Defect detection on videos using neural network

Roman Sizyakin[1,*], *Nikolay* Gapon[1], *Igor* Shraifel[1], *Svetlana* Tokareva[1], *Dmitriy* Bezuglov[1]

[1]DSTU, 344000, Rostov region, Rostov-on-Don, Gagarina 1, Russia

Abstract. In this paper, we consider a method for defects detection in a video sequence, which consists of three main steps; frame compensation, preprocessing by a detector, which is base on the ranking of pixel values, and the classification of all pixels having anomalous values using convolutional neural networks. The effectiveness of the proposed method shown in comparison with the known techniques on several frames of the video sequence with damaged in natural conditions. The analysis of the obtained results indicates the high efficiency of the proposed method. The additional use of machine learning as post-processing significantly reduce the likelihood of false alarm.

1 Introduction

With the rapid development of digital systems for reception, transmission and storage are a secondary analogue system of the video. Unfortunately, the analogue video sequence is such that over time the material recorded on them, are partially distorted. This is due to substandard equipment, improper storage of documents, as well as their physical damage. All this leads to the film abnormal values in the form of a single or group of pixels which value of brightness will significantly stand out from other pixels. The observation time of the abnormal data values, usually not more than a few frames in a video sequence, most often on one. Interference significantly degrades the visual quality of the video. Therefore, it is necessary to solve the task of automated detection of distorted values. In most cases, the problem of the digital restoration carried out by the artists for each frame manually, which significantly limits the speed of processing of archival recordings.

Dust removal based on applying some spatial or temporal processing algorithm frame. So, the spatial methods assume that the defect has a small area on the borders of the pronounced gradient values, that is, different from the rest of the image. The methods use grey-scale morphology (dilation, erosion, opening and closing) [1,2].

Some detectors are based on the analysis of some number of pixels, taken from the current, previous and subsequent frames. By this method, it is possible to carry SDIa, ROD and SROD detector. The peak detection index (SDI) is the most basic heuristic for detecting temporal discontinuities in the video sequence. This detector compares the brightness value of each pixel in the current frame $Y_{i,j}^{(k)}$ with the brightness values of the previous and subsequent frames of the video sequence. Then compute the minimum of the quadratic difference [3]. A method of ranking pixels (ROD) heuristic detector based on ordinal statistics [4]. A defect is detected as one of the ranked sequences exceeds a predetermined threshold T_l. The threshold T_l set by the user and determines the sensitivity of the detector. The difference is that the ROD from the SROD detector is that of the potential detector is not the median value and the difference between the local maximum and the local minimum [5]. The SROD detector compares the pixel value obtained as the result of motion compensation, the current pixel value. Damage is detected if the value of the current pixel is beyond the threshold.

Most of the defect detection methods using fixed threshold values to produce a mask with the presumed localisation of abnormal values. Incorrect assignment of a threshold may lead to an increase in the probability of false alarm at his fall, and to the insufficient detection while increasing its value.

2 Proposed method

The proposed method consists of three steps: motion compensated frames, the preprocessing of the detector, which is based on the ranking of pixel values, and classify all pixels with abnormal values using convolutional neural networks.

Video sequence assumes the following simplified mathematical model:

$$Y_{i,j}^{(k)} = (1 - d_{i,j}^{(k)}) \cdot S_{i,j}^{(k)} + d_{i,j}^{(k)} \cdot c_{i,j}^{(k)} \tag{1}$$

where $Y_{i,j}$ - the observed frame with strings $i = \overline{1,I}$ and columns $j = \overline{1,J}$, $k = \overline{1,K}$ - frame number, K - the number

---
* Corresponding author: roma_sz@mail.ru

of frames in a video sequence, $S_{i,j}$ - original (undamaged) image, $d_{i,j} \in \{0,1\}$ binary mask defects, which shows what site the image was corrupted, $c_{i,j}$ the mask, which contains values of brightness defects.

The scheme of the proposed algorithm is present in figure 1.
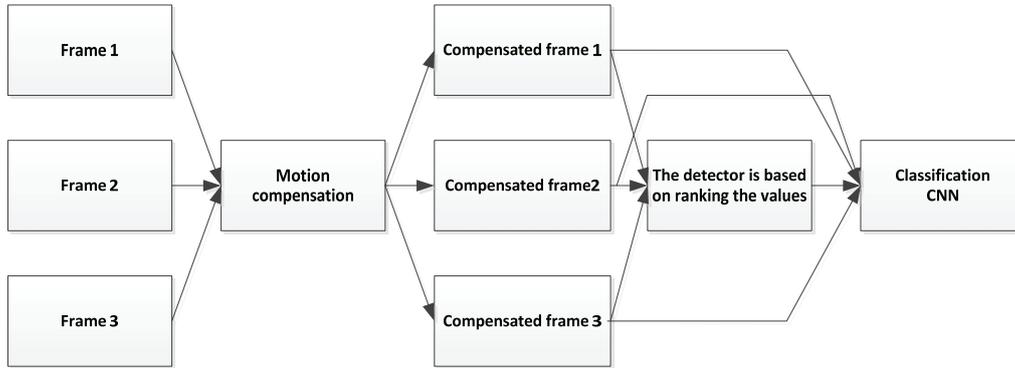

Fig. 1. The block diagram of defect detection on the video sequence.

Evaluation of motion compensation allows determining the direction of movement of the small blocks on the frame of the video sequence by finding them on adjacent frames. In this work as a metric to determine the similarity between the current and the desired block is used calculation evaluation of the sum of absolute differences (SAD):

$$SAD = \sum\sum \left| B_{current} - B_{required} \right| \rightarrow \min \qquad (2)$$

where $B_{current}$ - unit for which you want to estimate the motion vector, $B_{required}$ - a unit with a prospective new position.

This operation is schematically shown in figure 2.

a)   b)   c)
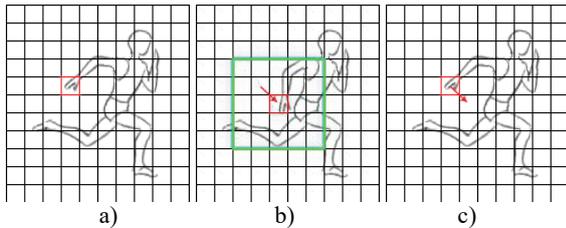
Fig. 2. Motion compensation a) current frame, b) - search for the similar block on the adjacent frame) - compensated frame.

Preprocessing is an important step when damage is detected. It allows to reduce the number of false alarms the detectors and to reduce the computational complexity of the algorithm since performance is an important indicator of the problem of damage detection for the dynamic signal.

As used pre-processing method proposed in [5]. The operation of this detector assumes that the brightness values in damaged areas of the frame that differs significantly from the brightness values in the same areas in adjacent frames. To determine the abnormal values, we use the ranking of pixel values in ascending order. This operation is schematically shown in Figure 3.

In the first stage of the method for each pixel of the current frame $Y^{(k)}$, six are selected spatially close pixels p_r, on adjacent frames $Y^{(k-1)}$ and $Y^{(k+1)}$. Then the selected pixels $p_r, r \in [1,6]$ ranked, where they are assigned sequence numbers d, moreover $d_1 \leq d_2 \leq ... \leq d_6$. Then the decision rule can be written as the following expression:

$$SROD = \begin{cases} d_1 - Y_{i,j}^{(k)}(R_{i,j}) & if \ d_1 - Y_{i,j}^{(k)}(R_{i,j}) > 0 \\ Y_{i,j}^{(k)}(R_{i,j}) - d_6 & if \ Y_{i,j}^{(k)}(R_{i,j}) - d_6 > 0 \\ 0 & else \end{cases} \quad (3)$$

The condition under which the pixel value is assumed to be abnormal is written as:

$$L_{SROD}(R_{i,j}) = \begin{cases} 1 & if \quad SROD > T \\ 0 & else \end{cases} \qquad where \ T \geq 0 \ (4)$$

where T - prior specified threshold value.

a)   b)
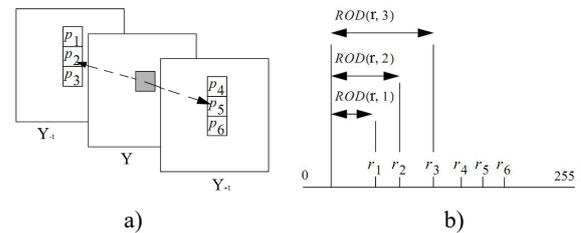
Fig. 3. A method of ranking the pixel values a) selection of pixel values spatially close to the alleged damage, b) – ranking of the selected pixel values.

The incorrect use of compensated frames what happens when the foreground object is moving or incorrectly chosen threshold value can lead to excessive false alarms. To reduce the number of false positives caused by these circumstances, and to reduce the dependence on the choice of the threshold value method is used for post-processing, based on machine learning.

The task of the classifier is to determine the class to which belongs each pixel of the marked unit on the mask obtained in the previous step.

The choice of convolutional neural networks as a post-processing based on two conditions: the first is attributed to the high-performance classification of two-dimensional signals, which is confirmed by modern studies [6], the second condition is that high speed, due to the lack of necessity for the computation of the special handles, which can take a considerable period.

To build a convolutional neural network this work uses a Google library the Flow Tensor [7]. This network has the convenient form of specifying variables and create the overall architecture has an interactive Visualizer logs, high speed of learning and classification using only CPU resources. Also, this convolutional neural network allows the use of an improved version of the method of error back-propagation "Adam optimiser" which allows you to achieve faster global extreme of a function.

At the initial stage of training, the neural network kernel in the convolutional layers is set to random values.

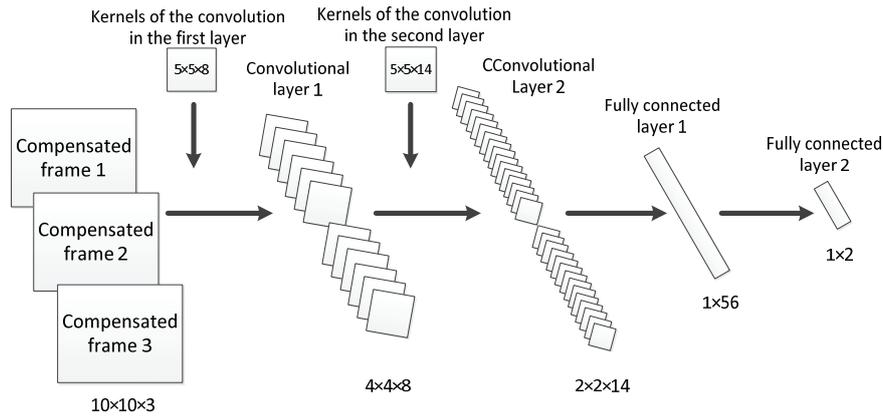The architecture of the convolutional neural network is presented in figure 4.

Fig. 4. The architecture of the convolutional neural network

Then for each input image are computed maps features by convolution of the image and a certain kernel. In a simplified form the function of a convolution can be represented as:

$$x_j^l = f\left(\sum_i x_i^{l-1} * k_j^l + b_j^l\right) \qquad (5)$$

where x_j^l - the map features j (for a layer l), k_j^l - the convolution kernel size 10×10 pixels, $f()$ - the activation function of the hidden layer, b_j^l - the shear factor, x_i^{l-1} - the map of features of the previous layer.

In the output layer as activation function the linear rectification [8]:

$$f(x) = \max(0, x) \qquad (6)$$

where x - the input card features.

As the activation function in layer two fully connected used "softmax" function, which allows determining the probability that a classified pixel between two classes. The activation function is written as:

$$y_j = \frac{e^{z_j}}{\sum_{i=1}^n e^{z_i}} \qquad (7)$$

where $n = 2$ and z - the output value for the current patch and the corresponding function cross entropy:

$$C = -\frac{1}{m}\left[\sum_{l=1}^m \sum_{j=1}^n t_j \log \frac{e^{z_j}}{\sum_{i=1}^n e^{z_i}}\right] \qquad (8)$$

where m the total number of patches, t_j - the indicator function.

Cross-entropy is used as an evaluation of loss of function in the method of backpropagation of error. We use a modification of the gradient descent method proposed in [9].

The essence of the modification is the addition of regularization parameter:

$$\Delta W_t = \eta(g_t + \rho \cdot W_{t-1}) + \mu \Delta W_{t-1}, \qquad (9)$$

$$g_t = \frac{D_t}{1-\beta} \cdot \sqrt{\frac{1-\alpha}{S_t}}, \qquad (10)$$

$$D_t = \beta \cdot D_{t-1} + (1-\beta) \cdot \nabla E_t; \quad D_0 = 0, \qquad (11)$$

$$S_t = \alpha \cdot S_{t-1} + (1-\alpha) \cdot \nabla E_t^2; \quad S_0 = 0, \qquad (12)$$

where ρ - regularization coefficient, α and β fixed values, equal 0.999 and 0.9.

Using this option helps to avoid overfitting by setting a "penalty" for excessive growth of the weights.

3 Experimental results

To validate the proposed method for automatic detection of lesions were used, the video sequence having damages obtained naturally. We use several frames taken from video sequence "Mistake engineer Cochin". Figure 5 shows the first test frame.
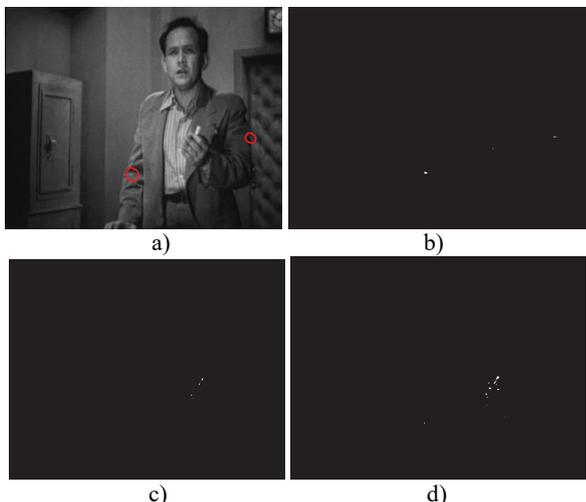
Fig. 5. The results of the methods for the detection of defects a) original frame, b) developed a method, c) - ROD detector d) SDI detector.

Results for the first test frame show that the proposed method allowed to detect more damage, in the absence of false alarms due to the possibility of using lower thresholds than in the known methods. The second test frame is shown in figure 6.

From these results, it can be concluded that known methods of damage detection are not always possible to detect interference with a low contrast compared to the background.
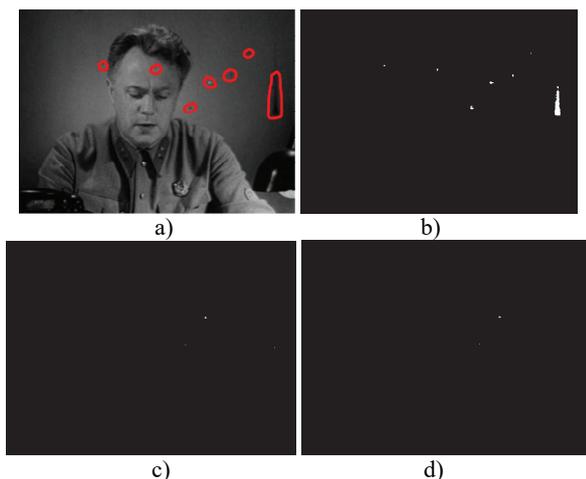
Fig. 6. The results of the methods for the detection of defects a) original frame, b) developed a method, c) ROD detector d) SDI detector.

The third test frame is shown in figure 7.
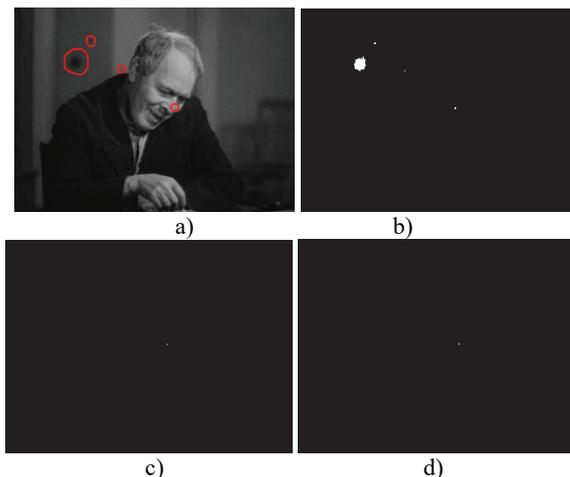
Fig. 7. The results of the methods for the detection of defects a) original frame, b) proposed method, c) ROD detector d) SDI detector.

Analysis of the obtained results indicates the high efficiency of the proposed method. The additional use of machine learning as a post-processing significantly reduces the error of the second kind. The error of the first kind, in most cases, the proposed method is lower than the known techniques of detecting the damage. The main part of the false alarms associated with the difficulty of motion compensation moving objects. This is because the object with fast motion is smeared, and find similar blocks in adjacent frames becomes almost impossible.

4 Conclusion

By conducted research we can draw the following conclusions:

- The method of defect detection on a video sequence showed high efficiency in the problem of damage detection with low contrast and complex shape.

- Use for additional classification on the post-processing made it possible to lower the threshold, which in turn led to a reduction in the probability of error of the second kind.

- Additional using classification based on machine learning reduced false alarm caused mistakes of the methods of motion compensation.

Reference

1. Tenze L., Ramponi G., and Carrato S., ICIP, Canada, pp. 660–663, 2000.

2. Joyeux L., Boukir S., Besserer B., and Buisson O., IVC journal, Elsevier, Vol 19 (8), pp. 503–516, May 2001.

3. Kokaram A.C., Springer Verlag, 1998.

4. Nadenau M. J. and Mitra S. K., Time–Varying Image Processing and Moving Object Recognition, Elsevier, pp. 27–35, 1997.

5. Roosmalen van P.M.B., Biemond J., and Lagendijk R. L., Signal Processing for Multimedia, 1999

6. Z. Lei, Y. Fan, D. Yimin, and J. Z. Ying, ICIP, 2016 IEEE International Conference, 2016.

7. Y. Wu, M. Schuster, Z. Chen, M. Wolfgang, M. Krikun, and etc., 2016, arXiv:1609.08144. [Online]. Available: https://arxiv.org/abs/1609.08144

8. A. Krizhevsky, I. Sutskever, and G. Hinton, Advances in Neural Information Processing Systems, p. 10971105, 2012.

9. P. D. Kingma and J. Ba, 2014, arXiv:1412.6980. [Online]. Available: https://arxiv.org/abs/1412.6980