Automatic assistance system for visual control of targets

V V Tulchinskii, V I Tulchinskii, S I Kondratyev, A L Boran-Keshishian
Admiral Ushakov Maritime State University, 93, Lenin Ave., Novorossiysk, 353924, Russian Federation
E-mail: tulchinskiyvadim@scf-groupe.ru

Abstract. The article discusses the possibility of creating a system for automating the process of searching and assessing the danger of targets using machine vision and existing ship systems. The technologies for processing images from the ship's camera system are analyzed: object recognition by a ready-made convolutional neural retrained network, as well as object detection by SURF algorithms. The parameters of the danger of goals within the framework of the problem to be solved have been analyzed and developed, and a structural analysis has been elaborated. The functionality of the automatic operation of the system as a closed circuit has been considered. To achieve the technical result in the system as an integrated part of the ship's equipment based on cognitive information processing, a composite information object from the ship's sensors and control systems was introduced. It is decomposed into blocks in which the associated logical information processing takes place using basic modules consisting of adaptive cells-approximators as learning elements capable of independently processing information. The following units are included in the local area network: an information support unit (ISU), a decision making unit (DMU), a command-organizational unit (COU) with an operator's workstation. The ISU is connected by a two-way communication output with aggregate LAN units, the other output is connected in series with a situation determination unit and the "model of the surrounding space" block. The survey was carried out on board a modern LNG gas carrier using ship systems.

1. Introduction
Contemporary navigation is a complex technical process, which, despite the abundance of modern technical means is, one way or another, based on some basic principles. Considering the neat passing of vessels – which is actually the main factor in the safe movement of the vessel along the passage route, it is possible to simplify the task, reducing it to two successive elements: the definition of targets and passing at a safe distance. In reality, each of these elements has few tools in its arsenal (Figure 1). Determination of targets can be visual, using radar (X or S bands) or using AIS (VHF data reception range). The parameters of the passing in quantitative terms can be obtained using ARPA or calculation from AIS data on ECDIS. Also, the parameters of passing can be obtained with the help of visual control of targets, using the general principles of visual control of targets.

Despite the contemporary technical means of navigation, visual identification of targets and subsequent visual target control (target angle control) today, as well as 100 years ago is, in many navigational situations, the most reliable tool in assessing the danger of a target. The use of visual observation to assess the situation is not just an additional tool, but still remains in fact the main one, which is enshrined in a number of both international guidelines and company procedures, and recommendations to navigators.
Despite this, the modern industry does not stand still. Within the framework of the global automation of complex technical processes, navigation is also undergoing changes. In the near future this is the prospect of introducing the concept of “One man on the Bridge”, in a more distant future - automated pilotage, navigation with small crew or without it. Already today, the SMS of various companies and international recommendations allow the use of Watch Level number 1 on the high seas in the daytime - only mate-on watch on the navigating bridge. Visually defining goals and continuously monitoring them requires the full attention of at least one person. At the same time, the navigator needs to perform a number of other tasks - ship control, control over the location relative to depths and navigational hazards. Also, part of the attention is paid to turning to ARPA for calculating the parameters of maneuvers, to ECDIS, to ship communication systems between mobile stations, and so on. In areas of intensive shipping or when targets suddenly appear on the high seas, situations sometimes arise when the lack of "eyes" is not a fictional problem. Of course, you can call an additional observer, but this, firstly, takes time, and secondly, this approach to the problem does not fit into the general concept of small crew navigation.

Machine vision is a whole area of science and technology that has already become a key in countless automated processes. Its use in promising unmanned and small-crew navigation is absolutely obvious. The only issue is the time and methods, approaches to solving the tasks and in search of the best ways to implement algorithms, build and integrate ship machine vision systems.

Despite the general complexity of the complex solution of the problems of passing of vessels using machine vision, the author believes that today there is a real possibility of using machine vision by using existing ship technical means to help the navigator to automatically determine dangerous targets. Such a new system does not require significant installation costs, since it is based on existing devices. The system does not make decisions, but only provides additional information for subsequent analysis and decision-making by the navigator. Figure 2 shows a diagram that includes the use of machine vision elements.

Figure 1. Contemporary search and passing of ships
In this article, the authors analyze the possibility of creating a ship's automatic assistance system for visual control of targets. The article considers:

- Technical prerequisites for the creation of the system
- Methods and approaches to solving the problem of identification of targets
- Simplified mathematical processing of target position parameters, characteristics of a dangerous target.
- General algorithm of the system's operation
- Application of associative-logical processing of concentrated information.

2. Technical prerequisites for the creation of the system
The technical base of the research object is a modern LNG gas carrier of automation level 2G, displacement 127760 mt, maximum length 299.06 m, width 46.4 m, draft 12.5 m. Maximum operating speed 19.48 knots. Date of construction 01/15/2021.

The following ship systems were used for the study:
- Marine radar JMR JR-7230-S3 / S [1];
- Onboard CCTV system (elements: server, display screen on the navigating bridge, digital controlled camera).

Characteristics of the digital camera CCTV system, located on the vessel's forecastle in the DP, on which various scenes were recorded for subsequent processing and analysis [2]:
- Model: MCT-300-FHW
- Focal length length: 4.44 ~ 142.6mm (Optical 32x):
- Approximation: 16 times
- Horizontal viewing angle: 62.8° (Wide) ~ 2.23° (Tele)
- Vertical viewing angle: 36.8° (Wide) ~ 1.26° (Tele)
- Configuration of CCTV system from “optical meter” (digital camera) to signal output means (control and display screen) is shown in Figure 3.

![Figure 2. Prospective search and passing of vessels](image-url)
Figure 3. The structure of the vessel’s CCTV system

The location of the "optical meter" (camera) is shown below in Figure 4.

Figure 4. The location of the optical meter on the vessel

All calculations and processing of fragments of the recorded video stream were carried out in the MatLab software environment. Various configurations of preset algorithms were used, as described below. Fragments of the video stream were recorded on 03/14/2021 at different times of the day during the approach and passage of the Singapore and Malacca Strait.

3. Methods and approaches to solving the problem of detection of targets

1) Detection of targets in the daytime. There are images of objects on the stage - various floating objects, large ships, fishing vessels, barges, etc. There is a large palette of colors in RGB space. The sizes of objects are easily distinguishable, it is possible to distinguish characteristics, to classify. The horizon is clearly visible. When coastal objects enter the frame, it is likely that additional algorithms are needed.

2) Detection of targets at night. Target lights are present on the stage. The most distinguishable are bright lights of fishing vessels, towers. Determining the images or sizes of objects is difficult (when using a conventional type of digital camera). Each object is pointwise clearly distinguishable. The scene is in a gray color space, the horizon is hardly distinguishable.

3.1 Daytime detection using a convolutional neural network.

Ready-made and pre-trained neural networks are often used in advanced machine vision tasks for object classification and search. However, there are no pre-trained networks that specifically meet the parameters of the problem. However, one of the most popular networks, AlexNet, has in its structure images of vessels and water bodies (from the ImageNet training dataset). Having configured the network in MatLab software environment, we will evaluate the accuracy of determining the objects. The area of interest for each image is pre-selected - above the ship's forecastle and below the horizon. Of the 30 images selected for analysis, 20% of the objects were detected correctly, 15% - with an error in the object type, and the rest of the data gave an incorrect definition. The general task is reduced not only to the recognition of the type of object, but to its general search on the scene object for the subsequent obtaining of pixel coordinates and further calculations. Therefore, we cannot be sure that, despite the incorrect classification of the object, the network correctly found the object of interest to us on the scene. Therefore, this option is not suitable for solving the task at hand.
Figure 5. An example of the received data. a - correct identification, b - partially correct one, d – incorrect one.

3.2 Daytime detection using a modified convolutional neural network pre-trained on a different dataset.

We can fine tune deeper layers in the network by training the network on a new dataset with a pretrained network as a starting point. Fine-tuning a network with transfer learning is often faster and easier than building and training a new network. The network has already learned a rich set of imaging features, but by fine-tuning the network, it can learn the features specific to our new dataset. (Figure 6). [3]

Network fine tuning is slower and requires more power than simple detection by the finished network, but since the network can learn to extract a different set of features, the resulting network can be more accurate. However, the new dataset should be moderately large. With a very large dataset, training may not be faster than training from scratch. Let’s retrain the AlexNet network on the CIFAR-10 dataset. The CIFAR-10 dataset consists of 60,000 32x32 color images in 10 classes, 6,000 images per class. There are 50,000 training images and 10,000 test images. The dataset is split into five training packages and one test package, each containing 10,000 images. The test suite contains exactly 1000 randomly selected images from each class. The training packages contain the remaining images in random order, but some training packages may contain more images from one class than another. In between, the training packages contain exactly 5000 images from each class. One of these classes in this dataset is vessel (ship, small craft).

Of the 30 images selected for analysis, 55% of the objects were identified correctly, and the rest of the objects were not detected.
3.3 Detection of an object (without classification) at night and during the day.

There are various descriptor algorithms that most quickly and accurately allow you to highlight the special points. Any object can be distinguished on the scene object from the horizon to the line of the vessel’s forecastle, due to the morphological homogeneity of the water surface image. It is particularly easier to do this in a space of gray shades.

It is proposed to use the so-called SURF method for detecting objects. Before these daytime images are necessarily translated into a space of gray tones.

The SURF method searches for feature points using the Hessian (the Hessian matrix). In general:

\[
H = \begin{bmatrix}
\frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\
\frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2}
\end{bmatrix}
\]  

(1)

where \( H \) is the Hessian matrix, \( f(x,y) \) - the function describing the change of the brightness gradient.

- For certain points, the direction of a greater change in brightness is calculated
- The scales of the Hessian matrix are sorted out by rank
- The descriptors formed are determined.

For the SURF method, the descriptor, as a rule, consists of 64 numbers of the Hessian matrix and has invariance with respect to image rotation. However, it is not scale-invariant (at multiple sizes). Therefore, the SURF method uses filters of various scales to calculate the Hessians. Then the gradient and scale are calculated for each feature point. The gradient at a point is calculated using Haar filters. The filter dimension is taken equal to 4s, where s is the point scale. After the feature points have been found, the SURF method calculates their descriptors. The descriptor is a set of 64 numbers for every 7th keypoint. Each number represents the difference in the gradient around the feature point. Each feature point represents the maximum of the Hessian, which guarantees the fact that in the vicinity of this point there will be areas with different gradients. Thus, the difference in descriptors for different feature points is ensured, as a result of which the invariance of the descriptor with respect to rotation is formed. The size of the area on which the descriptor is calculated is determined by the scale of the Hessian matrix, which ensures scale invariance. [4]

The quality of the definition depends on the "coarsening" parameter - the number of points to be selected. With different values of this parameter, more than 65% of the selected descriptors were located on real objects in all 30 studied images.

Figure 7. Daytime. Coarsening parameter value up to 100 points

Figure 8. Daytime. Coarsening parameter value up to 5 points.
Figure 9. Nighttime. Coarsening parameter value up to 5 points.

An empirically calculated graph of the dependence of the number of targets detected in the daytime and the coarsening parameter for their determination in the range of up to 65% is shown in Figure 10. At night, the value of the coarsening parameter is directly proportional to the approximate number of targets on the scene.

Figure 10. Coarsening parameter

Comparing the obtained results, it is obvious that simple SURF detection is more suitable for the purposes of the experimental system. Another important advantage is that this method does not require significant computing power.

4. Simplified mathematical processing of target position parameters, characteristics of a dangerous target

Having received and calculated the averaged coordinates of the object, it is possible to further track the parameters of its movement relative to the conventionally stationary area of the camera's capture. Let us assume that the averaged coordinates of the detected object are returned by the algorithm as \((x, y)_n\), where \(n\) iteration of each subsequent frame of the video stream is the pixel coordinates of the object in the frame coordinate system (vertically and horizontally).
Figure 11. Calculating target displacement

Considering that with small displacements in the area of scene objects (relative distances depend on the time interval between the analyzed frames), the values of these displacements are much less than the distances to objects, and the angle of position change is also not large, that is $D >> K \rightarrow D >> X$, since $K \approx X$. Then we have that $\beta = \frac{x}{2\pi D}$, whence:

$$\beta = \frac{180X}{\pi D}$$

(2)

in angular dimension.

The change in relative bearing over time and the displacement characteristic can be estimated as:

$$\beta' = \frac{d\beta}{dt} = \frac{d(\frac{180X}{\pi D})}{dt}$$

(3)

Thus:

$$\frac{d(\frac{180X}{\pi D})}{dt} > 0 \rightarrow \text{target bearing increases}$$

$$\frac{d(\frac{180X}{\pi D})}{dt} < 0 \rightarrow \text{target bearing decreases}$$

The greatest danger of the target can be assessed as:

$$\left\{ \begin{array}{l}
\frac{d\beta}{dt} < 0, \\
\frac{d(\frac{180X}{\pi D})}{dt} = 0, \\
D_i < G,
\end{array} \right.$$  

(4)

Where $G$ is the distance seen by the camera to the horizon line.

A target can be rated as conditionally safe if:

$$\left\{ \begin{array}{l}
\frac{d\beta}{dt} < 0, \\
\frac{d(\frac{180X}{\pi D})}{dt} \neq 0, \\
D_i \equiv G
\end{array} \right.$$  

(5)

The above reasoning is based on the "real" location of objects in a perspective projection with metric values. To translate the data from images into a perspective projection requires a number of complex calculations and use of the dataset to calibrate the camera. Within the framework of the problem under consideration, these calculations are absent and an approximation about the correspondence of the change in pixel values in the image to real metric values is accepted.
An example of determining the displacement parameters of 2 targets is shown below in Figure 12, where a is the first frame with 2 described targets; b - the second frame with the same described goals, the time between frames is chosen arbitrarily; c - imposition of frames determined by SURF algorithms and visualization of their relative movement on the scene. Target displacement values are announced in pixel coordinates. Getting rid of "confusion" of targets errors on successive frames should be ensured by reducing the time interval between frames. This interval is limited only by the camera parameters (frames per second, resolution).

![Figure 12. Determination of displacement parameters for two targets](image)

For the algorithm to work on each frame, it is necessary to determine:

- pixel distance from and to the horizon line \( G_n \). The line of the visible horizon can be found by one of the available algorithms for finding lines on a scene object, for example, Canny's edge detection algorithm, Hough transform, etc.
- pixel averaged coordinates of each detected object - target \((x_{ni}, y_{ni})\).
- the angle between the ship's DP and the direction to each of the targets on the frame \( \alpha_t \) is estimated based on the pixel coordinates of the target and the starting point of reference.

When the target is on the horizon, the algorithm provides for the possibility of using the distance to the target, which is periodically taken off the ship's radar. The radar distance is taken periodically, then the parameters of the target's movement along the axis \( y_{ni} \) are calculated automatically using data from the ship's heading and speed sensors. If \( D_{ni} \) is the distance to the target, taken when it was detected, then the distance to the target can be estimated as:

\[
\Delta_t = D_{(n+1)i} - D_{ni} - \int_{t_n}^{t_{(n+1)}} \cos(\alpha_t)v dt
\]

The body of the algorithm is logically divided into 3 parts:

1) analysis of the danger of the target when it is on the horizon and receiving periodic data on the distance to it from the ship's radar;
2) analysis of the danger of the target when it is on the horizon and there is no data on the distance to it from the ship's radar;
3) analysis of the danger of the target when it is located before the horizon line and the absence of data on the distance to it from the ship's radar;
When visually assessing the danger of a target, an important factor is also the dependence of the rate of change in the bearing or heading angle of the target on its position relative to the vessel. The closer the target is to traverse distances, the faster its bearing changes during a neat passing. Therefore, it is necessary to split the camera capture sector into smaller sectors. When detecting and tracking a target, it is necessary to check whether it belongs to a certain array. Depending on the sector of finding the target, a logical sector parameter $A_{si}, B_{si}, C_{si}$ is assigned for assessment (Figure 13).

Target displacement rate is assessed as:

1) $\beta_i' = \frac{\partial \beta_i}{\partial t}$ for the first variant of the algorithm (7)

2) $\beta_i' = \frac{\partial \beta_i}{\partial t} = \frac{d(180\pi X_i)e_{pixn}}{dt}$ for the second variant of the algorithm, where $G_{pixn} = G_nk$ (8)

3) $\beta_i' = \frac{\partial \beta_i}{\partial t} = \frac{d(180\pi X_i)e_{pixn}}{dt}$ for the first variant of the algorithm, where $G_{pixn} = Y_nk$ (9)

where, $k = f(Y_i, X_i, G_{pixn})$ – distortion coefficient, found experimentally.

The target approach speed is estimated as:

$$d_i = \frac{\partial Y_i}{\partial t}$$

The characteristic of the change in the displacement rate of the target for the 3rd part of the algorithm is estimated as [5]:

$$a_i = \frac{\partial \beta_i'}{\partial t}$$

![Figure 13. Target detection sectors and their logical parameters](image)

The danger analysis block for each part of the algorithm will accordingly be as follows:

1)  1. $\Delta_i > 0$
2. $\beta_i = 0$
3. $\beta_i' \in [0, C_N]$ (12)

The displacement indication block for each part of the algorithm will accordingly be as follows:

$$\beta_i'$$ target shifts to the right

$\beta_i'$ target shifts to the left (15)
5. General algorithm of the system’s operation

![Diagram of the system's operation]

**Figure 14.** Algorithm of the system’s operation
Figure 15. Panel for display and output of indication on the available monitor of the ship's CCTV

6. Consideration of the functionality of the automatic operation of the system as a closed circuit

In the analyzed system of automatic assistance for visual control of targets, the method for diagnosing the complex consists in using an automatic program mode: to check each functional unit of the complex, a closed circuit is created, the main elements of which are an optical meter (camera) and the unit being checked - a calculation and display unit. The structure of the closed circuit operation is designed in such a way as to get an answer to two questions: whether the tested unit is working or not working and what is its real error.

As a criterion for the prevalence of units, a functional is used, which takes into account the signs of the probability of a unit failure during the test $P_i$, the time of the unit check time $t_i$ and the conditional probability of communication of the $i$-th node with other devices [5]:

$$Q_i = 1 - \frac{n_i}{\sum_{i=1}^{k} n_i};$$  \hspace{1cm} (16)

Where $n_i$ is the number of links of a unit, calculated according to the graph of functional and logical links of the units of the complex, $k$ is the number of functional units.

The ranking functionality is defined as:

$$\Phi_i = Q_i \sum_{j=1}^{k} \gamma_{ij} q_{ij}, \quad i = 1, k;$$  \hspace{1cm} (17)

where $\gamma_{ij}$ - weight of the normalized probability of failure $P_i$ of the $i$-th unit,

$$q_{ij} = p_i / p_{max};$$  \hspace{1cm} (18)

$\gamma_{i2}$ - weight of normalized test time $t_i$ of the $i$-th unit;

$$q_{i2} = t_i / t_{max};$$  \hspace{1cm} (19)

The probability of failure during the test time $t_i$ is determined by the mean time of failure-free operation:

$$T_{0i} \cdot p_i = e^{-t_i/T_{0i}};$$ \hspace{1cm} (20)

7. Conclusion

To achieve the technical result in the system as an integrated part of the ship's equipment based on cognitive information processing, a composite information object from the ship's sensors and control systems was introduced. It is decomposed into blocks in which the associated logical information processing takes place using basic modules consisting of adaptive cells-approximators as learning elements capable of independently processing information.

Designation of functional blocks and units in the system:

1) CCTV
2) General block diagram of the control system
3) Case and system architecture
4) Radar
5) Block program of calculations
6) Block issuing indication (measurement result for the target)

Figure 16. Block diagram, as a useful model, of the integrated control of the system operation.

The cells are made of blocks: identification, assignment, calculation, analysis and indication, forming connections and weights of approximating functions, and having n-inputs and one output. There are also aggregate concentrator blocks, as a set of many inputs and outputs of weights, connections, hidden parameters of activators that convert signals in a cognitive system, in a local computer network, and local computer networks themselves, there can be several of them according to threshold positions. The local area network includes blocks: information support unit (ISU), decision making unit (DMU), command-organizational unit (COU) with operator's workstation, ISU is connected by an output with two-way communication with aggregate LAN units, the other output is connected in series with the situations determination unit and the "model of the surrounding space" block. There is a segment of an intellectual interface (II), an input learning unit (ILU) with a cognitive system monitor (CS) and an operator interface.

References
[1] JRC JMR-7230-S3 Instruction Manual, (JRC, Japan)
[2] Ship’s Drawing Hull 8006-8008. T-87 CCTV System. (HYUNDAI SAMHO HEAVY INDUSTRIES CO.,LTD.)
[3] MatLab 2018 Tutorial. Convolution Neural Networks
[4] MatLab 2018 Tutorial. SURF algorithm
[5] Potapov M K, Aleksandrov V V, Nasichenko P I 1980 Algebra and analysis of elementary functions (M.: Nauka)