[1]**M. P. Mukhina,**
[2]**I. V. Barkulova**

STRUCTURE OF AIDED CLASSIFICATION OF GROUND OBJECTS BY VIDEO OBSERVATION

[1,2]Educational & Research Institute of Information and Diagnostic Systems,
National Aviation University, Kyiv, Ukraine
E-mails: [1]marinamukhina79@gmail.com, [2]irochkaseden@gmail.com

Abstract—Analysis of classification structure by video observation has been done. It was formulated, that for feature extraction and their classification, normalized hypothesis for object feature detection, taking into account camera orientation and flight height, have being obtained. The system with aided classification based on probabilistic models, such as Bayesian classifier and Markov chain model, is proposed. The applied algorithm was used for detection by only two features related to Binary Large Objects (BLOB) analyses. Classification was done by two main feature parameters: area and center of mass. Feature vector contains the most informative components and allows the minimization of decision risks. Results have proven the reliability of classification during a number of video frames in the condition of non-full data descriptive space.

Index Terms—Feature vector; descriptive space; aided classification; Bayesian segmentation.

I. INTRODUCTION

Nowadays Unmanned Aerial Vehicle (UAV) can be used to solve many tasks that are not performed by aircraft because of various reasons. UAVs play an important role in mobile aerial monitoring operations and have been widely applied in diverse applications such as aerial surveillance, border patrol, resource exploration, and combat and military applications. Due to its mobility, UAV has also been deployed for search and rescue. One potential benefit of UAVs is that they could fill a gap in current border surveillance. In particular, technical capabilities of UAVs could improve coverage along remote sections of the borders. As most monitoring systems require detection and recognition of the object, object classification is a crucial process in UAV monitoring system.

The most works related to detection and classification of humans and vehicles by means of data obtained from UAV the motion is not the basic feature for object detection. On the one hand, it is good, because in some programs the static objects are also analyzed. On the contrary, it is difficult reliable approaches, especially, if you work with images, where the distance between pixel centers (GDB) is so great, and objects cover only few pixels. Object detection and surveillance for static camera is not a simple task. The series of problems such as illumination changes, shadows, disturbances, coverage one object by other, grows in real conditions.

Applying classification, it is needed to get two aims: firstly – find the difference between different classes of objects and secondly – support the fusion process of bounding-boxes, which are spatially connected. Because the detection of moving object is the reliable process and produces nearly no clutter detections, it is not necessary to consider clutter as a special class during classification process. So we are focused on two classes:

– moving objects (mostly ground vehicles like cars, trains, tanks);

– stationary objects (vehicles and landmark objects like buildings, nature objects, etc).

For further feature extraction and their classification, we have to obtain normalized hypothesis for object feature detection taking into account camera orientation and flight height.

II. OVERVIEW OF EXISTING CLASSIFICATION SYSTEMS

In article [1] the method of feature detection and their classification during moving object detection and consideration of classification system is introduced. The framework for moving object detection and system classification is proposed. Framework consists of modules such as feature detection, descriptor settings and classification.

Only very general features with little spatial information have been chosen due to high object distance and weak shape information, which made the system rejecting for example gradient-based features. Finally, the proposed system does not only identify an appropriate feature vector (descriptor), but also many features with weak separability.

As result, the moving target detection works reliably and stable, classification doesn't need to

separate between object and clutter but only between vehicles and people. The best achieved correct classification rate is of 96.08%. The moving target detection and classification approach is very fast and doesn't need any camera parameters or meta information.

Authors in [2] propose the algorithm of correlation comparing of black-white and color images of different dimensions and mostly solve the problem of template analyses instead of multi-hypothesis search.

In article [3] the important issues is discussed that needs to be resolved before fully automated outdoor surveillance systems can be developed. Interesting solutions to the problems of shadows, handling spurious objects; classification of objects based on recurrent motion and carried object detection are presented.

III. PROBLEM STATEMENT OF AIDED CLASSIFICATION

For aided classification of ground objects it is necessary to set the general model of video observation that includes the following components: mathematical model of camera; measurement models of basic navigation sensors of UAV; dynamic model of vehicle (UAV); coordinate systems in which both the target and UAV are estimated; prior data (cartographic data, templates, flight mission, etc).

Taking into account the specifics of video observation aboard UAV the following structure of aided classification is proposed (Fig. 1).

Operations on a stage of frame preprocessing are well known and good formalized. It is only to mention that the correction here means so called image rectification (transformation) to put the frame in accordance with georeferencing image database. This process mostly requires the information about flight height, camera orientation and world file of georeferencing image. Segmentation can be done as binary and color separation. Usually for aerial images it is done using Otsu method [4].

Among image features and their description there is a variety of methods. Most of them are divided into five main groups: textural, statistical, regional, contour and point features. All of them has different informativity and uniqueness and therefore must be carefully selected in presented template and in current frame. That's why the third stage of classification process is proposed to be different.

The term "informational reliability" is introduced for each detected feature and will be explained further. The idea is in forming the vector of frame features with careful selection of each to reach the compromise between informativity and computing costs.

And here the aided classification becomes important since provides not only well known algorithms based on training, but also using correlation analysis over time in terms of context-dependent observation of potential target.

Feature characteristics of frame and target are actually random processes which vary over time (or moments of observation), but must have strong interconnection with template ones. That's why the approach with analyses of correlation is proposed. It includes auto-correlation functions, cross-correlation functions, their forms and intervals of correlation together with moment characteristics of correlation functions.

Let's use the formalization in [5] for feature vector. Suppose that p is some feature detected in an image and associated with a descriptor:

$$\Phi(p) = \left\{ \phi_k(P) \mid k = 1, 2, ..., K \right\},$$

where for all K the feature vector provided by the kth descriptor is

$$\phi_k(p) = \left(f_{1p}^k, f_{2p}^k, ..., vf_{np}^k \right). \qquad (1)$$

By classification it is necessary to find the best correspondence between feature vector $\phi_k(p)$ of current frame and feature vector $\phi_k(q)$ of template of class object.

Usually it is done by calculating the distance between two vectors

$$d_k(p,q) = \left| \phi_k(p) - \phi_k(q) \right|.$$

Classification is done by finding the best match between features vectors. The efficient algorithm to perform this sorting is required to process as quickly as possible. The nearest-neighbor algorithms in the feature space of the image descriptors in Euclidean norm can be used for matching vector-based features. But optimality of the nearest neighbor algorithm depends on the data set characteristics.

The typical solution in the case of classification large datasets is to replace the linear search with an approximate matching algorithm that can offer speedups of several orders of magnitude over the linear search. This is, at the cost that some of the nearest neighbors returned are approximate neighbors, but usually close in distance to the exact neighbors.

Then it is necessary to take into account that there is a definite dependence on correct classification based on previous results of classification (on previous frame of video observation). Such classification is shown in Fig. 1 by step of context-dependent classification.
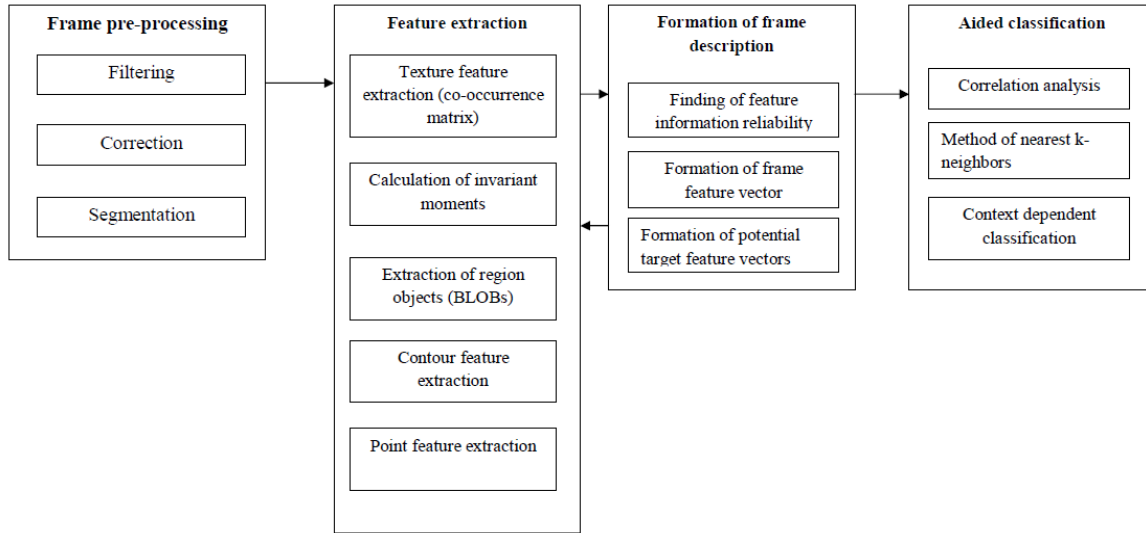
Fig. 1. Structure of aided classification

IV. PROPOSED ALGORITHM OF AIDED
CLASSIFICATION OF GROUND OBJECTS

Let's start from Bayesian classifier. Observation of feature vectors (1) is presented in sequence of N components $X = \left(\phi_k^1(p), \phi_k^2(p), ..., \phi_k^N(p) \right)$, and there are M classes of objects to be recognized $\omega_i, i = 1,..., M$. Solution is $\Omega_i = \left(\omega_i^1, \omega_i^2, ..., \omega_i^N \right)$, that is possible matching between sequence of observation and sequence of classes. The general number of such solutions is M^N. The problem is to solve to what class sequence is necessary to relate the sequence of observation.

Bayesian rule for this case is the following:

$$P(\Omega_i)p\left(X | \Omega_i \right) > P(\Omega_j)p\left(X | \Omega_j \right), \; i \neq j. \quad (2)$$

For solution Markov chain model is used to represent the probabilistic dependence of each class in a sequence:

$$P\left(\omega_i^k | \omega_i^{k-1}, \omega_i^{k-2}, ..., \omega_i^1 \right) = P\left(\omega_i^k | \omega_i^{k-1} \right). \quad (3)$$

Expression (3) is model of so called Markov chain of the first order, assuming that enough to have results of classification on previous step.

Two main assumptions are done. All classes are statistically independent. The probability density function in one class does not depend on others.

Thus, (3) can be re-written as follows:

$$P\left(X | \Omega_i \right) = \prod_{k=1}^{N} p\left(x_k | \omega_i^k \right). \quad (4)$$

Combining (3) and (4) Bayesian rule is obtained as following. For observation sequence $X = \left(\phi_k^1(p), \phi_k^2(p), ..., \phi_k^N(p) \right)$ the classification is done by class sequence $\Omega_i = \left(\omega_i^1, \omega_i^2, ..., \omega_i^N \right)$ in such a way to maximize the value.

$$\begin{aligned} &P\left(X | \Omega_i \right) P\left(\Omega_i \right) \\ &= P\left(\omega_i^1 \right) p\left(x_1 | \omega_i^1 \right) \prod_{k=2}^{N} P\left(\omega_i^k | \omega_i^{k-1} \right) p\left(x_k | \omega_i^k \right). \end{aligned} \quad (5)$$

The search requires the calculation of each last expression in (5), that is, total number of multiplication operations reaches $O\left(NM^N \right)$. But the optimization can be done in assumption that Ω_i and Ω_j differ from each other by only the last classes, i.e. $\omega_i^k = \omega_j^k$, then most of calculation can be eliminated since they are doubled.

The proposed algorithm was used for detection by only two features related to BLOB (Binary Large OBjects) analyses [6]. The image was first segmented in three basic color layers, and then combined into separate BLOB objects (Fig. 2).

Videos are image sequences over the time. A video is a function of image intensity (color) over time t. The camera takes continuously capture of the images. When the conversion of the video into frame is done, it gives the frames at different time period. Once the frames are ready then the processing is done on them.

After the video to frame conversion preprocessing is done on each frame to reduce the noise which is present in frame. The pre-processing is done using the mean filter, convolution filter, and median filter. The mask of the filter will multiply with the frame and noise will get removed so that

the result is accurate. Due to this only the background or foreground objects are preset and unwanted factors in the frame presents at the time of the capturing the images like dust.

Once the background frame [7] is initialized it will referred as the reference frame. There are many ways to obtain the initial background image. For example, take the first frame as the background directly, or the average pixel brightness of the first few frames as the background or using a background image sequences without the prospect of moving objects to estimate the background model parameters.

After the background frame is initialized the subtraction of the current frames and the reference frame is done for the moving object detection. The subtraction will be done pixel by pixel of the both frames. So that updated background is given to the reference frame. Then subtracted image is given to the segmentation.

After the subtraction of the frames the subtracted image is then segmented using the threshold value. This value is practically set or by Otsu method.

That subtracted frame gives the subtracted value of the each pixel that pixel value is compared with the threshold value if the subtracted pixel value is greater than the threshold value then it will represented by the 1 and if not greater then represented by 0. So the segmented image gives the moving object in white and the background is black. It will detect the moving object in the frame.

The segmented frame is now given to the morphological filtering for reducing the noise. The function of the morphological filtering is removing the small regions probably created by noise; fill up unnecessary cavities, smoothing boundaries, extracting edges. It will give pixel level operations. After the segmentation and morphological filtering the moving object is clearly seen in the frame and that will be the output of the system.

Classification was done by two main feature parameters: area and center of mass. Prior position and expected area size of target object were given and then analyzed among other candidates. Probability characteristics are presented in Fig. 3. The observations were done for period in 3 seconds with frame per second 30 (frames 476 to 566).
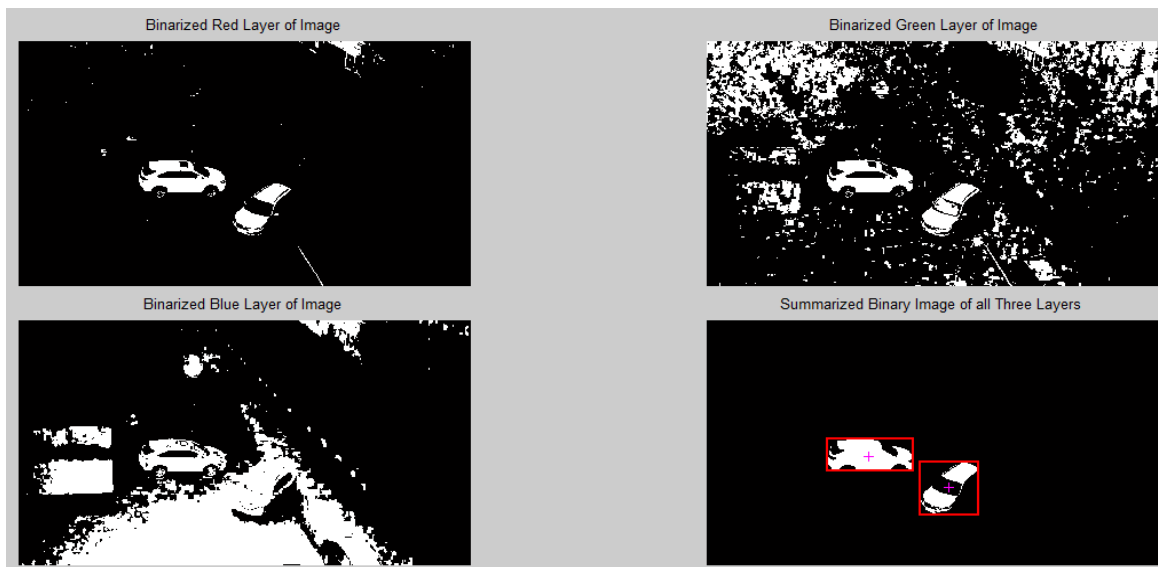

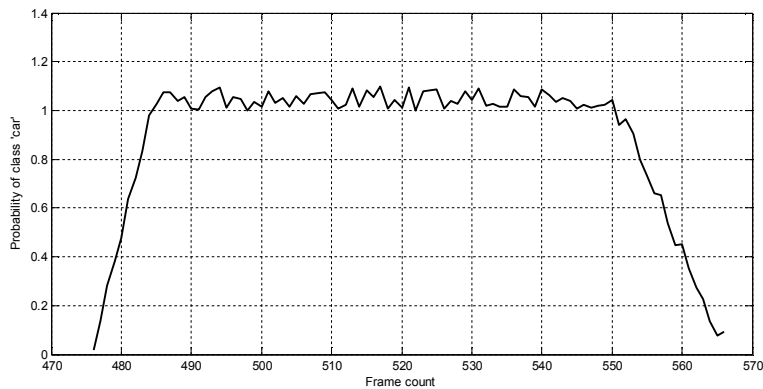Fig. 2. Binary Large OBjects segmentation

Fig. 3. Probability of class selection for series of video observation.

V. CONCLUSIONS

The experimental results of object classification by proposed structure prove high stability by careful selection of features. Accuracy in prediction of before detected object location is provided by context-dependent classification. The analysis of system prototype work in test conditions implemented in Matlab showed that average rate of a single frame (1280 x 720) processing is 300ms which makes it impossible to implement in real-time applications.

REFERENCES

[1] Teutsch, W. Krüger, and N. Heinze, "Detection and classification of moving objects from UAVs with optical sensors," *SPIE Defense, Security, and Sensing.* 2011, pp. 80501J–80501J.

[2] Syriamkin, and V. Shidlovkiy, *Correlation-extremal radio navigation systems*, Tomsk izdatelstvo, 2010, 316 p.

[3] Javed, and M. Shah, "Tracking and object classification for automated surveillance," in *European Conference on Computer Vision*, Springer, Berlin, Heidelberg, 2002, pp. 343–357.

[4] Jianzhuang, L. Wenqing, and T. Yupeng, "Automatic thresholding of gray-level pictures using two-dimension Otsu method," in *IEEE International Conference on Circuits and Systems*, 1991, Conference Proceedings, China, 1991, pp. 325–327.

[5] Hassaballah, A. Abdelmgeid, and H. Alshazly, "Image Features Detection, Description and Matching," *Image Feature Detectors and Descriptors.*" Springer International Publishing, 2016, pp. 11-45.

[6] Chen, Y. Lin, and T. Y. Chen, "Intelligent vehicle counting method based on blob analysis in traffic surveillance," in *IEEE Second International Conference on Innovative Computing, Information and Control*, 2007, pp. 238–238.

[7] Piccardi, "Background subtraction techniques: a review," in *IEEE international conference on Systems, man and cybernetics*, 2004, vol. 4, pp. 3099–3104.

Maryna Mukhina. Doctor of Engineering Science. Professor.
Aviation Computer-Integrated Complexes Department, Educational & Research Institute of Information and Diagnostic Systems, National Aviation University, Kyiv, Ukraine.
Education: National Aviation University, Kyiv, Ukraine, (2002).
Research area: Navigation and motion control.
Publications: more than 60 papers.
E-mail: marinamukhina79@gmail.com

Iryna Barkulova. Post-graduate student.
Aviation Computer-Integrated Complexes Department, Educational & Research Institute of Information and Diagnostic Systems, National Aviation University, Kyiv, Ukraine.
Education: National Aviation University, Kyiv, Ukraine, (2014).
Research area: Navigation and motion control.
Publications: more than 12 papers.
E-mail: irochkaseden@gmail.com

М. П. Мухіна, І. В. Баркулова. Структура допоміжної класифікації наземних об'єктів за відеоспостереженням
Проведено аналіз системи класифікації за даними відеоспостереження. Сформульовано, що для виділення ознак та їх класифікації, потрібно отримати нормалізовані гіпотези для виявлення ознак об'єкта, приймаючи до уваги орієнтацію камери та висоту польоту. Запропоновано структуру допоміжної класифікації на основі ймовірнісних моделей, таких як класифікатор Байєса та Марковські процеси. Запропонований алгоритм був застосований для виявлення лише двох ознак, що відносяться до BLOB аналізу. Класифікацію було проведено відповідно до двох основних характерних параметрів – площа та центр мас. Вектор ознак складається лише з найінформативніших складових та дозволяє мінімізувати ризики. Результати підтвердили надійність класифікації відповідно до набору відео кадрів у умовах неповних даних дескриптивного простору.
Ключові слова: вектор характерних ознак; дескриптивний простір; допоміжна класифікація; класифікатор Байєса; Марковські процеси.

Мухіна Марина Петрівна. Доктор технічних наук. Професор.
Кафедра авіаційних комп'ютерно-інтегрованих комплексів, Навчально-науковий інститут інформаційно-діагностичних систем, Національний авіаційний університет, Київ, Україна.
Освіта: Національний авіаційний університет, Київ, Україна, (2002).
Напрям наукової діяльності: навігація та управління рухом.
Кількість публікацій: понад 60.
E-mail: marinamukhina79@gmail.com

Баркулова Ірина Володимирівна. Аспірант.
Кафедра авіаційних комп'ютерно-інтегрованих комплексів, Навчально-науковий інститут інформаційних діагностичних систем, Національний авіаційний університет, Київ, Україна.
Освіта: Національний Авіаційний Університет, Київ, Україна, (2014).
Напрям наукової діяльності: навігація та управління рухом.
Кількість публікацій: понад 12.
E-mail: irochkaseden@gmail.com

М. П. Мухина, И. В. Баркулова. Структура вспомогательной классификации наземных объектов по видеонаблюдению
Проведено системный анализ классификации за данными видеонаблюдения. Сформулировано, что для выделения признаков и их классификации, нужно получить нормализированные гипотезы для обнаружения признаков объекта, принимая во внимание ориентацию камеры и высоту полета. Предложено структуру вспомогательной классификации на основании вероятностных моделей, таких как классификатор Байеса и Марковские цепи. Предложенный алгоритм был применен для обнаружения только двух признаков, что относятся к BLOB анализу. Классификация было произведена относительно двух основных параметров – площадь и центр масс. Вектор признаков состоит из наиболее информативных составляющих и позволяет минимизировать риски. Результаты подтвердили надежность классификации соответственно набору видео кадров в условиях неполных данных дескриптивного пространства.
Ключевые слова: вектор характерных признаков, дескриптивный простор, вспомогательная классификация, классификатор Байеса, Марковские цепи.

Мухина Марина Петровна. Доктор технических наук. Профессор.
Кафедра авиационных компьютерно-интегрированных комплексов, Учебно-научный институт информационно-диагностических систем, Национальный авиационный университет, Киев, Украина.
Образование: Национальный авиационный университет, Киев, Украина, (2002).
Направление навигация и управление движением.
Количество публикаций: более 60.
E-mail: marinamukhina79@gmail.com

Баркулова Ирина Владимировна. Аспирант.
Кафедра авиационных компьютерно-интегрированных комплексов, Учебно-научный институт информационно-диагностических систем, Национальный авиационный университет, Киев, Украина.
Образование: Национальный авиационный университет, Киев, Украина, (2014).
Направление навигация и управление движением.
Количество публикаций: более 12.
E-mail: irochkaseden@gmail.com