Method of fragment based tracking of displacement of a large areal object in images

V V Lopatina
Admiral Ushakov Maritime State University, 93, Lenin Ave., Novorossiysk, 353924, Russia
E-mail: int00h@mail.ru

Abstract. Automation of technological processes using computer vision systems necessitates fragment based tracking of the displacement of a large areal object in images. Examples of such processes are monitoring mooring operations, loading and unloading operations, analyzing the environment from the vessel's bridge, and tracking berthing. These processes are cyclical in the maritime transport industry and thus require methods for automatic data analysis in 24×7 mode. The paper proposes a method of fragment based tracking of the displacement of a large areal object in images in real time. The algorithm that underlies the method enables tracking the object fragment using convolution of pixel matrices and ensures data processing by frequency methods, including computational operations using Fourier images. The developed method is applicable for real-time tracking the object in images, even when it can be hardly distinguished from the background. The method of tracking the displacement of a large areal object proposed in the study can be used in software and hardware meters for longitudinal and vertical displacement, which are used in the maritime transport industry.

1. Introduction
Automation of technological processes using computer vision systems requires fragment based tracking of the displacement of a large areal object in images. Examples of such processes are monitoring mooring operations, loading and unloading a vessel, analyzing the environment from the vessel, monitoring berthing. These processes are cyclical in the maritime vehicle industry, which means they require methods for automatic data analysis in 24×7 mode.

Specific problems that require fragment based object tracking using computer vision methods during automation of processes in the seaport are as follows: tracking the vessel orientation near the berth during mooring and berthing, monitoring loading, and analyzing the environment from the vessel's bridge. These tasks are reduced to fragment based tracking of the displacement of a large areal object in images.

Let us consider the problems of high-precision positioning and steering of large-tonnage vessels, such as tracking the displacement of a vessel relative to objects and positioning the hull of a large-tonnage vessel relative to the berth using computer vision methods.

When solving these problems, it is necessary to perform accurate and contactless measurement of the longitudinal and vertical displacement of a moving object with regard to the peculiarities of its size, shape, speed of maneuvering, and speed of movement. For example, high-precision positioning of the vessel's
hull relative to the berth during loading/unloading and mooring operations. This problem should be solved regardless of the illuminance (day, night), weather conditions (rain, fog), changes in the relative position of objects (partial or complete overlap of the tracked object). When solving this problem, the object must be continuously monitored at all stages of measurements (approach of the vessel, mooring and departure from the berth). The problem is solved by computer vision methods.

Measurements begin when the vessel approaches the berth (maximum distance is 500 m), and continues during the vessel mooring and berthing (minimum distance is 2 m), and when the vessel leaves the berth. The speed of the vessel at all stages does not exceed 3 knots.

To obtain sharp images, a laser-optical meter is used at all stages of measurements [1]. It includes an array of digital video recording devices with variable focal length lenses and a microcomputer for primary processing of the data measured.

To eliminate ambiguity in the object choice and simplify recognizing the object of tracking, only part of the vessel's side is captured (Figure 1). This means that the input data array is a series of images of fragments of a large areal object, that is part of a vessel's side characterized by a small number of clearly distinguishable details.

![Figure 1](image)

**Figure 1.** Examples of images of fragments of the vessels' sides at a distance of 2–500 m: (a), (c) day, (b), (d) night.

All clearly distinguishable details are taken as objects of tracking: inscriptions, identification marks, welds, corrosion defects and differences in the color of the vessel's hull. In this case, features of the shape of the vessel's hull should be taken into account: uneven surface, protruding parts, curvature, etc. Shooting is performed with a frequency of 120 frames per second.

2. Method
Let us consider the problem of fragment-based tracking of the displacement of a large areal object in images in the following formulation.

There is a sequence of video frames of size \( w \times h \). Select the \( m \times n \) rectangular area \( W \) that contains the object of tracking.

To reduce the changes in brightness, normalize the range \([-1; 1]\).

Based on the selected rectangular area \( W \), create a training set of images \( \{W_1, W_2, ..., W_N\} \). To do this, apply affine transformations to the image (random image rotation in the range from -30 to 30 degrees). Increase the brightness of the dark pixels. To remove a sharp transition at the boundaries of \( W \), reduce the values closer to the boundaries to zero using the Henning window.

To calculate the first filter \( F_0 \), perform preprocessing.

For the resulting set of image \( N \), create a filter \( F_{i+1} \) to give a response for each image from the created set. To do this, calculate the sum of squares (matrix elements) of the pixel-by-pixel difference between the
window and the corresponding site of the object's location in the n-th image created from the frame with a true position of the object.

For example, if we take two adjacent video frames I₀ and I₁, frame I₀ is the frame where the rectangular area W₀ is selected, cut out rectangle W₁ of the same size in the same coordinates on frame I₁, implying that the object has shifted slightly. Consider the square of the sum of the difference between two rectangles:

\[ D(i, j) = \|W_i(i, j) - W_0\|_2 \]

where \( i \in [0, m] \), \( j \in [0, n] \).

Superpose W₀ with the object (i, j) in W₁ and bring the missing values to zero.

Then, find the minimum of the matrix D, its coordinates minus the coordinates of center W₁ will be the coordinates of the object on frame I₁.

For I₂, the coordinates of window W₂ are taken from I₁.

In center W at (x, y), the filter gives a response of 0. The response is centrally symmetric relative to (x, y), i.e. beyond the center point, the response will be greater than zero. Filter F ideally gives the result:

\[ G_N = 1 - \exp \left( \frac{(x - i)^2 + (y - j)^2}{\sigma^2} \right) \]

Minimize the response deviation from the desired one.

\[ D_n(i, j) = \|W^n(i, j) - F\|_2 \]

\[ F = \min_{F} \|D_n(i, j) - G(i, j)\|_2 \]

where \( n \in [1, N] \).

To implement the tracking algorithm using the convolution of matrices, multiply the entire window with the selected object by the corresponding part of the frame for each pixel of this frame. To simplify the algorithm, go to the frequency domain. Convolution of functions in the ordinary space (space that unites approaches based on direct manipulation of image pixels) is multiplication of their Fourier images. Let us apply the Fast Fourier Transform to the frames and perform the elementwise multiplication of their frequencies, and then transfer the result back into the matrix. The complexity of the algorithm will be \( O(mn \times \log mn) \), which is faster than the convolution of matrices \( O(m^2n^2) \).

To localize the object in the image, apply cross-correlation. Cross-correlation is a convolution with horizontally and vertically reflected F. Multiply the frequencies W by the complex conjugate matrix to the frequency matrix F.

\[ \tilde{F}(w, v) = \overline{F}(F(x, y)) \]

where \( \tilde{F}(w, v) \) is filter frequency matrix, \( \overline{F} \) is operator of two-dimensional discrete Fourier transform.

\[ \tilde{W}(w, v) = \overline{F}(W(x, y)) \]

where \( \tilde{W}(w, v) \) is frequency matrix of the fast Fourier transform applied to the image W.

\[ \hat{G}_{\text{conv}}(w, v) = \overline{F}(G_{\text{conv}}(x, y)) \]
where $\hat{G}_{\text{conv}}(w, v)$ is frequency matrix of the desired response.

$$G_{\text{conv}} = W \otimes F \rightarrow \hat{G}_{\text{conv}} = \hat{W} \otimes \hat{F}^i$$

where $W \otimes F$ is convolution of matrices, $\hat{W} \otimes \hat{F}^i$ is element wise matrix multiplication. Perfect response function:

$$G_{\text{conv}} = \exp \left( -\frac{(x-i)^2 + (y-j)^2}{\sigma^2} \right)$$

Since the convolutional metric is maximized, the response at $(x, y)$ will be greater than 0.

Filter frequency matrix for an image:

$$\hat{F}^i = \frac{\hat{G}_{\text{conv}}}{\hat{W}}$$

where $\frac{\hat{G}_{\text{conv}}}{\hat{W}}$ is elementwise division. Since the algorithm uses a set of images, the frequency matrix will be calculated for each image. After that, the resulting set of filters can be averaged as in [2]:

$$\bar{F}^i = \frac{1}{N} \sum_{i=1}^{N} \hat{F}^i = \frac{1}{N} \sum_{i=1}^{N} \frac{\hat{G}^i}{\hat{W}^i}$$

Take the filter frequencies that satisfy all images from the set on average:

$$\bar{F}^i = \min_{\bar{F}^i} \sum_{N=1}^{N} \left\| \hat{W}^i \otimes \bar{F}^i - \hat{G}^i \right\|_2$$

Find the minimum by taking the derivative with respect to the filter elements:

$$\frac{\sigma}{\sigma \bar{F}^i} \sum_{N=1}^{N} \left\| \hat{W}^i \otimes \bar{F}^i - \hat{G}^i \right\|_2$$

Another option is to take a filter that will give the maximum response to the entire set of images [3]. In this case, the filter for an image can be represented as:

$$\bar{F}^i = \frac{\hat{G}^i}{\hat{W}^i} = \frac{\hat{G}^i \otimes \bar{F}^i}{\hat{W}^i \otimes \hat{F}^i}$$

For a set of images:

$$\bar{F}^i = \sum_{N=1}^{N} \frac{\hat{G}^i \otimes \bar{F}^i}{\hat{W}^i \otimes \hat{F}^i}$$

Take the median filter as filter $\bar{F}^i$ to ensure the maximum response in the median image of the set.
After calculating $\tilde{F}^*$, calculate the response in the frequency domain:

$$\tilde{g}_{\text{conv}} = \tilde{W} \odot \tilde{F}^*$$

After that, transfer to the spatial domain and search for the maximum in matrix $G_{\text{conv}}$. Its coordinates will be the coordinates of the object $(i, j)$.

Let us denote the proposed algorithm by analogy with the well-established terminology as Median of Synthetic Exact Filters (MSEF).

3. Discussion

The developed method can be used for real-time tracking of objects in images, even when objects are hardly distinguished from the background. The tests were carried out using a laser-optical meter (2.7GHz core), the algorithm is written in Python using OpenCV. The maximum size of the tracking window is 64 × 64 pixels.

The algorithm results were compared with those of the following methods: On-line boosting [4], Multiple Instance Learning [5], Kernelized Correlation Filters [6], Median Flow [7], TLD (tracking, learning, detection) [8], GOTURN (Generic Object Tracking Using Regression Networks) [9], CSRT (Discriminative Correlation Filter with Channel and Spatial Reliability) [10].

| Algorithm                          | Number of frames processed at an initial rate of 120 frames per second | Average number of object losses (including jumps in object sharp coordinates) | Displacement of the area $W$ from the initial position (in pixels) |
|------------------------------------|------------------------------------------------------------------------|--------------------------------------------------------------------------------|-------------------------------------------------------------------|
| MSEF                               | 102 – 114                                                              | 6                                                                              | 3–7                                                               |
| MOSSE                              | 97 – 110                                                               | 4                                                                              | 16 – 20                                                           |
| ASEF                               | 85 – 101                                                               | 7                                                                              | 15 – 27                                                           |
| On-line boosting                   | 76 – 103                                                               | 9                                                                              | 13 – 24                                                           |
| Multiple Instance Learning         | 96 – 101                                                               | 12                                                                             | 27 – 34                                                           |
| Kernelized Correlation Filters     | 78 – 86                                                                | 22                                                                             | 20 – 29                                                           |
| Median Flow                        | 104 – 109                                                              | 17                                                                             | 17 – 30                                                           |
| TLD                                | 98 – 103                                                               | 19                                                                             | 12 – 28                                                           |
| GOTURN                             | 56 – 89                                                                | 34                                                                             | 13 – 35                                                           |
| CSRT                               | 34 – 46                                                                | 47                                                                             | 10 – 15                                                           |

The most significant parameter for measuring the longitudinal and vertical displacement of a moving object is the displacement of area $W$ (in red) from its initial coordinates (Figure 2). The erroneously determined object displacement (when $W$ is displaced from the initial position) indicates the error of the algorithm when determining the longitudinal and vertical displacement of the object. The error of the algorithm with an incorrectly selected filter implemented in a laser-optical meter can lead to a measurement error from several centimeters to several tens of meters, depending on the object's range.
The displacement in the developed algorithm is 3–7 pixels, which enables implementation of the algorithm as part of a software and hardware meter.

Figure 2. Examples of images of the tracked fragment of a vessel's side: (a) image of a vessel's fragment at a distance of 6 meters from the gauge, (b) object of tracking, (c) example of an erroneously detected object displacement.

4. Conclusion
The described method of tracking the object can be employed to create high-precision positioning systems and high-precision steering systems for various types of transport (sea, river, road, air and rail). Among other things, these systems solve the problems of stabilization of moving objects in various transport systems and their positioning.

The algorithm proposed in the study is used to solve the problem of positioning the hull of a large-tonnage vessel relative to the berth as part of the Monitoring System for Mooring Operations [11, 12].

References
[1] Lopatina V V, Senchenko V G 2020 Development of requirements for the optical system of a laser-optical measuring device of geometric displacements of a moving object Operation of Maritime Transport 3(96) 100–106
[2] Bolme D, Draper B, Beveridge J 2009 Average of Synthetic Exact Filters 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops) pp. 2105–2112 (Miami, FL)
[3] Bolme D, Draper B, Beveridge J, LuiY 2010 Visual object tracking using adaptive correlation filters 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 2544–2550 (San Francisco, CA)
[4] Grabner H, Bischof H 2006 On-line boosting and vision Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 1 260–267
[5] Babenko B, Belongie S 2009 Visual tracking with online Multiple Instance Learning 2009 IEEE Conference on Computer Vision and Pattern Recognition pp. 983–990
[6] Henriques J F, Caseiro R, Martins P, Batista J 2015 High-speed tracking with kernelized correlation filters IEEE Transactions on Pattern Analysis and Machine Intelligence 37(3) 583–596
[7] Kalal Z, Mikolajczyk K, Matas J 2010 Forward-backward error: Automatic detection of tracking failures Proceedings of the International Conference on Pattern Recognition pp. 2756–2759
[8] Kalal Z, Mikolajczyk K, Matas J 2012 Tracking-Learning-Detection IEEE trans-actions on pattern analysis and machine intelligence 34(7) 1409–1422
[9] Held D, Thrun S, Savarese S 2016 Learning to track at 100 FPS with deep regression networks Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 9905 749–765

[10] Lukezic A, Vojir T, CehovinZajcL 2018 Discriminative Correlation Filter Tracker with Channel and Spatial Reliability Int J Comput Vis 126 671–688

[11] Senchenko V, Lopatina V, Studenikin D, Butsanets A 2021 Technical Automation Tools for High-Precision Navigating of Sea and River Ships Proceedings of the XIII International Scientific Conference on Architecture and Construction 2020. Lecture Notes in Civil Engineering vol. 130 pp. 157–163 (Springer, Singapore)

[12] Popov A N, Kondratiev A I, Smirnov I O 2018 Algorithm for rapid prediction of the degree of collision hazard with ships and surface objects in the electronic navigation zone AGA 2018-19th Annual General Assembly (AGA) of the International Association of Maritime Universities (IAMU), pp. 442–449