaller, especially in the central of the profile where the outlier occurs, which is a little affected by the outliers, and there is no significant change in the upper and lower thresholds.

Observe Figures 10 and 11, compared with the moving mean method, the data gap between the two sides of the profile, the upper and lower threshold range, and the fluctuation of the Hampel identification method are significantly reduced. The upper and lower thresholds fluctuate greatly when the data interval on the right side of the profile is large, as shown on the right side of Figure 11. This kind of fluctuation is mainly caused by the increase of the data interval, which causes the difference between the data and the median to increase. Therefore, the large data interval on both sides of the profile will affect its outlier identification ability.

The value of $t$ in the Hampel identifier, namely, the scale factor for outlier detection, is an important parameter in the process of outlier detection. The value of $t$ directly determines the identification range of the outliers, and its identification range is $\pm t$ times MAD$_{se}$. When $t$ increases gradually, the range of the upper and lower thresholds will also increase exponentially. For the profile shown in Figure 5, when the $t$ value is 3 and 4, the leftmost data is detected as an outlier except for the middle field outlier. When the $t$ value reaches 5, as shown in Figure 12, the upper and lower thresholds

![Figure 10: Hampel identifier recognizes single-dot outlier ($i = 149$, $wl = 7$).](image1)

![Figure 11: Hampel identifier recognizes double-dot outlier ($i = 159$, $wl = 7$).](image2)
are significantly increased, and outliers are no longer detected on both sides of the profile. For the profile shown in Figure 6, the $t$ value is in the range of 3~12, the left starting data point can be effectively detected as an outlier except for the double-point outlier found in the middle. When the $t$ is 13, as shown in Figure 13, the outliers in the middle can be detected, but the outliers on the left cannot be detected. At this time, the upper and lower detection thresholds have increased significantly compared with Figure 11. Comparing the upper and lower threshold ranges on the right side of Figures 11 and 13, it is found that there is a huge difference. The recognition ability of outliers on the right side of Figure 11 had been seriously degraded, and the peak value of threshold on the right side of profile increased from 5 mm to 18 mm. For the profile start and endpoints, the Hampel identification method uses this point as the center to complement 0 and then judges, so the start value is easier to be judged as an outlier.

Moving window length is another important parameter that affects the identification ability of the Hampel method. When window length changes, the recognition range will also change accordingly, thus affecting the recognition effect of outliers. As for the profile shown in Figure 5, when the window length is gradually increased from 5 mm to 18 mm. For the profile start and endpoints, the Hampel identification method uses this point as the center to complement 0 and then judges, so the start value is easier to be judged as an outlier. In comparison with Figure 10, it is found that the upper and lower threshold ranges for the identification, especially on both sides of the contour, are significantly increased. For the contour shown in Figure 6, as the window width increases, the upper and lower threshold ranges will also gradually increase. When the window length is increased to 17, the outliers in the middle part of the profile and the left side is detected simultaneously; the data endpoint on the right side is also measured as the outliers. When the window width reaches 75, as shown in Figure 15, only the middle outliers and the right endpoint are detected as outliers. At this time, the upper and lower threshold ranges of profile judgment no longer conform to the profile trend and have no practical detection significance, but this shows the influence of window length change on outlier identification. Therefore, if the window width of the Hampel identification method increases gradually, the upper and lower threshold range for judging outliers also increases gradually, especially on both sides of the profile, which has a certain impact on the detection of outliers at the starting and stopping points on both sides of the profile.

In addition, multiple isolated outliers are likely to occur simultaneously in the measurement, and it is necessary to verify the ability to identify multioutliers for the Hampel identifier. Firstly, a dozen random isolated outliers were added on the profile of Figure 5, and then the Hampel identification method was used to identify the multioutliers; the recognition effect is shown in Figure 16. In the case of 3 times MADse and $wl = 7$, the Hampel identifier can effectively identify all the multioutliers, including different single-dot outliers and double-dot outliers, as shown in the figure. Moreover, the upper and lower threshold detection ranges are less affected by the multioutliers, which indicates the effectiveness of the Hampel identifier in identifying multioutliers. However, for the moving mean identifier, it is difficult to correctly identify multioutliers, and its upper and lower thresholds will also fluctuate greatly with each outlier, so it is not a good choice for multioutliers.

4. Analysis of Profile-Fitting Effect

The example in the previous section shows that the Hampel identification method can effectively eliminate the isolated outliers in the profile, including some initial outliers. We used the detection result when MADse was 3 times and window length was 7 to remove the outliers, as shown in
Figures 10 and 11. The profiles were fitted with the nonlinear least square method, and the original profiles with outliers retained were fitted too. Thus, the comparison of ellipse-fitting results before and after the removal of outliers is obtained, as shown in Figures 17 and 18.

Figures 17 and 18, respectively, show the comparison of ellipse-fitting effect before and after the outliers are eliminated in the profile shown in Figures 5 and 6. In the figure, the red thick point is the original data, the black frame is the outlier identified, the gray line is the ellipse contour fitted by the original data, the blue line is the ellipse contour fitted after the outliers are removed, “+” is the circle center of the ellipse fitted by the original data, and “*” is the center of the fitted ellipse after the outliers are removed. It can be clearly seen from the figure that, before the outliers are eliminated, the fitting results of the original profile data have been obviously out of reality, while after the outliers are removed by the Hampel method, the fitted ellipse conforms to the reality. Table 1 is the fitting ellipse parameters based on the measurement results. It can be seen from the figure and table that there are great differences between the fitted ellipses before and after the outliers are removed. For instance, before the outlier is removed, the ellipse of $i = 149$ has a major axis diameter of 13.1582 mm and an eccentricity of 0.8766; after
the outliers were removed, the major axis diameter was 19.8214 mm and the eccentricity was 0.237. For the data of $i = 159$, before the outlier is excluded, the major axis diameter of the ellipse is 12.6312 mm and the eccentricity is 0.9223, which is more flat; while the outliers were removed, the major axis diameter is 20.2996 mm and the eccentricity is 0.228, which is closer to the circle.

At the same time, the fitting error is analyzed after eliminating the outliers. The results are shown in Table 2.

For the profile of $i = 149$, before and after the elimination of the outliers, the sum of squares due to error (SSE) are $5.4535e-06$ and $5.7164e-07$; root mean square error (RMSE) was $1.2059e-04$ and $3.9148e-05$, respectively. For the profile of $i = 159$, the SSE is $8.4523e-06$ and $5.9981e-07$, and the RMSE is $1.5074e-04$ and $4.0318e-05$, respectively. It can be clearly seen that after removing the outliers, the sum of squares due to error and root mean square error of the fitting data are reduced by an order of magnitude, indicating that the removal of outliers greatly reduces the fitting error and the fitting results are more accurate.

5. Conclusions

In this paper, machine vision technology based on linear structured light is used to measure the external profile of
the bar. However, the measured profile data has obvious isolated outliers. Aiming at this kind of outliers, the real profile data were identified by the moving mean identification method and Hampel identification method, respectively. During the identification process, it is found that the sample mean and sample standard deviation adopted by the moving mean method are both estimated values, which are easily affected by outliers. The Hampel method is more robust by using the moving median deviation, that is, the median and absolute median deviations of data are used as outlier judgment indicators of location and distribution, respectively, so the recognition effect is better. At the same time, the influence of the different window lengths and detection threshold for the Hampel method on the identification effect is also discussed; furthermore, its ability to identify multioutliers was also tested. Then the nonlinear least squares method is used.

**Figure 17:** The comparison of profile fitting (i = 149).

**Figure 18:** The comparison of profile fitting (i = 159).
to fit the profile data before and after the outlier elimination, the main parameters of the fitted ellipse are obtained, and the fitting error is analyzed. Through the comparison and analysis of the fitting results, it is found that the data is contaminated by outliers, which will lead to the great deviation of profiler fitting results and obtain the wrong contour size parameters. The Hampel method can effectively identify the isolated outliers, which can effectively improve the accuracy and fitting effect of online measurement of bar profiles.

**Data Availability**

The scanning data of this study are not changed during the modification process.

**Conflicts of Interest**

The authors declare that they have no competing interests.

**Authors’ Contributions**

All authors equally contributed to this paper.

**Acknowledgments**

This work was supported by the Excellent Innovation Project for Graduate Students of Shanxi Province (20143021) and sponsored by the Fund for Shanxi “1331 Project” Key Subjects Construction.

**References**


