An Improved Optical Flow Algorithm Based on Mask-R-CNN and K-Means for Velocity Calculation

Yahui Peng, Xiaochen Liu, Chong Shen, Haoqian Huang, Donghua Zhao, Huiliang Cao, and Xiaoting Guo

Key Laboratory of Instrumentation Science & Dynamic Measurement, Ministry of Education, School of Instrument and Electronics, North University of China, Taiyuan 030051, China
College of Energy & Electric Engineering, Hohai University, Nanjing 210098, China
* Correspondence: guoxiaoting@nuc.edu.cn
† Yahui Peng and Xiaochen Liu have the equal contribution to this work.

Received: 13 April 2019; Accepted: 11 July 2019; Published: 13 July 2019

Abstract: Aiming at enhancing the accuracy and reliability of velocity calculation in vision navigation, an improved method is proposed in this paper. The method integrates Mask-R-CNN (Mask Region-based Convolutional Neural Network) and K-Means with the pyramid Lucas Kanade algorithm in order to reduce the harmful effect of moving objects on velocity calculation. Firstly, Mask-R-CNN is used to recognize the objects which have motions relative to the ground and covers them with masks to enhance the similarity between pixels and to reduce the impacts of the noisy moving pixels. Then, the pyramid Lucas Kanade algorithm is used to calculate the optical flow value. Finally, the value is clustered by the K-Means algorithm to abandon the outliers, and vehicle velocity is calculated by the processed optical flow. The prominent advantages of the proposed algorithm are (i) decreasing the bad impacts to velocity calculation, due to the objects which have relative motions; (ii) obtaining the correct optical flow sets and velocity calculation outputs with less fluctuation; and (iii) the applicability enhancement of the optical flow algorithm in complex navigation environment. The proposed algorithm is tested by actual experiments. Results with superior precision and reliability show the feasibility and effectiveness of the proposed method for vehicle velocity calculation in vision navigation system.

Keywords: velocity calculation; optical flow; pyramid LK; Mask-R-CNN; clustering algorithm

1. Introduction

The unmanned combat platform is a developing war mode in future wars, where a micro-unmanned aerial vehicle (UAV) occupies a significant position. An UAV is mission-centered design, which breaks away from the constraints of human safety, human physiology and psychology. It has a series of advantages, such as small size and light weight, strong mobility, good stealth, long idle time, large combat radius and large payload. Because of its relatively small size and good concealment, micro-UAVs can be widely used in future wars, which is conducive to effectively attack without the enemy’s preparedness. The biggest advantage of the UAV is that it does not require the human to be co-located with the aircraft. Thus, it is free from human physiological limitations and health damage. When performing some dangerous tasks, it does not need to worry about casualties. Therefore, it is suitable for performing dull, dirty and dangerous missions, which is in the so-called “3D” environment. However, the premise of the UAV consummate flight during the execution of the mission is precision navigation. Precise aircraft navigation information such as velocity is the basis of the stability control of the UAV [1–5]. The most commonly techniques used to obtain vehicle velocity include the inertial navigation system (INS) [6–8], global position system (GPS) [9,10], geomagnetic...
navigation and a GPS/INS [11–13] integrated navigation system. However, the methods mentioned above have problems such as integral and cumulative error, signal loss and the possibility to be interfered by continuously changing electromagnetic signal. With the advantages of low cost, rich navigation information, and strong anti-electromagnetic interference capability, vision navigation has undergone a great development in recent years and can be a good alternative to the common navigation system [14,15].

To obtain precise vehicle velocity, optical flow algorithm is an important method used in vision navigation. The so-called optical flow is a series of information of an object in changing images formed on the retina of a human or insect when they observe the moving object. In flight, birds and insects are inspired by optical flow to control the velocity of flight, avoid obstacles, land and take off autonomously [16]. The concept of optical flow was first proposed by Gibson [17]. Later, Poggio and Reiehardt presented an approach to compute the motion of each pixel in an image [18], which can be considered a rough flow method. The first practical optical flow model was established by the classical work of Horn and Schunck (HS) [19]. It assumes that the brightness of a pixel keeps constant during a short time interval, which is known as the brightness constancy assumption (BCA) [20]. Optical flow not only provides important information about the unknown environment, but also helps to determine the direction and the velocity of the aircraft. Due to this characteristic, optical flow is one of the primary techniques for UAV navigation. Optical flow can be widely used for obstacle detection and collision avoidance. Meanwhile it can help to calculate the velocity of global movement.

Nowadays, the mainstream optical flow calculation techniques include the frame differential method [21], pyramid Lucas Kanade (LK) algorithm and convolution network-based optical flow algorithms [22]. Aiming at a nonlinear controller for a vertical take-off and landing of UAV, the measurement of the average optical flow is exploited to enable hover and landing control on a moving platform. The average optical flow is obtained from a textured target plane. An embedded inertial measurement unit provides additional information for derotation of the flow and to have precise output based on pyramid LK algorithm [23]. To calculate dense optical flow, rich descriptors are integrated into the variational optical flow [24]. A dense optical flow field, which almost has the same high accuracy as variational optical flow, can be estimated. Reaching out to new domains of motion analysis, the solution of the dense sampling problem can be done [25]. Based on the derived objective function in optical flow calculation, which formalizes the median filtering heuristic, Sun and his team integrate flow estimations over large spatial neighborhoods and preserve motion details well [26]. Basically, the optical flow calculation methods mentioned above improve the accuracy of outputs by ameliorating the calculation process.

In the Horn and Schunck algorithm [19], it is assumed that the optical flow field satisfies the constraints of global smoothness based on satisfying the basic optical flow equation. Smoothing constraints are applied to the whole optical flow field. It is based on dense optical flow computation. Each object point of the figure should be calculated, in order to obtain the optical flow vector of the points. However, in spare optical flow computation, only the point which can describe the motion characteristics of moving objects should be calculated. So, if the dense optical flow computation is used in our algorithm, the image processing speed of this method is sometimes slower than the input speed of video image sequence. The system processing is not real-time, and the algorithm calculation is complex. Also, spare optical flow computation improves the speed and efficiency of the operation. LK algorithm is one of the optical flow algorithms that are based on spare optical flow computation. The accuracy of dense optical flow is better than that of spare optical flow computation. The goal what we want to reach is to improve the accuracy of optical flow and velocity calculation. So, LK algorithm is more suitable for the experiment in the research.

In the semi-visual odometry developed by the Scaramuzza lab [27], pyramid LK algorithm is also applied to some extent. According to the assumption of invariant gray level, frame-to-frame pose estimation is realized by minimizing the gray value error of pixels. At the same time, the pyramid processing method is used to do the estimation. It starts from the top of the pyramid, and the results of
the upper layer are taken as the initial values of the next layer, which are the same as the idea of the Pyramid LK. This method can improve the accuracy of the algorithm.

However, these algorithms are not adaptable enough to accurately calculate the velocity of aircraft in complex environments, where people and other objects like animals and cars, which have motions relative to the ground, may lead to extra pixel movement. Environment noises resulting from interference of moving objects may result in wrong optical flow values and reduce the accuracy of vehicle velocity calculation. Optical flow method judges the movement of objects according to the movement of pixels between images. With the ground as the reference frame, if the objects on the ground remain unchanged during the UAV flight, the calculation result only includes the UAV flight speed. However, if the target takes the ground as the reference frame and has relative motion, the calculation result of optical flow method will include the target moving at the same time, thus introducing errors. When the velocity error is large, the navigation will be greatly disturbed.

During the velocity calculation, there may be some moving objects inevitably in the two adjacent images captured by the UAV, such as people, animals, vehicles, etc. Indeed, motions captured by the UAV can be divided into two parts; global motions caused by the changes of background and local motions caused by moving foreground in the scene. As these motions have different speeds, the local motion may have serious impact on optical flow values. Considering these complex environments in optical flow calculation, this paper proposes an improved pyramid LK optical flow algorithm by combining Mask-R-CNN (Mask Region-based Convolutional Neural Network) [28] and K-Means [29] to reduce the impacts of extra moving pixels. The main contributions of the paper are as follows:

1. Train the weights of Mask-R-CNN based on COCO dataset [30] to create moving target recognition;
2. Based on the trained NN (Neural Network) model, recognize the objects with motions relative to the ground, and cover them with mask and preserve the background at the same time;
3. Use pyramid LK algorithm to deal with the processed images;
4. Add a mask on the moving object to enhance the similarity between background pixels and object pixels, so only the global motion is completely retained, and the error caused by the moving object can be reduced [31].

As the original image is processed by Mask-R-CNN, some information of the image is lost. Thus, singular optical flow values may appear when calculated by pyramid LK algorithm. Then, the K-Means method is used to cluster the optical flow values to get correct optical flow data, which are calculated based on global motion in the processed images. It is the basic principle to obtain precise vehicle velocity.

The paper is organized as follows: the improved optical flow algorithm based on Mask-R-CNN and K-Means is given in Section 2. Section 3 gives the results of experiment and comparisons to show that the proposed algorithm can enhance the accuracy and reliability of vehicle velocity calculated by optical flow. The paper ends with a conclusion in Section 4.

2. Algorithm Design

When pyramid LK algorithm is used for velocity calculation, local moving pixels may sometimes lead to wrong optical value. Therefore, large singular values would occur during calculation. Further, it affects the precision of UAV velocity in vision navigation system. So, Mask-R-CNN is introduced to perform object instance segmentation, in order to recognize and cover the objects with local movements. Additionally, K-Means clustering method can eliminate the singular optical values calculated by pyramid LK algorithm. The process is shown in Figure 1.

2.1. Pyramid LK Algorithm

The optical flow algorithm can be used to calculate the instantaneous velocity by comparing the corresponding features in two adjacent images. It estimates the optical flow value based on the assumption that the brightness of the feature pixel points does not change [32] between two adjacent images. The instantaneous velocity is called the optical flow vector.
where the solution of equation that means the moving velocity is called optical flow.

Equation (1) can be expanded by Taylor series:

\[
E(x + u\Delta t, y + v\Delta t, t + \Delta t) = E(x, y, t)
\]  

where \(x\) and \(y\) represent the horizontal and vertical positions of pixel points respectively, and \(E(x,y)\) represent the brightness of point \((x,y)\) at time \(t\).

The brightness along the trajectory does not change. The equation can be expanded as:

\[
\frac{dE(x, y, t)}{dt} = 0
\]  

The differential form of Equation (2) is:

\[
\frac{\partial E}{\partial x} \frac{dx}{dt} + \frac{\partial E}{\partial y} \frac{dy}{dt} + \frac{\partial E}{\partial t} = 0
\]  

When the gray value of the image changes with the three variables of \(x, y, t\), the left side of the Equation (1) can be expanded by Taylor series:

\[
E_x u + E_y v = -E_t
\]  

where \(\frac{dx}{dt} = E_x\), \(\frac{dy}{dt} = E_y\) denote the horizontal and vertical gradient of pixel; \(\frac{dt}{dt} = E_t\) is the time gradient of a pixel; \(\frac{dx}{dt}\) and \(\frac{dy}{dt}\) are the horizontal and vertical components of the velocity vector, which is called optical flow.

Due to the negligible time, motions between two frames can be regarded as linear in a relatively short time, thus \(\frac{dx}{dt} = u\), \(\frac{dy}{dt} = v\).

If the equation is set up with \(N\) identical pixels, the following can be obtained:

\[
\begin{bmatrix}
E_{x1} & E_{y1} \\
E_{x2} & E_{y2} \\
\vdots & \vdots \\
E_{xN} & E_{yN}
\end{bmatrix}
\begin{bmatrix}
u \\
v
\end{bmatrix} =
\begin{bmatrix}
E_{t1} \\
E_{t2} \\
\vdots \\
E_{tN}
\end{bmatrix}
\]  

where the solution of equation that means the moving velocity \(u\) and \(v\) is easy to calculate.
The precision of LK algorithm is not adequate when processing two adjacent images. To solve this problem, the Gaussian image pyramid technique is introduced, which satisfies the small motion hypothesis by continuously subdividing the velocity and improves the noise immunity of the algorithm. The core idea of the pyramid LK algorithm is to construct the image to be compared into an image sequence with a resolution ranging from small to large. Firstly, the image with a low resolution is calculated, and then the result is iterated as the initial value to the image with a high spatial resolution. That is, from the top layer of the pyramid to the bottom layer until the lowest value of optical flow value is calculated. The process of pyramid LK algorithm is displayed in Figure 2.

**Figure 2. Pyramid LK algorithm.**

**Step 1**: for the adjacent reference frame and the current frame in the image sequence, the levels of different resolutions are sampled according to the decreasing precision, and the original image is taken as the base of the pyramid \((L = 0)\). When the image is divided into a certain level, the motion parameters with large displacement will become small enough to meet the constraints of optical flow calculation.

**Step 2**: calculation, and in turn down from the top layer set \(g_L\) as \(L\) initial estimate optical flow vector, \(\Delta f_L\) as optical flow calculated in \(L\) layer as a result, the mapping relationship of between the layers is:

\[
g_{L-1} = k(g_L + \Delta f_L) \quad (6)
\]

**Step 3**: set the total number of pyramid layers as \(N\) and initialize the optical flux \(g_{N-1} = 0\) at the top of the pyramid. The motion parameter of the original image can be obtained by calculating from the top down as:

\[
f = g_0 + \Delta f_0 = \sum_{L=0}^{N-1} k_L \Delta f_L \quad (7)
\]

**Step 4**: by taking the modulus of the two velocity components obtained from equation (9), the global motion parameter estimation results can be obtained:

\[
\begin{aligned}
x &= |u| \\
y &= |v|
\end{aligned} \quad (8)
\]

2.2. K-Means

Clustering algorithm can be used in pattern recognition, image processing, machine learning and statistics. One of the most popular clustering methods is the K-Means algorithm, where clusters are identified by minimizing the clustering error. When using the K-Means algorithm for clustering, it does not need to know the category of data in advance, but only needs to classify according to the distance or similarity between data. So, it is an unsupervised clustering algorithm [33].

K-Means clustering algorithm is an algorithm based on partition, and the data is divided into \(k\) preset clusters, \(k < n\). Each cluster is represented by a clustering center, denoted by \(\{C_i\}_{i=1}^{k}\). After obtaining the clustering center of each cluster, all data are mapped to a certain clustering center, that is, the cluster where the clustering center is added. Finally, all data will be divided into corresponding clustering centers, and \(k\) clusters will be obtained, denoted as \(\{S_i\}_{i=1}^{k}\).
Assuming the data-set as $D = \{ x_i \}_{i=1}^n$, $x \in \mathbb{R}^d$, and the objective function is:

$$I(S) = \arg\min \sum_{i=1}^n \sum_{x \in S_i} \|x - c_i\|_2$$

(9)

$$c_i = \frac{1}{|S_i|} \sum_{x \in S_i} x$$

(10)

where $|S_i|$ means the quantity of the sample in category $S_i$.

Generally, K-Means clustering algorithm adopts iterative strategy to solve the local optimal solution of the objective function. The clustering process is shown in Table 1.

| Step 1 | Set a predetermined RoI for each point in this feature map to achieve multiple candidate points: $N$ points: |
|--------|--------------------------------------------------|
| Step 2 | Calculate which category each point belongs to $K$ center points: Find all the data points that belong to this category Modify the coordinates to the coordinates of the center point |
| Step 3 | K-Means algorithm does not converge |
| Step 4 | Output |

The result of pyramid LK output is optical flow value, denoting the movement of pixel. The K-Means clustering algorithm is used to cluster the results into three categories; namely, smaller value, moderate value and larger value. The moderate value is selected as the correct result, which reduces the bad influence of the wrong optical flow value and reduces the error of speed.

2.3. Mask-R-CNN

Mask-R-CNN is a conceptually simple and flexible method for object instance segmentation. A region proposal network (RPN) is utilized for region of interest extraction [31]. While predicting the box offset and class for each region of interest, it outputs a binary mask. FCN (Fully Convolutional Network) can be applied to each region of interest (RoI) for the prediction of a segmentation mask, since the mask directly presents the correspondence between pixels by convolution [34–36]. The framework of Mask-R-CNN is given in Figure 3.

Figure 3. Mask-R-CNN framework for instance segment.
The specific procedure of the Mask-R-CNN is as follows.

**Step 1:** input of the normalized captured image into the pre-trained network in order to get the corresponding feature map.

**Step 2:** set a predetermined RoI for each point in this feature map to achieve multiple candidate RoIs.

**Step 3:** feed the candidate RoI into the RPN network for binary classification (foreground or background) and Bounding Box regression; part of the candidate RoIs are then filtered out.

**Step 4:** RoIAlign operation is performed on the remaining RoIs (that is, the pixel of the original image and the feature map is first corresponded, and then the feature map is corresponded to the fixed feature).

**Step 5:** classify these RoIs, do Bounding Box Regression, and make MASK generation (FCN operation in each RoI);

Ultimately, there are three branches generated, which contain the information to predict reg-layer, cls-layer and object mask. Then, the first two branches are used for bounding-box classification and regression in parallel. The third branch is used to output the binary mask of the objective feature \([36,37]\).

### 2.4. Pyramid LK Algorithm Based on Mask-R-CNN and K-Means

The framework has been introduced in the top of Section 2. In this part, the process of the improved algorithm is given in detail. In this paper, vehicles and people, which would often produce locally moving targets in the road taken by the UAV, are presupposed as recognizable and masked targets, as well as trees and buildings, which would not move, are under special circumstances. Therefore, when recognizing, targets such as trees and buildings are not taken as objects of recognition.

**Step 1:** input of the normalized image captured by UAV into the main network as Figure 4.

![Figure 4. Two adjacent images captured by UAV: (a) the first image; (b) the second image.](image)

**Step 2:** feature extraction and generation of regions of interest.
The image is sent to the main network to extract the data, and then the region proposal network is used to find the region of interest. Subsequently, a layer called RoIAlign is adopted, which accurately aligns the extracted features with the input to improve the accuracy of the object mask \([38]\).

**Step 3:** propose the box offset, the class, and the mask. The result of Mask-R-CNN is displayed in Figure 5.

**Step 4:** determine the color of the mask. Several background pixels are sampled around the target. The RGB (Red Green Blue) values of the pixels are extracted, and then the average values of R, G and B channels are obtained respectively. These three values are used to increase the RGB value of the mask.

**Step 5:** cover the objects that have relative motion with masks in a similar color as shown in Figure 6. The introduction of similar color masks can bring a part of the correct optical flow. There are still erroneous optical streams at the edge of the mask. If the positions of the object are ignored, plenty of the correct optical flow values are abandoned. Therefore, the playback method of changing the pixel color to benign color retains more correct optical flow value than the playback method of directly
ignoring the pixel position. The results of this method are more accurate. In this way, the similarity between pixels can be enhanced and extra pixels movement has less impact on optical flow calculation.

![Figure 5. The results of Mask-R-CNN.](image)

![Figure 6. Cover the objects with similar color.](image)

**Step 6:** input two consecutive frames of processed images into the pyramid LK algorithm.

**Step 7:** by selecting appropriate N and K, the calculated optical flow direction in the N layer can be obtained from Equation (5), and the initial optical flow estimation vector of the next layer can be obtained by substituting it into Equation (6). At last, the iterative calculation of Equation (7) is carried out once to obtain the optical flow estimation vector of the original image sequence and obtain the optical flow value result \([39,40]\).

**Step 8:** input the optical flow matrix calculated by pyramid LK algorithm into the K-Means clustering algorithm and select the class closest to the real optical flow value.

2.5. **Practical Analysis of the Algorithm**

The flight speed of the UAV and the sampling rate of the camera affect the accuracy of the algorithm. These two experimental parameters are mutually compatible in order to make the experiment satisfy the conditions of optical flow calculation. If the flight speed of the aircraft is slow, the higher the frequency of image acquisition and the smaller the movement between the two frames at the same flight speed, the more accurate the calculation will be. If the sampling rate of the camera is too low, the flight speed of the aircraft should be reduced appropriately to satisfy the computational condition of continuous small motion between two frames of images. Thus, if these parameters are suitable, the algorithm can be useful in velocity calculation. Mask-R-CNN recognition weights need powerful GPU (Graphics Processing Unit) training when they are trained, but when the training weights are used directly for target recognition, only the CPU (Central Processing Unit) can be used to process them. Thus, the algorithm can be applied for real-time practical experiments.
3. Experiments and Evaluation

3.1. Experimental Equipment

To validate the performance of the proposed algorithm for velocity calculation in vision navigation, a practical experiment based on an UAV was carried out in the campus of North University of China, Taiyuan, China. In the experiment, the flight speed of our UAV was about 8.4 m/s. The optical flow sensor was calibrated by the method proposed in reference [41], and the internal and external parameters of the sensor were obtained. The NovAtel ProPak6 GPS was used in this experiment as the high-precision speed and position reference. After calculation, about 50 pixels were moved between the two frames. The parameters of cameras and reference are shown in Table 2. The experimental facilities are shown in Figure 7.

Table 2. Parameters of experimental equipment.

| Optical Flow Camera | GPS |
|---------------------|-----|
| Sensor brand        | Sony IMX179 |
| Sensor category     | CMOS |
| Lens size           | 1/3.2 inch |
| Pixel size          | 1.4 µm |
| Focal length        | 10 mm |
| Sampling rate       | 30 Hz |
| Resolution          | 1280 × 760 |
| Brand               | ProPak6 |
| Position accuracy   | 1 cm + 1 ppm |
| Velocity accuracy   | 0.03 m/s |
| Time accuracy       | 20 ns |

Figure 7. Experimental facilities.

3.2. The Evaluation of the Improved Optical Flow Algorithm for Velocity Calculation

To evaluate the performance of the proposed method in moving objects recognition, the addition of similar color masks, as well as the final velocity calculation based on optical flow, three kinds of typical application circumstances are considered in our experiments, which are explained in detail as follows. As the main purpose of our algorithm is to weaken the influence of the moving objects on the velocity calculation based on optical flow, the application circumstances are cautiously chosen as different cases exist for different moving objects.

Besides, the proposed pyramid LK algorithm based on MASK-R-CNN and K-Means (LK + MCR + KM) is investigated to evaluate the performance of velocity calculation, with three other algorithms
pyramid LK algorithm (LK), pyramid LK based on Mask-R-CNN (LK + MRC), and pyramid LK based on K-Means (LK + KM) for comparison.

In the experiment, in order to verify the experimental results more conveniently and rationally, only one positive direction (X direction or Y direction) of the pixel movement is carried out in one experiment. Therefore, the accuracy of the results can be demonstrated by decomposing the vectors and extracting only the values of the experimental direction and comparing the decomposition results with the standard values. If the calculation results are correct, the direction of the vector should be the same as the direction of the moving pixels. That is, there are no components in other directions. Then, the velocity error can be calculated.

3.2.1. Experiment One in the Normal Application Circumstance

Experiment 1 is executed in normal application circumstance with many cars and few persons. Mask-R-CNN helps to recognize the objects and the RGB of the masks, as it made it the same as the background. The progress of dealing with the images is shown in Figure 8.

![Figure 8](image-url)

**Figure 8.** (a) Image in experiment 1; (b) the result of instance segmentation; (c) covering the objects with similar color masks.

The comparison of the four different algorithms in velocity calculation is illustrated in Figure 9. Meanwhile, in order to prove the superiority of the proposed algorithm, the quantitative comparisons with other algorithms are listed in Tables 3 and 4.

| Algorithm                        | LK          | LK + Mask-R-CNN | LK + KM | LK + Mask-R-CNN + KM |
|----------------------------------|-------------|-----------------|---------|----------------------|
| RMSE                             | 2.1632      | 2.1121          | 1.1781  | 0.6845               |
| STD                              | 1.5276      | 1.5522          | 0.8628  | 0.5011               |

Table 3. Velocity errors of four algorithms in the X direction.
To prove the superiority of the proposed algorithm, the quantitative comparisons with other algorithms are listed in Tables 3 and 4. The addition of the Mask-R-CNN algorithm eliminates the large deviation caused by relative motion because it covers the objects, while the K-Means clustering algorithm makes the velocity curve closer to the basic value and reduces the impact of the wrong optical flow. It means that the proposed algorithm performed better than the other three algorithms and can be applied to the enhancement of velocity calculation precision.

It can be easily seen from Tables 3 and 4, that compared to the normal pyramid LK algorithm and pyramid LK based on K-Means algorithm, the performance of the proposed method was largely improved, which is illustrated by the increase in the accuracy the decrease in the dispersion degree of data. Specifically, decrease of the standard deviation means that the results are more average and concentrated, which indicates that the velocity result are more stable. The least RMES obtained by the proposed method implies that its velocity calculation accuracy is superior to the other three algorithms and can be applied to the enhancement of velocity calculation precision.

3.2.2. Experiment Two in the Circumstance with Many Moving Objects

Experiment 1 considers the scene mainly with many cars. Compared with people, the moving direction of the cars is relatively assured. In order to verify the algorithm application in complex circumstance, experiment 2 is carried out in normal light with many people and other moving objects. Compared with the first experiment, experiment 2 is more complicated, because people have uncertain moving velocities and directions. The progress of dealing with images is shown in Figure 10.

The comparison of the four different algorithms in velocity calculation is illustrated in Figure 11. To prove the superiority of the proposed algorithm, the quantitative comparisons with other algorithms are listed in Tables 5 and 6.

Figure 9 shows the velocity error comparison after the joint of clustering algorithm and Mask-R-CNN. Also, it indicates the improvement of velocity accuracy in X and Y direction in four algorithms. The addition of the Mask-R-CNN algorithm eliminates the large deviation caused by relative motion because it covers the objects, while the K-Means clustering algorithm makes the velocity curve closer to the basic value and reduces the impact of the wrong optical flow. It means that the proposed algorithm performed better than the other three algorithms and can be applied to the enhancement of velocity calculation precision.

Table 4. Velocity errors of four algorithms in the Y direction.

|             | LK                  | LK + Mask-R-CNN     | LK + KM               | LK + Mask-R-CNN + KM |
|-------------|---------------------|---------------------|-----------------------|----------------------|
| RMSE        | 1.6447              | 1.4352              | 1.0739                | 0.5677               |
| STD         | 1.1686              | 1.0039              | 0.7867                | 0.4155               |

Table 5. Velocity errors of four algorithms in the X direction.

|             | LK                  | LK + Mask-R-CNN     | LK + KM               | LK + Mask-R-CNN + KM |
|-------------|---------------------|---------------------|-----------------------|----------------------|
| RMSE        | 3.6609              | 2.6738              | 2.0163                | 1.1772               |
| STD         | 2.5952              | 1.8290              | 1.4750                | 0.8150               |
uncertain moving velocities and directions. The progress of dealing with images is shown in Figure 10.

Figure 10. (a) The image in experiment 2; (b) the result of instance segmentation; (c) covering the objects with similar color masks.

The comparison of the four different algorithms in velocity calculation is illustrated in Figure 11.

To prove the superiority of the proposed algorithm, the quantitative comparisons with other algorithms are listed in Tables 5 and 6.

Table 5. Velocity errors of four algorithms in the X direction.

|                | LK       | LK + Mask-R-CNN | LK + KM | LK + Mask-R-CNN + KM |
|----------------|----------|-----------------|---------|-----------------------|
| RMSE           | 3.6609   | 2.6738          | 2.0163  | 1.1772                |
| STD            | 2.5952   | 1.8290          | 1.4750  | 0.8150                |

Although, there are more moving objects in the second scene than in the first one, the results are more fluctuant. However, it still performs better than the original algorithm. In complex environment, the proposed algorithm still has a good performance in velocity calculation. It can still recognize cars, people and other objects that are moving. Then, it calculates the optical value and the high of the images using only global motions. It reduces the quantity of singular optical value and makes the velocity error curve smoother and steadier than the original pyramid LK algorithm. The proposed algorithm also applies to where many people are walking. Also, it performs well in enhancing the accuracy and stability of vehicle velocity calculation. So, the improved pyramid LK algorithm is suitable for velocity calculation in complex environment with higher precision.
3.2.3. Experiment Three in the Circumstance with Dim Light

Working in dim-light environment is a great test for the UAV vision navigation system. It should be proved whether it can still keep high-precision velocity calculation in the environment with insufficient light. So, the last experiment is carried out in dusk with dim light to verify the proposed algorithm application in this environment. The effect of the instance segmentation and masks covering is shown in Figure 12.

![Figure 12](image1.png)  
(a) The image in experiment 3; (b) the result of instance segmentation; (c) covering the objects with similar color masks.

It can be directly seen from Figure 12 that in dim light, Mask-R-CNN can recognize moving objects even though it is difficult for human eyes to do so. It then covers them with similar color masks, thus enhancing the similarity between pixels. The comparison of the four different algorithms in velocity calculation is illustrated in Figure 13. To prove the superiority of the proposed algorithm, the quantitative comparisons with other algorithms are listed in Tables 7 and 8.

![Figure 13](image2.png)  
(a) Velocity error in X direction of experiment 3; (b) velocity error in Y direction of experiment 3.
In the third experiment with dim light, the result of the other three algorithms is more fluctuant. The proposed algorithm performs well with stability and few errors. In dark environment, the proposed algorithm can still recognize the moving objects and cover them with dark masks. Then, it calculated the optical value and vehicle velocity by the images. It can be seen from Figure 13 that the original pyramid LK algorithm calculates the velocity with more errors in dark environment than that in normal light. However, the proposed algorithm can determine the correct pixel movement even in dim light and calculate the speed of the vehicle with higher precision.

4. Conclusions

Velocity is one of the main parameters in visual navigation and the improvement of its accuracy is beneficial for promoting the development of visual navigation. An improved optical flow algorithm based on Mask-R-CNN and K-Means is studied in this paper to enhance the performance of velocity calculation in visual navigation. By integrating the technique of Mask-R-CNN and K-Means, the proposed algorithm can effectively rule out the outliers of optical flow values by reducing the effect of the dynamic objects in UAV captured images with similar color mask. More importantly, it can largely improve the accuracy of velocity with relatively correct optical flow values. The performance of the proposed algorithm is validated by an unmanned aerial vehicle flight experiment in the campus of North University of China. The experimental results demonstrate that dynamic objects on captured images can effectively be recognized and covered, and its side effect on optical flow and velocity calculation can be largely reduced, resulting in more optimal accuracy and reliability compared with traditional optical flow-based methods.

Author Contributions: Y.P. and X.L. contributed equally to this paper. Funding acquisition, C.S., H.C., H.H. and X.G.; investigation, Y.P.; methodology, Y.P., X.L., C.S., H.H. and X.G.; software, Y.P.; writing—original draft, Y.P.; writing—review and editing, C.S., D.Z., H.C. and X.G.

Funding: This work was supported in part by the National Natural Science Foundation of China (61603353, 51705477, 61703098), and the Pre-research Field Foundation (6140518010201) and the Scientific and Technology Innovation Programs of Higher Education Institutions in Shanxi (201802084), the Shanxi Province Science Foundation for Youths (201601D021067), the Program for the Top Young Academic Leaders of Higher Learning Institutions of Shanxi, the Young Academic Leaders Foundation in North University of China, Science Foundation of North University of China (XJJ201822), the 15th Graduate Science and Technology Project of North University of China (20181537), and the Fund for Shanxi “1331 Project” Key Subjects Construction, National Science Foundation of Jiangsu Province (BK20160699).

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Guo, X.; Tang, J.; Li, J.; Wang, C.; Shen, C.; Liu, J. Determine turntable coordinate system considering its non-orthogonality. Rev. Sci. Instrum. 2019, 90, 033704. [CrossRef] [PubMed]
2. Crivelli, T.; Fradet, M.; Conze, P.H.; Robert, P.; Perez, P. Robust Optical Flow Integration. IEEE Trans. Image Process. 2015, 24, 484–498. [CrossRef] [PubMed]
3. Shen, C.; Yang, J.; Tang, J.; Liu, J.; Cao, H. Note: Parallel processing algorithm of temperature and noise error for micro-electro-mechanical system gyroscope based on variational mode decomposition and augmented nonlinear differentiator. *Rev. Sci. Instrum.* 2018, 89, 076107. [CrossRef] [PubMed]

4. Li, P.; Sheng, G.; Zhang, X.; Wu, J.; Xu, B.; Liu, X.; Zhang, Y. Underwater terrain-aided navigation system based on combination matching algorithm. *Ista Trans.* 2018, 78, 80–87. [CrossRef] [PubMed]

5. Cao, H.; Zhang, Y.; Shen, C.; Liu, Y.; Wang, X. Temperature Energy Influence Compensation for MEMS Vibration Gyroscope Based on RBF NN-GA-KF Method. *Shock Vib.* 2018. [CrossRef]

6. Zhang, C.S.; Elaksher, A. An Unmanned Aerial Vehicle-Based Imaging System for 3D Measurement of Unpaved Road Surface Distresses. *Comput. Aided Civ. Infrastruct. Eng.* 2012, 27, 118–129. [CrossRef]

7. Shen, C.; Song, R.; Li, J.; Zhang, X.; Tang, J.; Shi, Y.; Liu, J.; Cao, H. Temperature drift modeling of MEMS gyroscope based on genetic-Elman neural network. *Mech. Syst. Signal Process.* 2016, 72, 897–905.

8. Nemra, A.; Aouf, N. Robust INS/GPS Sensor Fusion for UAV Localization Using SDRE Nonlinear Filtering. *IEEE Sens. J.* 2010, 10, 789–798. [CrossRef]

9. Seo, S.-H.; Lee, B.-H.; Im, S.-H.; Jee, G.-I. Effect of Spoofing on Unmanned Aerial Vehicle using Counterfeited GPS Signal. *J. Position. Navig. Timing* 2015, 4, 57–65. [CrossRef]

10. Shen, C.; Liu, X.; Cao, H.; Zhou, Y.; Liu, J.; Tang, J.; Guo, X.; Huang, H.; Chen, X. Brain-like Navigation Scheme based on MEMS-INS and Place Recognition. *Appl. Sci.* 2019, 9, 1708. [CrossRef]

11. Noureldin, A.; El-Shafie, A.; Bayoumi, M. GPS/INS integration utilizing dynamic neural networks for vehicular navigation. *Inf. Fusion* 2011, 12, 48–57. [CrossRef]

12. Guo, X.; Tang, J.; Li, J.; Shen, C.; Liu, J. Attitude Measurement based on Imaging Ray Tracking Model and Orthographic Projection with Iteration Algorithm. *ISA Trans.* 2019. [CrossRef] [PubMed]

13. Bhatt, D.; Aggarwal, P.; Devabhaktuni, V.; Bhattacharya, P. A novel hybrid fusion algorithm to bridge the period of GPS outages using low-cost INS. *Expert Syst. Appl.* 2014, 41, 2166–2173. [CrossRef]

14. Lu, Y.; Xue, Z.; Xia, G.; Zhang, L. A survey on vision-based UAV navigation. *Geo-Spat. Inf. Sci.* 2018, 21, 21–32. [CrossRef]

15. Cao, H.; Zhang, Y.; Han, Z.; Shao, X.; Gao, J.; Huang, K.; Shi, Y.; Tang, J.; Shen, C.; Liu, J. Pole-Zero-Temperature Compensation Circuit Design and Experiment for Dual-mass MEMS Gyroscope Bandwidth Expansion. *IEEE/ASME Trans. Mechatron.* 2019, 24, 677–688. [CrossRef]

16. Guo, H.; Zhu, Q.; Tang, J.; Nian, F.; Liu, W.; Zhao, R.; Du, F.; Yang, B.; Liu, J. A temperature and humidity synchronization detection method based on microwave coupled-resonator. *Sens. Actuators B* 2019, 281, 434–440. [CrossRef]

17. Gibson, J.J. *The Perception of the Visual World*, 1st ed.; Houghton Mifflin Company: Boston, MA, USA, 1950.

18. Poggio, T.; Reiehardt, W. Visual control orientation behavior in the fly: Part II. Towards underlying neural interactions. *Q. Rev. Biophys.* 1976, 9, 377–438. [CrossRef] [PubMed]

19. Horn, B.; Schunck, B. Determining optical flow. *Artif. Intell.* 1981, 17, 185–203. [CrossRef]

20. Tu, Z.; Xie, W.; Zhang, D.; Poppe, R.; Veltkamp, R.C.; Li, B.; Yuan, J. A survey of variational and CNN-based optical flow techniques. *Signal Process. Image Commun.* 2019, 72, 9–24. [CrossRef]

21. Baker, S.; Scharstein, D.; Lewis, J.P.; Roth, S.; Black, M.J.; Szeliski, R. A Database and Evaluation Methodology for Optical Flow. *Int. J. Comput. Vis.* 2011, 92, 1–31. [CrossRef]

22. Niu, Y.; Xu, Z.W.; Che, X.J. Dynamically Removing False Features in Pyramidal Lucas-Kanade Registration. *IEEE Trans. Image Process.* 2014, 23, 3535–3544. [CrossRef] [PubMed]

23. Herisse, B.; Hamel, T.; Mahony, R.; Russotto, F.-X. Landing a VTOL Unmanned Aerial Vehicle on a Moving Platform Using Optical Flow. *IEEE Trans. Robot.* 2012, 28, 77–89. [CrossRef]

24. Wang, Z.; He, G.; Du, W.; Zhou, J.; Han, X.; Wang, J.; He, H.; Guo, X.; Wang, J.; Kou, Y. Application of Parameter Optimized Variational Mode Decomposition Method in Fault Diagnosis of Gearbox. *IEEE Access* 2019, 7, 44871–44882. [CrossRef]

25. Brox, T.; Malik, J. Large Displacement Optical Flow: Descriptor Matching in Variational Motion Estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* 2011, 33, 500–513. [PubMed]

26. Sun, D.; Roth, S.; Black, M.J. A Quantitative Analysis of Current Practices in Optical Flow Estimation and the Principles behind Them. *Int. J. Comput. Vis.* 2014, 106, 115–137. [CrossRef]

27. Forster, C.; Pizzoli, M.; Scaramuzza, D. SVO: Fast semi-direct monocular visual odometry. In *Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA)*, Hong Kong, China, 31 May–7 June 2014; pp. 15–22.
28. Peng, Y.A.; Chen, Z.X.; Wu, Q.M.J.; Liu, C.Y. Traffic flow detection and statistics via improved optical flow and connected region analysis. *Signal Image Video Process*. 2018, 12, 99–105. [CrossRef]
29. Varol, G.; Laptev, I.; Schmid, C. Long-Term Temporal Convolutions for Action Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 2018, 40, 1510–1517. [CrossRef] [PubMed]
30. He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007.
31. Wang, Z.; Du, W.; Wang, J. Research and Application of Improved Adaptive MOMEDA Fault Diagnosis Method. *Measurement* 2019, 140, 63–75. [CrossRef]
32. Chung, K.-L.; Lin, K.-S. An efficient line symmetry-based K-means algorithm. *Pattern Recognit. Lett.* 2006, 27, 765–772. [CrossRef]
33. Ren, S.Q.; He, K.M.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 2017, 39, 1137–1149. [CrossRef]
34. Lu, S.; Ren, C.; Zhang, J.; Zhai, Q.; Liu, W. A Novel Approach to Droplet’s 3D Shape Recovery Based on Mask R-CNN and Improved Lambert(-)Phong Model. *Micromachines (Basel)* 2018, 9, 462. [CrossRef] [PubMed]
35. Wang, Z.; Zhou, J.; Wang, J.; Du, W.; Wang, J.; Han, X.; He, G. A novel Fault Diagnosis Method of Gearbox Based on Maximum Kurtosis Spectral Entropy Deconvolution. *IEEE Access* 2019, 7, 29520–29532. [CrossRef]
36. Shen, C.; Zhang, Y.; Tang, J.; Cao, H.; Liu, J. Dual-optimization for a MEMS-INS/GPS System during GPS Outages Based on the Cubature Kalman Filter and Neural Networks. *Mech. Syst. Signal Process.* In press.
37. Liu, W. Video Face Detection Based on Deep Learning. *Wirel. Pers. Commun.* 2018, 102, 2853–2868. [CrossRef]
38. Wang, Z.; Zheng, L.; Wenhua, D. A novel method for intelligent fault diagnosis of bearing based on capsule neural network. *Complexity* 2019. [CrossRef]
39. Von Luxburg, U. A tutorial on spectral clustering. *Stat. Comput.* 2007, 17, 395–416. [CrossRef]
40. Wang, Z.; Wang, J.; Cai, W.; Zhou, J.; Du, W.; Wang, J.; He, G.; He, H. Application of an Improved Ensemble Local Mean Decomposition Method for Gearbox Composite Fault diagnosis. *Complexity* 2019. [CrossRef]
41. Zhang, Z.Y. A new camera calibration technique based on circular points. *IEEE Trans. Pattern Anal. Mach. Intell.* 2000, 22, 1330–1336. [CrossRef]