ANIMAL DETECTION USING A SERIES OF IMAGES UNDER COMPLEX SHOOTING CONDITIONS

A. G. Zotin [1], A. V. Proskurin [1,] *

[1] Institute of Computer Science and Telecommunications, Reshetnev Siberian State University of Science and Technology, Krasnoyarsk, Russian Federation - zotin@sibsau.ru, proskurin.av.wof@gmail.com

Commission II, WG II/5

KEY WORDS: Animal Detection, Background Modeling, Camera Traps, MSR Algorithm

ABSTRACT:

Camera traps providing enormous number of images during a season help to observe remotely animals in the wild. However, analysis of such image collection manually is impossible. In this research, we develop a method for automatic animal detection based on background modeling of scene under complex shooting. First, we design a fast algorithm for image selection without motions. Second, the images are processed by modified Multi-Scale Retinex algorithm in order to align uneven illumination. Finally, background is subtracted from incoming image using adaptive threshold. A threshold value is adjusted by saliency map, which is calculated using pyramid consisting of the original image and images modified by MSR algorithm. Proposed method allows to achieve high estimators of animals detection.

1. INTRODUCTION

In the last two decades, wildlife monitoring is performed using camera traps, which are cameras with infrared motion sensors (O'Connell et al., 2011). Camera trap creates a sequence of images with time lapse of several seconds between images after motion sensor is triggered. Since the range of the sensor is relatively small, the camera traps also are set to capture images at regular intervals. That leads to the generation of a huge amount of images, up to 75% of which do not contain animals (Swanson et al., 2015; Swinnen et al., 2014), and significantly increases labor costs. Three approaches are used to automatically detect animals in images and decrease manual work.

The first approach uses a deep neural network to divide images into two classes: with and without animals (Norouzzadeh et al., 2018). This approach demonstrates high accuracy, but requires about a million of manually classified images that make this approach difficult to use widely. It is also hard to obtain information about location of animals.

The second approach utilizes frame differencing to find motion areas. Part of these areas can be detected due to changes in light and swaying of vegetations, and also correspond to the position of animals in the previous image. A trained classifier is used to distinguish between such areas and areas with animals (Castelblanco et al., 2017). This approach allows to localize the animal in the image, but still requires a large number of images to train classifier.

The third approach for animal detection is based on background modeling of scene using a series of images. For each pixel of the current image, its difference from the background model is estimated. If the difference is less than a threshold, then this pixel is considered part of the animal. This approach allows to detect and localize an animal without use of classifiers, but construction of the background model is difficult due to such factors as swaying trees and leaves, changing sunlight with dimming of some places and lightening others, etc. Additionally, the very presence of animals in the images can distort the generated background model.

In this paper, we propose a method for detection of animals based on background modeling with adaptive threshold segmentation. The threshold adjustment is based on a saliency map calculated for different levels of pyramid consisting of the original image and images modified by Multi-Scale Retinex (MSR) algorithm. The generation of the background model, the preprocessing of the analyzed images, and the calculation of the saliency map are performed using the modified MSR algorithm. Also the preceding selection of images without motion is performed for the generation of the background model.

The rest of the paper is organized as follows. In Section 2, the related works in background subtraction are briefly reviewed. The description of our proposed framework is presented in Section 3. Some empirical results and discussions are demonstrated in Section 4. Finally, conclusions are drawn in Section 5.

2. RELATED WORKS

Over the past two decades, a large number of methods have been proposed for motion detection, some of which are discussed in reviews (Bouwmans et al., 2010; Bouwmans, 2014; Bouwmans et al., 2019). One of the basic approaches to generate a background model is based on the assumption that all values taken by a pixel at a particular point are generated by a random variable with a certain probability density function. In this case, it is enough to estimate the parameters of this function to determine whether the new pixel value belongs to the

* Corresponding author

https://doi.org/10.5194/isprs-archives-XLII-2-W12-249-2019 | © Authors 2019. CC BY 4.0 License.

background or not. In most cases, it is assumed that the probability density function is Gaussian, and two parameters are adaptively evaluated: the mean value and the variance (Wren et al., 1997). The mean value is considered as a background model that is evaluated recursively:

$$B_t = \alpha B_{t-1} + (1-\alpha) I_t \tag{1}$$

where B_t, I_t = background model and image at time t, respectively
α = constant (usually equal to 0.05)

Variance is used to determine whether a pixel belongs to the background (in this case, the difference between B_{t-1} and I_t is less than the threshold) or not. This method is simple and efficient in calculations.

Methods based on Σ-Δ (sigma-delta) motion detection (Manzanera, 2007) are also quite popular due to the high speed of calculations. The essence of the methods lies in the recursive nonlinear adjustment of the background model by increment (decrement) at a constant value if it is smaller (larger) than the current image. Often, the values -1, 0, 1 are used as valid constants.

The methods listed above are fast, simple to develop, and demonstrate satisfactory results in good indoor shooting conditions. However, in the case of dynamic background and high noise, more complex methods are required. Among them, the most popular is based on the use of a Gaussian Mixture Model (GMM) (Stauffer et al., 1999). At each pixel GMM estimate the mean and variance of a number of Gaussians which also have weights indicating their persistence. If observed value matches a Gaussian, its weight increases and parameters are updated using running average. If the sum of matched Gaussians' weights is above a given threshold, the pixel is classified as background.

Other methods are developed based on GMM. In (Heikkilä et al., 2004) it is proposed to model each pixel as a group of adaptive local binary pattern histograms that are calculated over a circular region around the pixel to leverage neighborhood information. In paper (Chen et al., 2014) minimum spanning tree aggregation technique is used to integrate pixel-based and region-based background models to suppress the noisy background estimates obtained from GMM. Algorithm FTSG (Wang et al., 2014b) uses GMM augmented by flux tensor motion detection which significantly improves the quality of stopped objects detection.

One of the drawbacks of GMM methods is the assumption that the background is seen much more often than moving objects, and its dispersion is much lower. In relation with this, the GMM methods are difficult to use for images captured by camera traps. Another approach to build a background model is based on keeping the last 100 values for each pixel as examples of background (Wang et al., 2007). The new value of the pixel refers to the background if it corresponds to the majority of the kept values. The new found background value is saved as a new example, and replaces the oldest. In the paper (Van Droogenbroeck et al., 2014), it is proposed to replace one of the kept values randomly after classifying a pixel as a background. This increased the accuracy of detection and reduced the number of kept examples to 20.

Another approach to background modeling was proposed in (Kim et al., 2005). Authors represent each pixel as a codebook containing a set of statistics, such as minimum, maximum and average values, frequency or number of occurrences, etc. These statistics are used to assess whether a new pixel value belongs to the background by analogy with GMM. Algorithm (Wang et al., 2014a) uses a small number of code words obtained by clustering observed values using running average and efficacy counters. Algorithm proposed in (St-Charles et al., 2015) uses mixed codebooks based on color and local binary features and regulates its own internal parameters using feedback mechanisms.

In all of the above methods, no attempt is made to solve the problem of changes in illumination in the analyzed images before motion detection. Various methods can be used to solve the problem of uneven illumination. These methods can be divided into the following classes: intensity transformations, histogram transformations and Retinex method. Intensity transformations use a wide set of specific functions such as linear function, logarithmic function, or power function, including γ-correction (Gonzalez and Woods, 2008). A histogram transformations approach modifies local histograms in dark and bright areas according to a desired shape (Raju et al., 2013). The Retinex is the most advanced technique which is able to simulate some of the adaptation mechanism of the human vision system under complex luminance conditions. Many image enhancement techniques based on Retinex theory have been reported in the literature such as Single Scale Retinex (SSR) algorithm, Multi-Scale Retinex (MSR) algorithm, Multi-Scale Retinex with Colour Restoration (MSRCR) and their modifications (Liu et al., 2016; Liao et al., 2017; Zotin, 2018).

3. PROPOSED METHOD

Our proposed method of animal detection and localization is based on a background modeling of scene under complex shooting conditions. Fast algorithm for image selection without motion is designed as the first step of background image construction. The idea is that images without motion less likely to contain animals, and thus constructed background image would be more accurate. Since uneven illumination has a great influence on the background model, we use MSR algorithm, which utilizes wavelet transform to speed up the calculations (Zotin, 2018). In this algorithm, artifacts such as halo or stairs may appear in an obtained image due to the large values in the high-frequency components. We modified the MSR algorithm by adding two coefficients, which change the process of high-frequency component computation. For adaptive thresholding, we apply a saliency map of image, which allows to adjust threshold for noise suppression. Saliency map of image is calculated using pyramid consisting of the original image and images modified by MSR algorithm.

The workflow of proposed method is presented in Figure 1. It can be divided into six steps. First, images without motion (empty images) are selected using rough motion maps. Second step is images preprocessing by modified MSR algorithm. Third step is generation of the background image description. Fourth step is formation of saliency map for each image in series. Fifth step is background subtraction from the analyzed image with saliency map usage for threshold calculation. Sixth step is post-processing of a background subtraction result. In the following subsections, we describe each step in details.

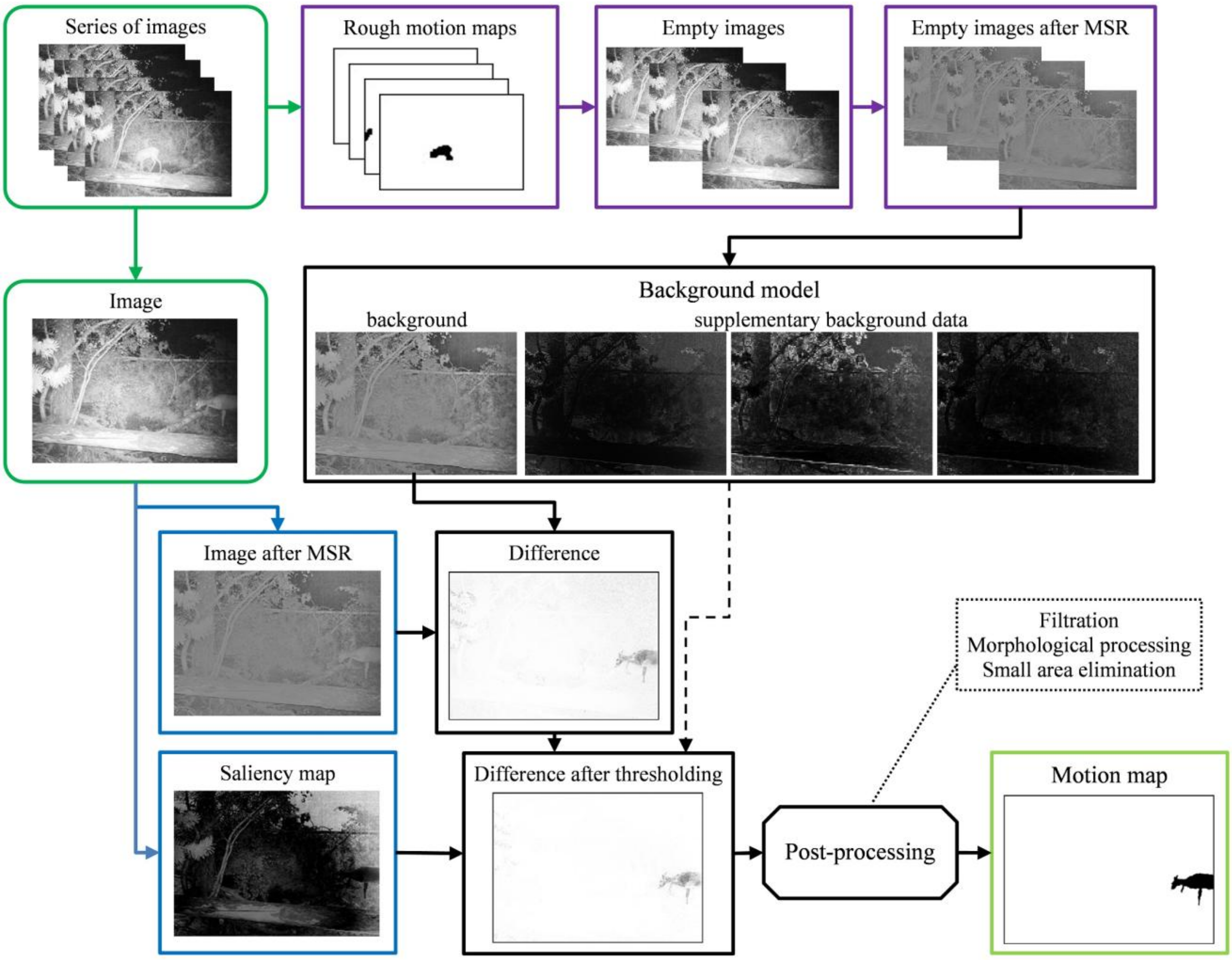

Figure 1. A scheme of the proposed method. Motion maps and differences are inverted.

3.1 Fast Selection of Images without Motion

The proposed algorithm for the fast selection of empty images allows to detect the movement of animals based on information about the difference of three consecutive images. This information is used to filter the areas corresponding to the position of the animals in the previous image. Also images are represented as a set of blocks to reduce computational costs and the number of background motion detections. Additionally, it is proposed to normalize the pixel values of the block based on the average pixel value in the vicinity of this block to solve the problem of false detection due to changes in light.

In the first stage, all images are converted into grayscale, and for each image (I) an integral image (Ii) is calculated. The value of integral image pixel with coordinates (ix, iy) is calculated as follows (Bay et al., 2008):

$$Ii(ix,iy)=\sum_{x'=1}^{ix}\sum_{y'=1}^{iy}I(x',y') \qquad (2)$$

Integral images allows to get the sum of the pixel values of a rectangular area of arbitrary size for a fixed time, which is used later to reduce the computation time.

At the second stage, the images are divided by a grid into square blocks of size $s \times s$ pixels. For each image a map of blocks BM is created, where the value of pixel with coordinates (x, y) is equal to the average pixel value of the corresponding block ($B_{x,y}$) of the original image:

$$BM(x,y)=\frac{1}{s^2}\sum_{x',y'\in B_{x,y}}I(x',y') \qquad (3)$$

Since the movement of the branches in the background occurs with limited amplitude, all oscillations will fall into one block with a sufficiently large value of s, which will significantly reduce the number of false detections. It was found experimentally that the best result is achieved by s = 32 for images of 2592 × 1844 pixels. This value is chosen as the main one.

In addition, at this stage the pixel values of the map of blocks are normalized to reduce the negative influence of possible illumination change in two consecutive images. For normalization of each pixel with coordinates (x, y), pixels from the neighborhood ($N_{x,y}$) with a radius of neighborhood of 1 (kernel 3 × 3) are used:

$$BMN(x,y)=\left[BM(x,y)-\frac{1}{9}\sum_{x',y'\in N_{x,y}}BM(x',y')\right]+127 \qquad (4)$$

where BMN = map of blocks with normalized illumination

The idea of the proposed normalization is based on the assumption that within a small region the illumination change affects all the pixels evenly. Thus, the values of the pixels of the central block will remain constant relative to the pixels of the neighboring blocks during shading or lightening of the scene.

At the next stage, the map of differences (SB_t) is computed between two consecutive images t and $t - 1$ by subtracting the corresponding values of blocks:

$$SB_t(x,y) = \left| BMN_t(x,y) - BMN_{t-1}(x,y) \right| \tag{5}$$

The resulting difference is used to build rough motion map (MM_t) using an adaptive threshold based on the difference between three consecutive images:

$$MM_t(x,y) = \begin{cases} 255 & \text{if } SB_t(x,y) > Th_t(x,y) \\ 0 & \text{otherwise} \end{cases}, \tag{6}$$

$$Th_t(x,y) = \left[\frac{1}{9} \sum_{x',y' \in N_{x,y}} \frac{SB_t(x',y') + SB_{t-1}(x',y')}{2} + \right.$$
$$\left. + \frac{1}{w \cdot h} \sum_{x'=1}^{w} \sum_{y'=1}^{h} \frac{SB_t(x',y') + SB_{t-1}(x',y')}{2} + 1 \right] \cdot mul$$

where w, h = width and height of maps of differences
mul = threshold multiplication factor, chosen the same for all series of images.

The resulting motion map can contain the area (called “ghost”) corresponding to the position of the animal in the previous image. This area is filtered using the motion map of the previous image:

$$fMM_t(x,y) = \begin{cases} 0 & \begin{array}{l} \text{if } MM_t(x,y) = 255 \text{ and} \\ MM_{t-1}(x,y) = 255 \end{array} \\ MM_t(x,y) & \text{otherwise} \end{cases} \tag{7}$$

where fMM_t = motion map after filtering of ghosts

Since the calculation of a fMM_t requires two previous images, for the first image we consider the third and second images as previous ones.

At the last stage, the obtained motion maps are processed using morphological operators of erosion and dilation. If all pixels of fMM_t are equal to 0, then we select corresponding image I_t for the next step of our method. The number of selected images can vary or be fixed. If there are a few of empty images, then images with animals can be used excluding detected motion areas.

3.2 Illumination Enhancement

Uneven illumination has a great influence on the background model formation, especially if the images were taken in the morning, evening and at night. In this regard, a modified MSR algorithm is used. During the MSR illumination enhancement there is the likelihood of artifacts, which occur mainly in fragments with high local contrast in the original image. These artifacts can adversely affect the result of the background model formation and lead to false positives at the stage of animal detection. Artifacts are characterized by high values of high-frequency wavelet components. To reduce the number of artifacts we use two coefficients to correct the intensity of the high-frequency components. The first (k_{div}) is responsible for uniform correction over the entire range of values, and the second (k_h) for linear correction in case of exceeding the threshold of local contrast (T_H). In order to optimize the computational process during the software implementation the Look-Up Table is used to calculate detail component of wavelet transform in accordance with equation 8. In the case of processing a one-dimensional discrete signal $S=\{s_j\}_{j\in Z}$, the detail (H) component will be formed depending on the difference of neighbouring pixels (S_{2j}–S_{2j+1}) using equation 9. It should be noted that the inverse wavelet transform remains unchanged.

$$H_{LUT}(i) = \begin{cases} \dfrac{i}{k_{div}} & \text{if } \quad i \le T_H \\ \dfrac{(i - T_H) \cdot k_h + T_H}{k_{div}} & \text{if } \quad i > T_H \end{cases} \tag{8}$$

$$H_j = \operatorname{sgn}(S_{2j} - S_{2j+1}) \cdot H_{LUT}\left(\left| S_{2j} - S_{2j+1} \right| \right) \tag{9}$$

The response of the MSR function (R_{MSR}) typically gives both negative and positive values, and the obtained range limits will be arbitrary. Depending on the image in the distribution of the output values, the average value can be shifted relative to zero.

$$R_{MSR}(x,y,\boldsymbol{\sigma}) = \sum_{k=1}^{n} w_k \cdot \log\left(\frac{I_{x,y}}{I_{x,y} * G_{x,y}(\sigma_k)} \right) \tag{10}$$

where $I_{x,y}$ = the intensity value at pixel (x,y)
n = the number of scales
$G_{x,y}(\sigma)$ = Gaussian
$*$ = convolution operator
$\boldsymbol{\sigma} = \{\sigma_1, \sigma_2, \dots \sigma_n\}$ = set of the blurring coefficients

To obtain the images with the desirable mean brightness value of output image it was decided to use the mechanics based on Autolevels algorithm taking into account the cut-off of the boundary values with the adaptive adjustment of the range based on the boundary thresholds and desired range size. Thus, the calculation of output image brightness value (I_{MSR}) of a pixel with the desired range of visualization (I_{TR}) conducted by using equation 11. The values outside of desirable range are clipped.

$$I_{MSR}(x,y) = Cl\left(\frac{R_{MSR}(x,y,\boldsymbol{\sigma}) - R_{avg}}{(R_{\max} - R_{\min}) \cdot Pr} \cdot I_{TR} + k_{ofs} \right) \tag{11}$$

where $R_{MSR}(x,y)$ = MSR function response of pixel (x,y)
R_{avg} = average value of MSR function
R_{min} = minimum value of MSR function
R_{max} = maximum value of MSR function
k_{ofs} = brightness offset
$Cl()$ = the cut-off function for values outside the desirable range

Figure 2 shows the example of the illumination enhancement of fragment with high local contrast using the MSR algorithms. In the figure, the appearance of artifacts in some places in the case of using the processing without modification of wavelet transform can be seen. Figures 3, 4 demonstrate results of uneven illumination correction for subset of three images taken during day and night correspondingly.

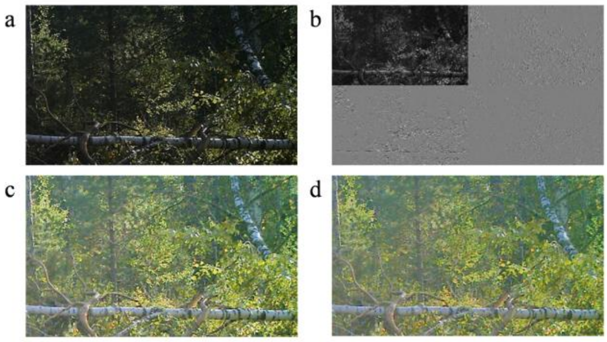

Figure 2. Example of the illumination enhancement of area with high local contrast: a) original image, b) wavelet transform, c) MSR result, d) MSR with wavelet transform modification

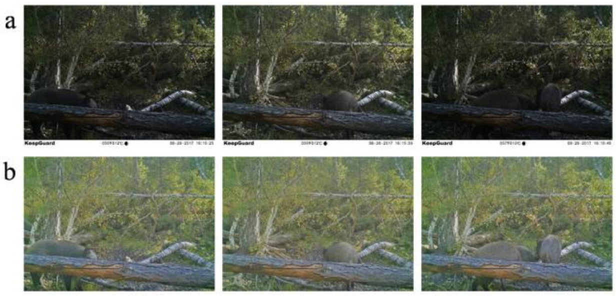

Figure 3. Example of illumination enhancement of day images subset: a) original images, b) processed images

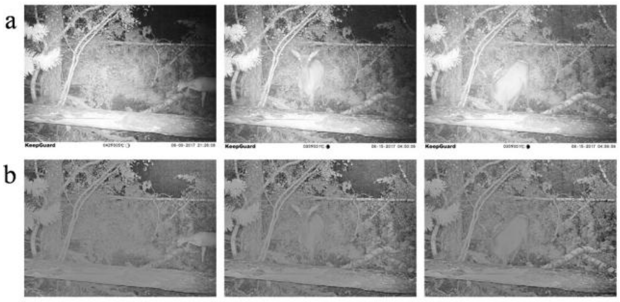

Figure 4. Example of illumination enhancement of night images subset: a) original images, b) processed images

3.3 Generation of the Background Image

Processed images by MSR algorithm are utilized to create background image. To generate a description of background model we use information about brightness, color and such statistics information as pixel wise standard deviation (PSD), block based standard deviation (BBSD) and its variance. Depending on characteristics of input images, i.e. captured at nighttime or daytime, mean value is calculated as grayscale image or color image correspondingly.

Examples of background model maps are shown in figure 5. For better visual interpretation of maps containing information of PSD and mean value of BBSD multiplication by 5 is used, and for map of BBSD variance the values are multiplied by 20.

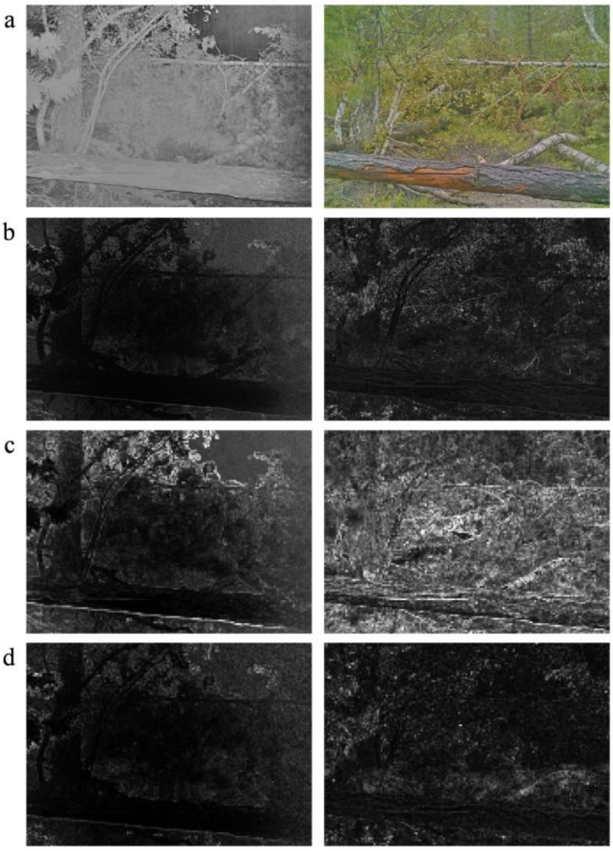

Figure 5. Examples of background model maps: a) mean value, b) pixel standard deviation, c) mean of block based standard deviation, d) variance of block based standard deviation

Calculation of basic characteristics of background map conducted in following way. Firstly, for each pixel the mean value of brightness ($E1_{x,y}$) is calculated. Then, a calculation of brightness standard deviation ($Ds_{x,y}$) in pixels among the set of images are conducted. After this, we calculate resulting maps of mean brightness (color) $E_{x,y}$ and PSD of brightness $D_{x,\mathrm{y}}$ including only pixels, which value is less than $2{\cdot}Ds_{x,y}$. Apart from these maps, we generate additional maps containing mean value of BBSD (BSD_b) and variance of BBSD ($VBSD_b$). In our implementation the default block size is 16×16 pixels, however blocks with other sizes can be used.

3.4 Formation of Saliency Maps

In the images captured at night, there are dark areas with a lot of noise arising from the characteristics of the backlight or flash of camera trap. Saliency maps allow to easily identify them. Also saliency map allow to determine areas with strong light in the daytime if image have uneven illumination.

The formation of the saliency maps in proposed method is based on the brightness and chromatic component (Favorskaya et al., 2016). Thus, color models CIE Lab, YUV, YCbCr, and similar color models can be used.

We propose to use the pyramid of saliency maps with 2-3 levels. Thus level 1 saliency map calculated by using the original image, level 2 is the map formed by using result of applying modified MSR-function to original image. Level 3 of saliency map can be obtained by using image processed by modified MSR-function two times, or set different parameters of MSR function. Numbers of used levels are decided based on characteristics of original image. The separated saliency maps are normalized to fit the range [0…255] with following merging into a single saliency map.

Examples of generated saliency maps for daytime and night time images are presented in figure 6. As can be seen from examples, saliency maps with usage of two levels gives more detailed features of saliency objects in observed scene.

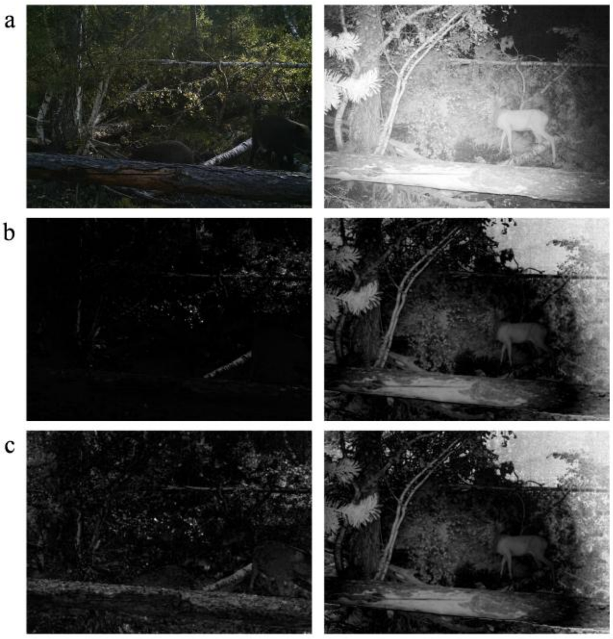

Figure 6. Example of saliency maps: a) original images, b) saliency map level 1, c) saliency map with 2 levels

3.5 Background Subtraction and Post-processing

Background subtraction is the last step on which mask with the animals or their observed parts are formed. During this step the difference between analysing image and background model is calculated according to equation 12.

$$Df_{x,y} = \left| Ia_{x,y} - E_{x,y} \right| \tag{12}$$

where $Ia_{x,\mathrm{y}}$ = value of pixel (x,y) of pre-processed image
$E_{x,\mathrm{y}}$ = value of mean brightness map

Obtained map of difference is filtered using threshold ($Ts_{x,\mathrm{y}}$) which is defined for each pixel (equation 13). The threshold is based on values of saliency map, standard deviation map and support map of animal positioning probability (Smap). Smap sets the value of the k_{HS} coefficient and aims to reduce the likelihood of false positives. The introduction of this map allows to set the coefficient values taking into account animal observation areas. Thus, in most cases, the probability of an animal appearing in the upper part of the image is small, since in most cases treetops, sky or distant objects are observed. It is possible to adjust the coefficient coefficients in such way that the coefficient value will be high in the upper region and in the lowest one, but small in the central region and in region close to the lower part of image.

$$Ts_{x,y} = D_{x,y}\left(\frac{Sal_{x,y}}{100} + k_{HS} + k_{Base} + k_{BSD} \right), \tag{13}$$

$$k_{BSD} = \begin{cases} \dfrac{dBSD_b}{BSD_b} & \text{if } dBSD < VBSD_b \cdot k_{sdf} \\ \dfrac{2 \cdot dBSD_b}{BSD_b} & \text{otherwise} \end{cases},$$

$$dBSD_b = \left| iBSD_b - BSD_b \right|$$

where $Sal_{x,y}$ = value of saliency map of analysing image at pixel (x,y)
k_{Base} = coefficient defined by user, in range [0..1]
k_{BSD} = coefficient based on the variance of BBSD
$iBSD_b$ = value of BBSD of analysed image
k_{sdf} = coefficient of maximal variance of BBSD

Utilizing difference between analyzed image and background model, and filtering threshold a map (*Fmap*) are formed, which afterwards changed to binary form (*Bmap*).

$$Fmap_{x,y} = \begin{cases} Df_{x,y} & \text{if } Df_{x,y} > Ts_{x,y} \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

In order to find out the regions with possible appearance of animal and reducing noise, a post processing is used according to expression 15.

$$NBmap_{x,y} = \begin{cases} 1 & \text{if } \sum\limits_{x',y' \in M_{x,y}} Bmap_{x',y'} \geq T_f \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

where $NBmap_{x,\mathrm{y}}$ = new value of binary map at pixel (x,y)
$Bmap_{x,\mathrm{y}}$ = old value of binary map at pixel (x,y)
$M_{x,\mathrm{y}}$ = structural element centred around pixel (x,y)
T_f = threshold defining effect of filters

If T_f equals 1 then result of processing will be similar to morphological operation dilation. It can be used to fill the holes (missing pixels) in a continuous object and makes an object to grow by size correspondently to structural element of filter. In case when T_f equal size of structural element the result of filtration will be similar to morphological operation erosion. The erosion operation is complement of the dilation operation in context with the operation effect. That is erosion operation causes object to lose its size. Thus, the erosion operation removes those structures which are lesser in size than that of the structuring element. So it can be used to remove the noise and eliminate small objects from resulting binary mask. In order cases filter produce more smooth edges of possible object with filing or removing pixel.

Apart from filtering and morphological processing, different approaches can be used to eliminate unwanted small fragments. For example, fragments with small size can be removed and objects areas at a short distance can be merged.

4. EXPERIMENTS AND RESULTS

In order to verify the validity of our method, we experimented with sets of images captured by camera traps in different regions of “Ergaki” natural park. There are 5 camera traps, which can capture images in daytime color and nighttime formats with different spatial resolution (Figure 7). Camera 1 resolution is

2592×1944 pixels (Figure 7a), camera 2 resolution is 3264×2448 pixels (Figure 7b), resolution of cameras 3 and 4 is 4000×3000 pixels (Figures 7c and 7d, respectively), and camera 5 resolution is 1920×1080 pixels (Figure 7e). We extracted about 200 daytime and 200 nighttime images from each camera, and created set with overall size of 2071 images.

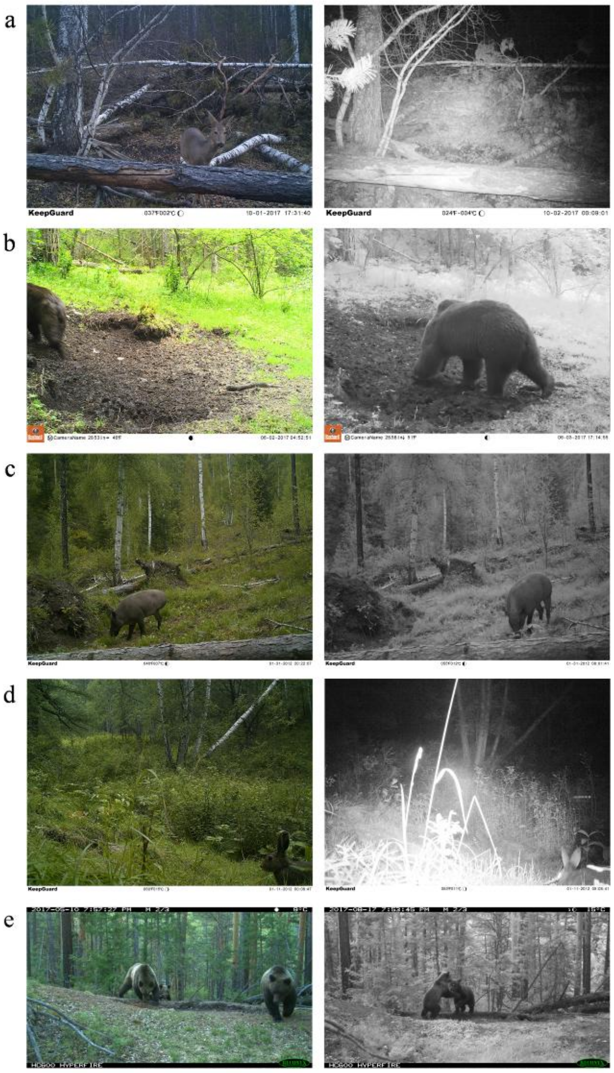

Figure 7. Examples of experimental images captured by different camera traps at day and night

For each image ground truth were manually created in the form of binary masks segmenting the background and foreground pixels. These masks are compared with motion maps created by the proposed method. To evaluate the effectiveness, the precision, recall, and F-measure are used.

$$F-measure=\frac{2\cdot Precision\cdot Recall}{Precision+Recall} \tag{16}$$

Four versions of the method were compared: in the first version (M-EP), the first two steps (selecting of empty images and preprocessing images by modified MSR algorithm) of the method were excluded; in the second (M-E), step 1 was omitted; in the third (M-P), step 2 was omitted; in the fourth (M all), all steps of the method were performed. All versions of the method were evaluated separately for day and night images in addition to evaluation over the entire set of images. The results of evaluation are shown in tables 1, 2 and 3.

Version	Day	Night	Day + Night
M-EP	0.716	0.693	0.704
M-E	0.817	0.783	0.800
M-P	0.803	0.821	0.812
M all	0.881	0.925	0.903

Table 1. Comparison of animal detection by *precision* among four versions of the proposed method

Version	Day	Night	Day + Night
M-EP	0.917	0.844	0.880
M-E	0.892	0.877	0.884
M-P	0.875	0.889	0.882
M all	0.851	0.895	0.873

Table 2. Comparison of animal detection by *recall* among four versions of the proposed method

Version	Day	Night	Day + Night
M-EP	0.804	0.761	0.782
M-E	0.852	0.827	0.840
M-P	0.837	0.853	0.845
M all	0.865	0.909	0.887

Table 3. Comparison of animal detection by *F-measure* among four versions of the proposed method

It can be seen, that using empty image selection and illumination enhancement by modified MSR algorithm as preliminary steps of background modeling increase precision by 20% and F-measure by 10%. Examples of created motion maps are shown in Figure 8.

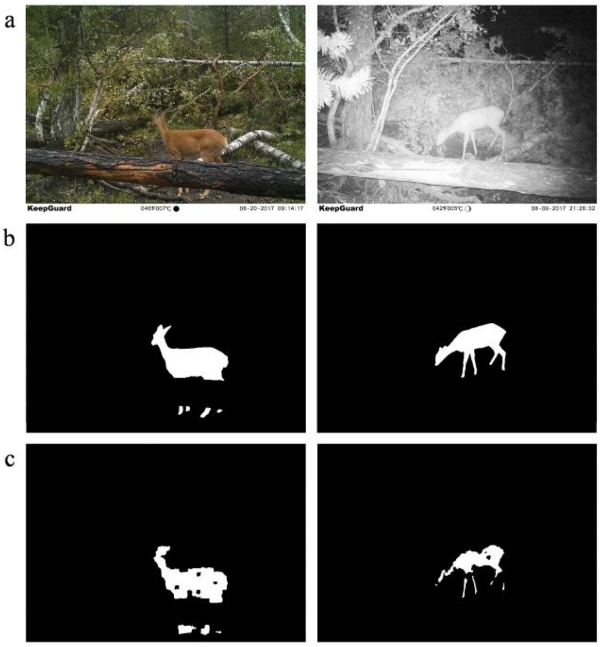

Figure 8. Example of animal detection using proposed method: a) original images, b) ground truth, c) motion maps

5. CONCLUSION

In this paper, we proposed the method of animal detection based on a background modeling. In the first step, images without motion are selected for creation of background model. For this purpose fast algorithm of motion detection based on information about difference of three consecutive images was developed. The introduction of this step prevents usage of the foreground for creation of the background model and increase precision of animal detection. Since uneven illumination also greatly influence on the created background model, in the second step we applied modified MSR algorithm to all images. Usage of these two steps increased precision of animal detection by 20% and F-measure by 10%. Additionally, we use adaptive threshold, which calculated based on saliency map, that allows to suppress noise arising from the characteristics of the backlight or flash of camera traps. Saliency map of image is calculated using pyramid consisting of the original image and images modified by MSR algorithm. For experiments, 2,071 images captured by camera traps at daytime and nighttime with different shooting condition were used. The method shows the precision of animal detection 0.903 and F-measure 0.887.

ACKNOWLEDGEMENTS

The reported study was funded by Russian Foundation for Basic Research, Government of Krasnoyarsk Territory, Krasnoyarsk Regional Fund of Science, to the research project 18-47-240001.

REFERENCES

Bay, H., Ess, A., Tuytelaars, T., Gool, L.V., 2008. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3), pp. 346-359.

Bouwmans, T., Baf, F. El, Vachon, B., 2010. Statistical background modeling for foreground detection: A survey. *Handbook of Pattern Recognition and Computer Vision*, 4, pp. 181-199.

Bouwmans, T., 2014. Traditional and recent approaches in background modeling for foreground detection: An overview. *Computer Science Review*, 11-12, pp. 31-66.

Bouwmans, T., Garcia-Garcia, B., 2019. Background subtraction in real applications: challenges, current models and future directions. *arXiv preprint*. arxiv.org/abs/1901.03577.

Castelblanco, L.P., Narvaez, C.L., Pulido, A.D., 2017. Methodology for mammal classification in camera trap images. *Proc. SPIE 10341, Ninth International Conference on Machine Vision (ICMV 2016)*, 10341. doi.org/10.1117/12.2268732.

Chen, M., Yang, Q., Li, Q., Wang, G., Yang, M., 2014. Spatiotemporal background subtraction using minimum spanning tree and optical flow. *European Conference on Computer Vision 2014*, 8695, pp. 521-534.

Favorskaya M.N., Buryachenko V.V., 2016. Fast salient object detection in non-stationary video sequences based on spatial saliency maps. Smart Innovation, *Systems and Technologies*, 55, pp. 121-132.

Gonzalez, R.C., Woods, R.E., 2008. *Digital Image Processing. 3rd ed.*, Pearson Prentice Hall, Upper Saddle River, NJ.

Heikkilä, M., Pietikäinen, M., Heikkilä, J., 2004. A texture-based method for detecting moving objects. *British Machine Vision Conference*, pp. 187-196.

Kim, K., Chalidabhongse, T., Harwood, D., Davis, L., 2005. Real-time foreground-background segmentation using codebook model. *Real-Time Imaging*, 11(3), pp. 172-185.

Liao, S., Piao, Y., Li, B., 2017. Low illumination color image enhancement based on improved Retinex, *Proc. SPIE 10605, LIDAR Imaging Detection and Target Recognition 2017*, 10605, doi.org/10.1117/12.2295105.

Liu, H., Sun, X., Han, H., Cao, W., 2016. Low-light video image enhancement based on multiscale Retinex-like algorithm. *2016 Chinese Control and Decision Conference (CCDC)*, pp. 3712-3715.

Manzanera. A., 2007. Σ-Δ background subtraction and the Zipf law. *Progress in Pattern Recognition, Image Analysis and Applications. CIARP 2007. Lecture Notes in Computer Science*, 4756, pp. 42-51.

Norouzzadeh, M.S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M.S., Packer, C., Clune, J., 2018. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences*, 115(25), pp. 5716-5725.

O'Connell, A.F., Nichols, J.D., Karanth, K.U. (Eds.), 2011. *Camera Traps in Animal Ecology*. Springer Japan, Tokyo, 271 p.

Raju, A., Dwarakish, G.S., Reddy, V., 2013. A comparative analysis of histogram equalization based techniques for contrast enhancement and brightness preserving. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 6(5), pp. 353-366.

Stauffer, C., Grimson, E., 1999. Adaptive background mixture models for real-time tracking. *IEEE International Conference on Computer Vision and Pattern Recognition*, 2, pp. 246-252.

St-Charles, P.L., Bilodeau, G.A., Bergevin, R., 2015. A self-adjusting approach to change detection based on background word consensus. *2015 IEEE Winter Conference on Applications of Computer Vision*, pp. 990-997.

Swanson, A.B., Kosmala, M., Lintott, C.J., Simpson, R.J., Smith, A., Packer, C., 2015. Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. *Scientific Data 2*. doi.org/10.1038/sdata.2015.26.

Swinnen, K.R.R., Reijniers, J., Breno, M., Leirs, H., 2014. A novel method to reduce time investment when processing videos from camera trap studies. *PLoS ONE*, 9(6): e98881. doi.org/10.1371/journal.pone.0098881.

Van Droogenbroeck, M., Barnich, O., 2014. ViBe: A Disruptive Method for Background Subtraction. *Background Modeling and Foreground Detection for Video Surveillance*, pp. 7.1-7.23.

Wang, B., Dudek, P., 2014a. A Fast Self-Tuning Background Subtraction Algorithm. *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 401-404.

Wang, H., Suter, D., 2007. A consensus-based method for tracking: Modelling background scenario and foreground appearance. *Pattern Recognition*, 40(3), pp. 1091-1105.

Wang, R., Bunyak, F., Seetharaman, G., Palaniappan, K., 2014b. Static and moving object detection using flux tensor with split Gaussian models. *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 414-418.

Wren, C., Azarbayejani, A., Darrell. T., Pentland, A., 1997. Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), pp. 780-785.

Zotin A., 2018. Fast algorithm of image enhancement based on multi-scale Retinex. *Procedia Computer Science*, 131, pp. 6-14.

Revised March 2019