Methods of automatic correction of image parameters in computer vision systems

Dluzhevskyi A O

Department of Computerized Control Systems, National Aviation University, 1 Liubomyra Huzara ave., Kyiv 03058, Ukraine

Corresponding author e-mail: dlandrew@ukr.net

Abstract. Video surveillance – one of the most popular and effective security measures. Video surveillance systems are being introduced in public utilities, transport, hotel industry, industry, public institutions and many others. However, they are deployed not only with security objectives. However, when deploying surveillance systems, there is a problem with the selection of the correct camera settings. This problem lies in the fact that a specialist can carry out a basic setting, which is performed under certain environmental conditions. But over time, the parameters of the external environment may change (for video surveillance systems, the main interference factor is lighting, which affects the brightness and contrast of the image). Task – to develop the project of intellectual video surveillance system that will be able to analyse data from up to five IP-cameras and count people who pass in front of these cameras. Provide collected data sending on the remote server. This video surveillance system has to use developed method of automatic correction of image.

1. Introduction

Taking into account the volume of information that are taking place in different areas of our lives, including industrial, commercial and household sectors, we can reach a conclusion that the task of gathering this information, its processing and analysis are becoming more complex every year.

Most of this information relates to human behavior and is visual. This information can be very useful not only in terms of security, identification and prevent dangerous situations, but also in terms of drawing up statistics that can be further processed by decision-making systems etc.

Using human resources to analyze such information can be very inefficient due to significant influence of "human factor" and involves considerable expense of resources. Only thanks to modern hardware and software tools, task processing and analysis of such information may be effective.

Thus, the use of intellectual video analytics is a good addition to the already existing systems of video surveillance and encourages the introduction of new video surveillance systems at enterprises of different size and direction. These companies can be:

– Shops. An example of an analysis of the length of queues to prevent excessive queues by opening the extra cash desk;

– Shopping centers. The analysis of the ways in which the people walk can be useful for advertising, and analysis the number of people give ideas about the most popular outlets and general trends in trade;

– Educational institutions. Analysis of the number of students will give the presentation about how many students visit lessons in general, and certain subjects in particular case. That will give ideas about how effectiv the educational system is and about the quality of teaching;

– Offices. After analyzing the habits of workers, it is possible to get information about the time that the worker spends at work, the statistic of being late and early leaving from the workplace;

– Factory. Analysis of the quality of work, using the example of packers with taking into account such parameters as: speed packaging, the number of products packaged for change and comparing it with the norms, errors the packer has made during the change, and so on;

Also intelligent video surveillance system can be used elsewhere. For example:

– Public places. Defining objects left unattended in public places and transport, etc.;

– Highways. Determination of traffic jams formation in order to alert road users, collect statistics, etc.

Although intelligent video surveillance system can be used in video systems of various types, but the largest distribution IP network surveillance is beginning to acquire. The reason is that such systems are relatively inexpensive and easy scaling. Network video surveillance system can easily be used in a small office as well as at the big enterprises and factories.

Today, there are different approaches to solve problems of automatic image identification. Great number of approaches is conditioned due to the volume of data being processed to form a base of knowledge in recognition systems. It is also conditioned due to difficulties in defining of object's edges, located in the image. Separately, there is a problem related to the human factor. This kind of problems affect on the quality of decision-making during continuous operation. A promising way to overcome these difficulties is the automation of image processing that can significantly reduce the amount of routine and labor-intensive work related to transform the images used. This can increase the objectivity of used visual signs detection of the studied images. But to solve this problem, input data should be presented in prepared format that takes away the load (and therefore the demands on hardware) from pattern recognition system and allows us to execute processes in parallel.

The process of identification of objects in computer systems is aimed at creating identifiers for objects in a computer system. There are a number of problems, in the process of identification. One of them - a high impact of noise at the input. While the input signal has a high level of noises, it is not possible to uniquely identify an object. This includes both the classification of the object and determination of the specific instance.

To be able to correctly identify objects, the noise has to be removed from input signal. Such noises can be peak and Gaussian noises. Purification of the input data from noise is performed by using various filters. Depending on the type of noise some types of filters may be slightly more effective than others.

In practice, to reduce noise level different methods of filtration can be used. There can be such filters as: linear, median, and used adaptive. Filtration method should be chosen depending on the nature of the noise in the signal. Using algorithms for finding the optimal preprocessing method, depending on the type of noise in the input will reduce the number of problems related with the identification process. A system must be able to estimate the optimal filtering methods to select an optimal method of filtration.

2. Principles of video image processing

Video surveillance - one of the most popular and effective security measures. Video surveillance systems are being introduced in public utilities, transport, hotel industry, industry, public institutions, sports and leisure centers, commercial organizations and projects of the "safe city" level. However, they are deployed not only with security objectives.

The equipment offered by leading vendors, allows you to create systems of any sizes. In addition to fixed and PTZ cameras SD and HD [1], it can be infrared imagers, video servers, disk arrays, equipment for data channels, and so on. The modern network cameras have built-in video analytics and supports multiple video streams, while the software solves various problems of video surveillance and includes a system for monitoring and controlling recording and video client supporting a variety of devices.

Video analytics systems allow to identify abandoned or taken away objects and cases of "roaming" (when the object is in a particular area than the specified time), they are able to monitor adherence to the route, to fix the intersection lines to analyze the flow of visitors and count objects (this data can be used in marketing research).

Some manufacturers install the means of video analytics on cameras and encoders [2]. As a result, simultaneously with the video stream metadata is received that can be used by operators and recorded into storage for later rapid search by tags. However, the analytical capabilities of the camera itself are limited by resources of its processor. The most advanced systems of analytics are developed by specialized companies and are installed on the powerful dedicated servers.

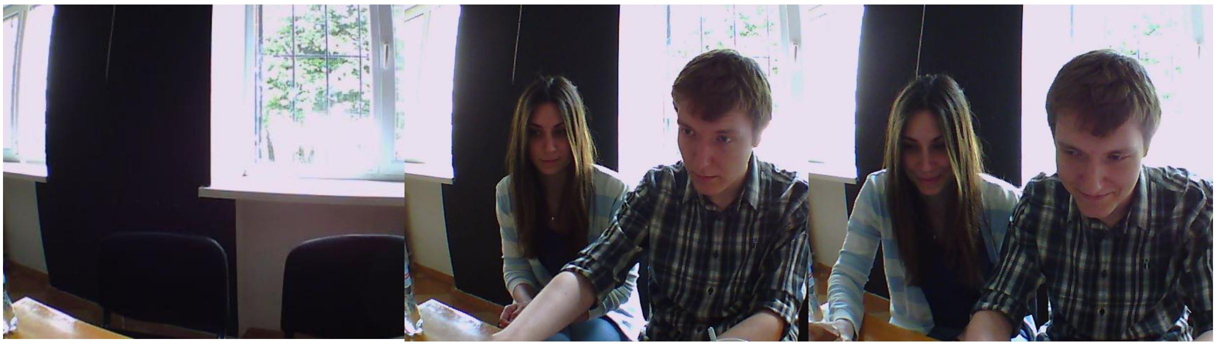
Video analytics - a good assistant for operator, whose efficiency decreases rapidly with the passage of time, some researches show that it can miss up to 95% of the events after 20 minutes of continuous monitoring. Analysis Tools video suggest that it is necessary to take a close look at the computer screen, which displays the incident, defined as an alarm system. The prospect of the next few years - tracing routes, and search video files recorded by several cameras.

2.1. Technique of motion detection on video image

Video motion detection algorithm, which was used, is based on the technique of image correlation. It is possible to represent video image from surveillance camera as an array of images (Figure 1).

$$A = [a1, a2, a3, \dots, an]$$

where a is frame, n is serial number of the frame.



A B C

Figure 1. Example of video image as a sequence of frames.

The algorithm operates with two frames. The first of them will be called undeformed (basic), let’s denote it’s pixel intensity matrix as F, the second frame is deformed relative to the base, so that there is a change in pixels intensity. Let’s denote it’s pixel intensity matrix as G. Firstly, from the frame array A we select it’s element and assume it is undeformed. So:

$$F = a_i$$

Where i is serial number of selected element

a_i is pixel intensity matrix of i-th element

Selection of an element can be either random or have a can be choosen using a certain logic. Selection of frame can have influence on detection of distarted zone on the image. Then we take the next frame:

$$G = a_{i+1}$$

Now, let’s build a pixel intensity difference matrix:

$$R(i,j) = |F(i,j) - G(i,j)| \quad (1)$$

Elements of intensity matrix $R(i,j)$ are not equal to 0 correlate with deformed image by indexes. Now if we find related fields in matrix R – we get areas on the image, where change of pixel intensity take place.

For the task of related fields detection an algorithm, based on ABC mask, was used. The first step of this algorithm is preparatory. This algorithm is able to operate with binary arrays only. So, firstly, it is needed to convert array *R* to binary form (Figure 2). We will use a threshold binarization.

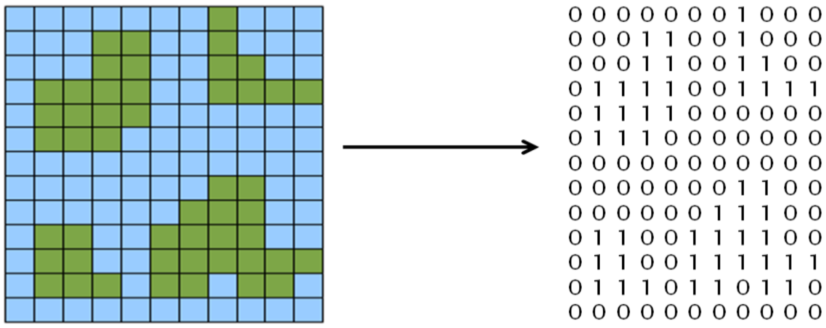


Figure 2. Array binarization.

The task of related fields detection can be solved in different ways. Some techniques are based on template usage, other have recursive realization.

Each technique has a number of advantages and disadvantages, so templates can be quite time-consuming, and the use of recursion is quite controversial in terms of the use of stack memory of the program

The field is called related if each pixel from this field has a neighbor from the same field. Given algorithm is four-connected but eight connected implementation is possible

The idea of this algorithm is based on using a corner – ABC mask (Figure 3) [3].

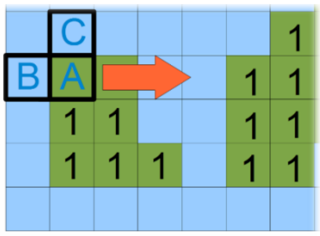


Figure 3. ABC mask and direction of consecutive scanning of image.

The benefit of this algorithm is that it is single-pass and non-recursive. Passing through the image by this mask is performed from left to right and from top to bottom. It is assumed that there is no objects beyong the borders of the image. So, in the case, when B or C gets there - some additional check during scanning is needed [4].

There can be 5 possible positions of the mask on the image (Figure 4).

1) when all of three mask components are not marked. In this case the pixel is just passed;

2) when only element A is marked – this case means that new object with new number is created;

3) when only element B is marked – in this case current pixel (pixel A) is marked by the mark from element B;

4) when only element C is marked – in this case current pixel A is marked by the mark from element B;

5) in this case we talk about that marks (object numbers) B and C are related. In some cases it is possible to build a graph of equivalence of such marks. This graph can be further processed. But it is also possible to go another way: in the case, if B is not equal to C we change all processed pixels, marked as C to mark B.

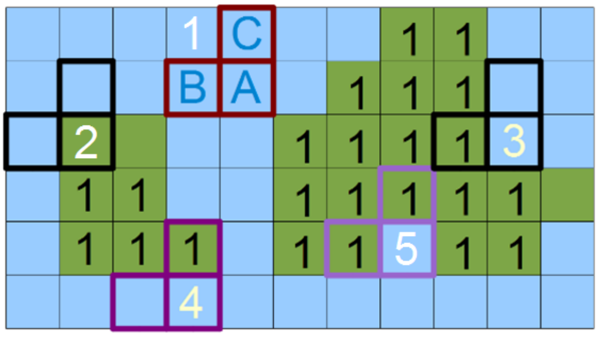


Figure 4. Five possible positions of ABC mask.

The disadvantages of related fields detection using this method is the fact that the field numbers are not in consecutive order.

For further analysis of objects in the video image, metadata is formed with respect to each image. This metadata contains various information about the objects in the image. Such information can be the position of the object on the image, object's ID, etc.

Position the object on the image can be represented as a rectangle which is circumscribed around the object. Thus, the position of the object will be expressed by begining coordinates of circumscribed rectangle (i, j) and its size in two-dimensional space (w, h).

$$O = f(i, j, w, h)$$

where *i* is the first appearance of the element in the array, according to rows

j is the first appearance of the element in the array, according to columns

w is maximal width of the object in pixels

h is maximal height of the object in pixels

To find the initial coordinates of objects is search of array elements from top to bottom and left to right. Found objects are stored in the list. We call it *Obj*. There are some possible scenarios:

1) The current array element is equal to 0. This means that we are in the area of the array with no objects;

2) The current array element is equal to 0. This means that we are in the area of the array with an object; Element mark *(k)* is checked;

2.a) During the element's label check it was found that any object with such label has been registred in the list yet: $O(k) \notin Obj$;

2.b) During the element's label check it was found that an object with such label has been already registred in the list *Obj*: $O(k) \in Obj$.

If $O(k) \notin Obj$ mens, that this is the first appearance of this object found in the array. It's initial and final coordinates are recorded (istart,jstart та iend, jend correspondingly). And $O(k)$ is added in the list Obj;

If $O(k) \in Obj$ the coordinates of current element are checked corresponding to recorded ones (istart,jstart,iend,jend) of the object *O(k)* using such conditions:

- If *j* <*jstart*, then a new extreme left margin of the object *O(k)* was found. This means that we need to "move" the initial coordinate of the object according to index j: jstart = j;

- If j> jend, then found a new extreme right border and to "move" the initial coordinate of the object according to index *j: jend = j*.

Also, we assume that this is the last row of the array, which contains elements with the labels the of this object. So *iend* = *i*; If the next line will have more elements of the object - the final value iend will be overwritten in any case

To test this algorithm, two images were taken [5]:

- entirely homogenious background (as basic image);

- Various geometric shape, printed on the solid background of the previous image (as deformed image).

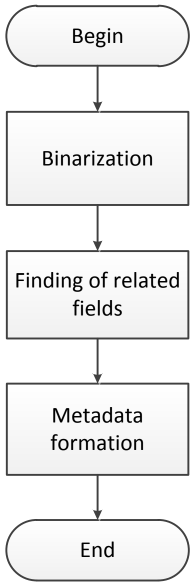
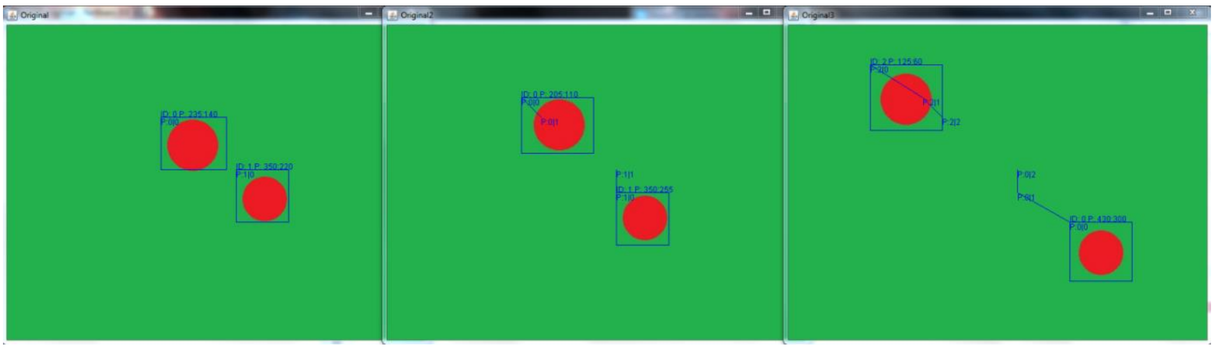


Figure 5. Steps, needed to form frame.

It is possible to perform different actions over the received metadata. For example, if you find one and the same objects in the elements of array A - you can track their movement through frames. This can be useful for creating algorithms that analyze the behavior of objects in the video image.

To find the same object in different frames it is proposed to consider object $Obj_m(a_2)$ as object $Obj_n(a_1)$ if the distance between them is less than between the object $Obj_n(a_1)$ and any other object in the list Obj. Here $Obj(a_p)$ – is the list of objects in a scene with a number *p*; $Obj_m(a_p)$ - object with index *m* frame with number *p*.

For a demonstration of this approach (Figure 6) with the metadata for each new frame also dots connected by a line are displayed. These dots represent previous position of each object in the image.



A B C

Figure 6. Location of the same objects in different frames.

The availability of data about the position of objects in the previous frames allows us to conduct additional analysis of the events occurred in the image, such as the direction of the object. Based on data collected, events of line intersection (Figure 7) and entry into restricted area (Figure 8) have been implemented. This finds its application in systems of computer vision and video surveillance.

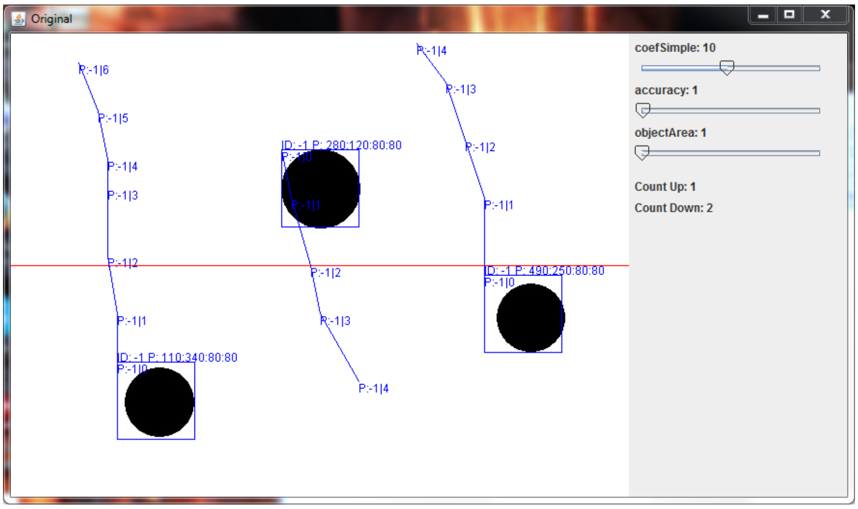


Figure 7. Line intersection.

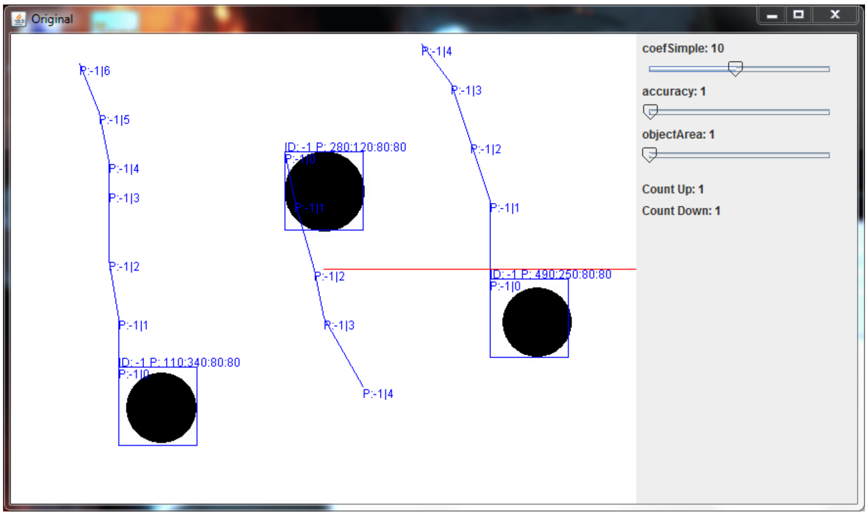


Figure 8. Restricted area intersection.

2.2. Noise of a digital image

Described algorithm of detection of areas with deformation has certain disadvantages. When using this algorithm for image analysis that has been received from Web-cameras in real time (Figure 9), the quality of deformed zones detection significantly decreases. The main reason for this phenomenon - the presence of noise on the video frames.

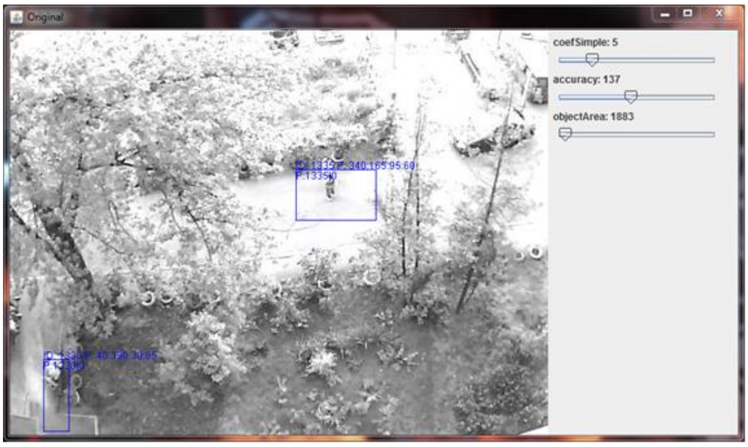


Figure 9. Image from Web-camera.

The source of the noise is the matrix of the camera. The algorithm was stable on test frames due to the lack of noise. The amount of noise depends on many factors. Here are some of them:

- Illumination of the scene
- Possibilities of the camera's matrix
- The number of colors

Using empirical method, it was determined that the noise level in the black and white image is significantly less. This is because in black and white image the noise affects only one channel – brightness of white. In color images noise affect simultaneously on the three channels: R, G, B (red, green and blue). Therefore, noise in color images makes much more change in the intensity of the pixels.

The noise in the video image is dynamic [6]. That is different frames have different noise. Even if the is no action on the scene, the algorithm described before will detect deformation because of the change in the pixel intensity, caused by the noise.

Image noise is random (not present in the object imaged) variation of brightness or color information in images, and is usually an aspect of electronic noise. It can be produced by the sensor and circuitry of a scanner or digital camera. Image noise can also originate in film grain and in the unavoidable shot noise of an ideal photon detector. Image noise is an undesirable by-product of image capture that adds spurious and extraneous information.

The original meaning of "noise" was and remains "unwanted signal"; unwanted electrical fluctuations in signals received by AM radios caused audible acoustic noise ("static"). By analogy unwanted electrical fluctuations themselves came to be known as "noise". Image noise is, of course, inaudible [7].

The magnitude of image noise can range from almost imperceptible specks on a digital photograph taken in good light, to optical and radioastronomical images that are almost entirely noise, from which a small amount of information can be derived by sophisticated processing (a noise level that would be totally unacceptable in a photograph since it would be impossible to determine even what the subject was).

3. Techniques of image parameters analysis and correction

Preprocessing and preparation of images is equally important and difficult step in the process of image analysis, pattern recognition or decision-making [8]. Properly prepared image guarantees stable work of image recognition and identification algorithms [9].

The process of identification of objects in computer systems is aimed at creating identifiers for objects in a computer system. There are a number of problems, in the process of identification. One of them - a high impact of noise at the input. While the input signal has a high level of noises, it is not possible to uniquely identify an object. This includes both the classification of the object and determination of the specific instance [10].

To be able to correctly identify objects, the noise has to be removed from input signal. Such noises can be peak and Gaussian noises. Purification of the input data from noise is performed by using various filters. Depending on the type of noise some types of filters may be slightly more effective than others.

In practice, to reduce noise level different methods of filtration can be used. There can be such filters as: linear, median, and used adaptive. Filtration method should be chosen depending on the nature of the noise in the signal. Using algorithms for finding the optimal preprocessing method, depending on the type of noise in the input will reduce the number of problems related with the identification process. A system must be able to estimate the optimal filtering methods to select an optimal method of filtration.

3.1. Reasons of digital noise occurence

Digital noise is noticeable in the image in the form of superimposed masks of pixels of random color and brightness. On cameras with color filter array (the majority of digital cameras belong to this type) color noise usually has a visually larger grain than the pixels in the images. This is a side effect of the algorithm for obtaining a full color image

For the three matrix systems or matrix without filter the noise will be more fine-grained. In the color image noise can have a different intensity for different image channels. This visually paints it. The noise in pictures shot with the use of glow lamp has predominantly yellow and blue hues, not green and purple. The fact is that, although initially all pixels are equally exposed to noise, but after applying white balance, blue channel of image changes, and therefore the noise it increases stronger.

Noise is noticeable on plain areas, and in particular - in the dark areas of the image. As a rule, in electronics, we usually speak about signal-to-noise ratio. Visually, we can compare the noise of different matrices as follows: bring the two paired test photos to the same size and the same brightness, and after this visually evaluate color noise.

Use case analysis was performed in order to reach high-level requirements for the system. The first step in use case modeling is to identify actors who interact with the system (either provide input or receive output from the system). For a typical DVR system, three actors were identified. The first is the user who interacts with it to record and watch video streams. The second is the DVR data service that is required to update schedule information on the system. The third is the television provider (typically a cable or satellite TV company) used by the customer and whose signal is connected to the DVR unit [10].

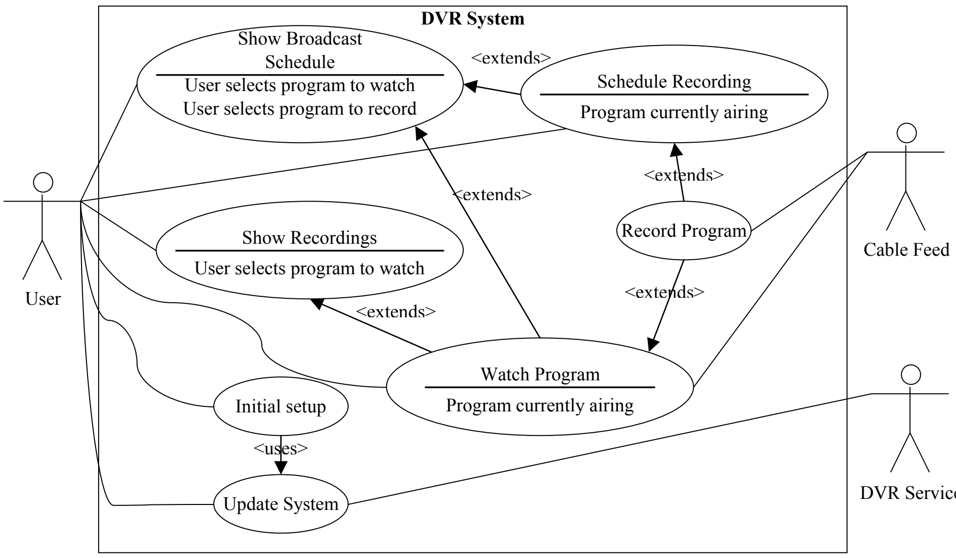


Figure 10. Use case diagram of DVR program module.

So the purpose of the class diagram can be summarized as:

– analysis and design of the static view of an application;
– describe responsibilities of a system;
– base for component and deployment diagrams;
– forward and reverse engineering [10].

As an example, an activity diagram of digital video recording module of the system is given (Figure 11).

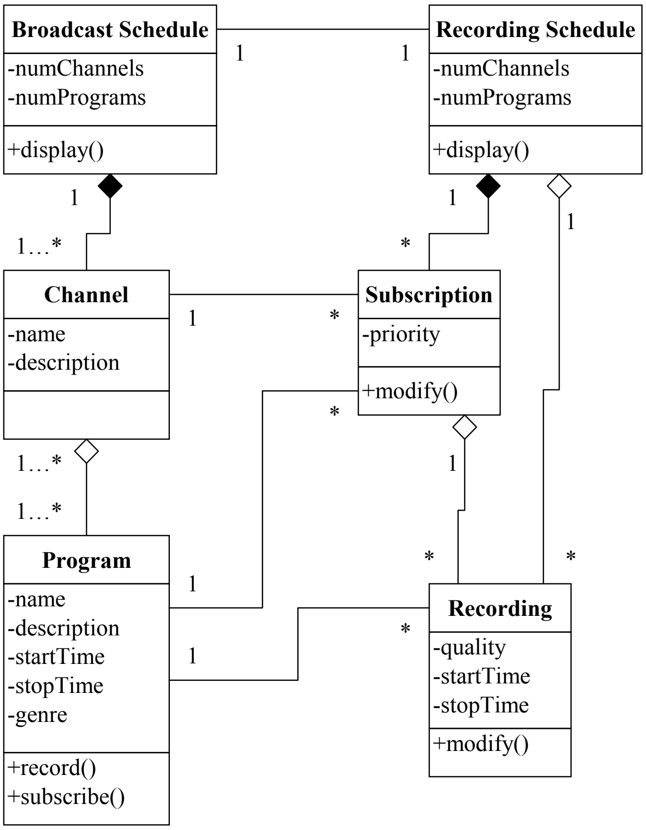


Figure 11. Class diagram of DVR program module.

Using OpenCV library and Java programming language the intellectual video surveillance system was build. The main window of the program is shown below (figure 12). The system recognizes groups of people on a outdoor scene. Such kind of scenes are considered to be the most problematic and difficult because of lots of moving objects, such as shadows and foliage.

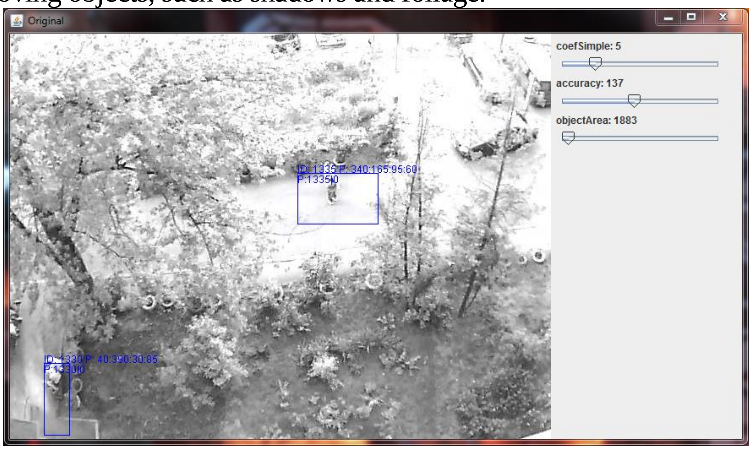


Figure 12. Two groups of people are detected on a outdoor scene.

Concussions

The process of identification of objects in computer systems is aimed at creating identifiers for objects in a computer system. There are a number of problems, in the process of identification. One of them - a high impact of noise at the input. While the input signal has a high level of noises, it is not possible to uniquely identify an object. This includes both the classification of the object and determination of the specific instance.

To be able to correctly identify objects, the noise has to be removed from input signal. Such noises can be peak and Gaussian noises. Purification of the input data from noise is performed by using various filters. Depending on the type of noise some types of filters may be slightly more effective than others.

In practice, to reduce noise level different methods of filtration can be used. There can be such filters as: linear, median, and used adaptive. Filtration method should be chosen depending on the nature of the noise in the signal. Using algorithms for finding the optimal preprocessing method, depending on the type of noise in the input will reduce the number of problems related with the identification process. A system must be able to estimate the optimal filtering methods to select an optimal method of filtration.

Increasing brightness allows us to neglect some low-level noise with tiny grains. Changing a color scheme to grayscale prevents the noise to affect different channels of image, this makes further image processing much easier. And applying different kind of filters allows us to fight agains noises that are caused by a variety of reasons.

Also the idea of how to evaluate the quality of image filtration was proposed. To test the technique of filter efficiency evaluation a series of experiments were performed. During these experiments it was shown that given method allows us to estimate the effectiveness of a particular method of filtration.

The results of the system's work are the events (messages) that can be transferred to the system operator or video recorded in the video archive for later retrieval. In addition, video analysis generates metadata that is a data structure that describes the content of each frame of the video sequence. The metadata includes information such as location and object identifiers (usually in the form of alarm frames), the trajectory and velocity of objects, data division or merging of objects, information on the begining and ending of the alarming situation. Metadata can be written in the video archive and reproduced along with the video.

The software product has been testing during the development stage and is currently testing on the real conditions.

References

[1] Dalal N and Triggs W 2005 Histograms of Oriented Gradients for Human Detection *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* vol 3 (San Diego: IEEE Computer Society) pp 886-893

[2] Dalal N and Triggs W 2005 Histograms of Oriented Gradients for Human Detection *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* vol 3 (San Diego: IEEE Computer Society) pp 886-893

[3] Rakesh Chandra Joshi, Mayank Joshi, Adithya Gaurav Singh and Sanjay Mathur 2018 Object Detection, Classification and Tracking Methods for Video Surveillance: A Review *2018 4th International Conference on Computing Communication and Automation (ICCCA)*

[4] Miranda A A, Le Borgne Y A and Bontempi G 2008 New Routes from Minimal Approximation Error to Principal Components *Neural Processing Letters* vol 27 pp 197-207 https://doi.org/10.1007/s11063-007-9069-2

[5] Tran D and Sorokin A 2008 Human activity recognition with metric learning *European Conference on Computer Vision* (Marseille: Springer) pp 548-561

[6] Neeraj K Kanhere and Stanley T Birchfield 2010 A Taxonomy and Analysis of Camera Calibration Methods for Traffic Monitoring Applications *IEEE Transactions on Intelligent Transportation Systems (Volume: 11, Issue: 2, June 2010)* pp 441-452

[7] Miranda A A, Le Borgne Y A and Bontempi G 2008 New Routes from Minimal Approximation Error to Principal Components *Neural Processing Letters* vol 27 pp 197-207 https://doi.org/10.1007/s11063-007-9069-2
[8] Tran D and Sorokin A 2008 Human activity recognition with metric learning *European Conference on Computer Vision* (Marseille: Springer) pp 548-561
[9] Bradley D and Roth G 2007 Adaptive thresholding using the integral image *Journal of Graphics Tools*. Volume 12, Issue 2 pp 13-21 NRC 48816
[10] Dluzhevskyi A and Artamonov Y 2018 Technique of Image Preparation for Processing in Video Surveillance Systems *Polish journal of science* #7 vol. 1(Warszawa) pp 31-34 (https://www.poljs.com/wp-content/uploads/2018/09/POLISH-JOURNAL-OF-SCIENCE-%E2%84%967-VOL.-1-2018.pdf)