

## Background extraction method for analysis of natural images captured by camera traps

M. N. Favorskaya<sup>a</sup>, Dr. Sc., Tech., Professor, [orcid.org/0000-0002-2181-0454](https://orcid.org/0000-0002-2181-0454), [favorskaya@sibsau.ru](mailto:favorskaya@sibsau.ru)

V. V. Buryachenko<sup>a</sup>, PhD, Tech., Associate Professor, [orcid.org/0000-0003-1151-1159](https://orcid.org/0000-0003-1151-1159)

<sup>a</sup>Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarsky Rabochy Av., 660037, Krasnoyarsk, Russian Federation

**Introduction:** Automatic detection of animals, particularly birds, on images captured in the wild by camera traps remains an unsolved task due to the shooting and weather conditions. Such observations generate thousands or millions of images which are impossible to analyze manually. Wildlife sanctuaries and national parks normally use cheap camera traps. Their low quality images require careful multifold processing prior to the recognition of animal species. **Purpose:** Developing a background extraction method based on Gaussian mixture model in order to locate an object of interest under any time/season/meteorological conditions. **Results:** We propose a background extraction method based on a modified Gaussian mixture model. The modification uses truncated pixel values (in low bites) to decrease the dependence on the illumination changes or shadows. After that, binary masks are created and processed instead of real intensity values. The proposed method is aimed for background estimation of natural scenes in wildlife sanctuaries and national parks. Structural elements (trunks of growing and/or fallen trees) are considered slowly changeable during the seasons, while other textured areas are simulated by texture patterns corresponding to the current season. Such an approach provides a compact background model of a scene. Also, we consider the influence of the time/season/meteorological attributes of a scene with respect to its restoration ability. The method was tested using a rich dataset of natural images obtained on the territory of Ergaki wildlife sanctuary in Krasnoyarsk Krai, Russia. **Practical relevance:** The application of the modified Gaussian mixture model provides an accuracy of object detection as high as 79–83 % in the daytime and 60–69 % at night, under acceptable meteorological conditions. When the meteorological conditions are bad, the accuracy is 5–8 % lower.

**Keywords** – background subtraction, natural scene, gaussians mixture model, animal detection, background model.

**Citation:** Favorskaya M. N., Buryachenko V. V. Background extraction method for analysis of natural images captured by camera traps. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2018, no. 6, pp. 35–45. doi:10.31799/1684-8853-2018-6-35-45

### Introduction

Monitoring of animals in the wild using camera traps is one of the promising ways for monitoring animal behavior in the wildlife sanctuaries and national parks. Camera traps provide tremendous amount of information capturing any motion in a scene. Some camera traps produce a set of still images of moving object (animal, bird, or human) through 5–8 s, while other devices deliver a short movie. In this article, we deal with a set of still images, which are automatically marked by current date and time. Each camera trap has own stationary position in a predefined place, such as animal trails, watering places, and so forth. The stored amount of such information, for example for a half of a year, can achieve several terabytes from dozens of camera traps that makes impossible to process them manually.

For recognition of animal species or analysis of animal behavior, we need to process the original images sometimes of low quality in such manner that allows us to separate a visual object of interest from cluttered background. The well-known scene background challenges make this task dif-

ficult for solving. Among them, it is worth noting the cluttered background, occlusions, color shadows, moving background (for example, fluttering of leaves or waving trees), illumination changes within a day, flash shooting within a night, season changes, and meteorological impacts. At the same time, a scene remains unchanged principally and it is profitably regarding the computational costs to store a background pattern of a particular scene with a possibility to transform it into another state depending on the time/season/meteorological attributes.

Our contribution is twofold. First, we propose a simplified background extraction method based on the modified Gaussian Mixture Model (GMM). The modification uses the truncated pixel values (in low bites) in order to decrease dependence from illumination changes and shadows with following creation and processing the binary masks instead of real intensity values. The proposed method separates a scene into persistent (trunks of growing and/or fallen trees) and non-persistent (snow, foliage, grass, sky, lake, river, and Earth surface for boreal forests) textured regions. The persistent textured regions serve as the landmarks in any season with

a non-changeable distribution, while the distributions of the non-persistent textured regions are changed respect to a current season. Such approach provides a compact background model of a scene. Second, we consider the influence of the time/season/meteorological attributes of scene respect to restoration ability. Note that we need not to provide a high accuracy of the proposed background models because the goal is to detect the location of the object of interest.

## Related work

Background subtraction method compares the current image with a reference image called background model. However, this method has many disadvantages because of illumination changes, shadows, occlusions, noise, and dynamic background [1]. All these impacts make unreasonable to employ this method in many applications. In such cases, the background extraction algorithms are necessary.

In [2], one can find the detailed survey on traditional and recent background models with the complete classification from basic models to domain transform models. Traditional background models are classified in the following categories:

- basic models (average calculation, median processing, and histogram analysis);
- statistical models (Gaussian models, support vector models, and subspace learning models);
- cluster models ( $K$ -means models, codebook models, and basic sequential clustering);
- neural network models (general regression Neural Network (NN), multivalued NN, competitive NN, dipolar competitive NN, self organizing NN, and growing hierarchical self organizing NN);
- estimation models (Wiener filter, Kalman filter, correntropy filter, and Chebychev filter).

In last decade, appearance of visual content from mobile devices and Internet videos requires the development of background subtraction methods for challenging environments. The recent background models are classified in the following categories:

- advanced statistical background models (mixture models, hybrid models, nonparametric models, and multi-kernels models);
- fuzzy background models (fuzzy background modeling, fuzzy foreground detection, fuzzy background maintenance, fuzzy features, and fuzzy post-processing).
- discriminative subspace learning models (discriminative subspace models and mixed subspace models);
- Robust Principal Components Analysis (RPCA) models (RPCA via principal component pursuit, RPCA via outlier pursuit, RPCA via spar-

sity control, RPCA via sparse corruptions, RPCA via log-sum heuristic recovery, RPCA via iteratively reweighted least squares, Bayesian RPCA, and approximated RPCA);

- subspace tracking (Grassmannian Robust Adaptive Subspace Tracking Algorithm (GRASTA), transformed-GRASTA,  $l_p$ -norm robust online subspace tracking, and Grassmannian online subspace updates with structured-sparsity);

- low rank minimization (contiguous outliers detection, direct robust matrix factorization, direct robust matrix factorization-row, probabilistic robust matrix factorization, and Bayesian robust matrix factorization);

- sparse models (compressive sensing models, structured sparsity, dynamic group sparsity, dictionary learning, and sparse error estimation);

- transform domain models (fast Fourier transform, discrete cosine transform, Walsh transform, wavelet transform, and Hadamard transform).

Not all from mentioned above methods are suitable for the monitoring task in the wild. Advanced statistical models and codebook models are among the most promising methods.

Background extraction is the cornerstone of background subtraction method. One of the traditional methods suitable for natural scene analysis is a temporal median filter method [3]. It requires a durable observation during training step. A median value of the certain pixel points extracted from  $K$  frames is taken as the background pixel value in this point. The improvement of this method called as the average method supposes to calculate the average value instead of the median value. The incremental form of the average method is often used for real-time application, when for each pixel  $k$  the background model is update using equation

$$B_{k+1} = \frac{n-1}{n} B_k + \frac{1}{n} I_k, \quad (1)$$

where  $B_k$  and  $B_{k+1}$  are the intensities in the current background model and new background model, respectively;  $n$  is the number of frames;  $I_k$  is the intensity in current frame.

The incremental method has lesser computational cost respect to the temporal median filter method and provides better extraction result. In [4], it was shown that if  $n = 100$  and more the incremental method has become a running average background learning method:

$$B_{k+1} = (1-\alpha) B_k + \alpha I_k, \quad (2)$$

where  $\alpha$  is the learning factor,  $\alpha = 0.01$ ,  $\alpha = 0.1$ , or another experimental constant.

This method is widely used in practice; however, it is prone to generate the ghosts.

The Gaussian mixture model for background estimation was proposed in [5]. In the GMM, the pixel's intensity values over a time are modeled by a single Gaussian or as a mixture of several Gaussians. The background pixels are identified by comparing the pixel values and mean values of models. Many improvements of the GMM are available in literature [6–11]. The GMM is appropriate for complex natural scenes including tree branches shaking, water rippling, etc. The disadvantages of the GMM are the high computational complexity and necessity to store the Gaussian model parameters.

Difficulties in building a proper mathematical model, which describes the probability density function of pixel values, led to development of a non-parametric approach for background modeling. In [12], a Kernel Density Estimation (KDE) was proposed with the main idea to evaluate the intensity density of pixels directly from sample history values that made this method sensitive to detection of moving objects. A nonparametric background generation model for on-line surveillance was proposed in [13]. First, the statistics of background variations without training samples were estimated. Second, the background was generated using a heuristic framework. The combination of the KDE and GMM was offered in [14] in order to estimate accurately the density function of background. Nonparametric estimation methods adapt to fast changes' detection in a scene. At the same time, they have unsatisfactory background building in situations, when several moving objects have different speeds.

In [15], the background was modelled using a codebook algorithm. This method is referred to cluster models. For each pixel, a codebook consisting of one or more codewords is constructed based on a color distortion metric together with brightness bounds. Generally, the clusters represented by codewords do not correspond to single Gaussian or other parametric distributions. If the color distortion of incoming pixel to some codewords is less than the threshold and its brightness lies within the brightness range of that codeword, then this pixel is classified as background, otherwise, it is classified as foreground. The codebook algorithm estimates a background over a long period with a limited memory. The original algorithm was improved in several ways. Thus, a multilayer codebook model was proposed in [16], which removed most of the dynamic background and significantly increased the computational efficiency.

A universal sample-based background subtraction algorithm called as Visual Background extractor (ViBe) was developed in [17]. A classification model was based on a small number of correspondences between a candidate value and the corresponding background pixel model. The ViBe can be initialized with a single frame under assumption that

neighboring pixels share a similar temporal distribution. Also, an original mechanism for updating the background model over time for a set of frames was presented. Hereinafter, the extensions of the ViBe approach were proposed in order to eliminate the ghosts [18]. The ViBe method has advantages in the computation speed and detection effect but does not invariant to frequent background changes.

A robust background extraction algorithm called as Neighbor-based Intensity Correction (NIC) method was offered in [19]. The NIC method identified and modified the motion pixels from the difference of the background and the current frame. The first frame was considered as an initial background and updated by the pixel intensity from new frame based on the analysis of neighborhood surrounding. In the intensity modification procedure, the comparison of the standard deviation values calculated from two pixel windows was executed. Finally, the foreground is detected by the background subtraction algorithm with an optimal threshold calculated by the Otsu method.

Two universal modifications, such as dynamic background estimation and complementary learning, were implemented in GMM, ViBe, and codebook algorithms for complex dynamic background modelling and accurate foreground objects [20]. Combining the complementary learning technique, these improved algorithms had good performance on the detection of dynamic background including waving tree, rippling water, and fountain.

The approach, when the background model was augmented with an explicit foreground model, was developed in [21]. Thus, two statistical models (background and foreground) were used in a closed loop. A background model is periodically updated to account for illumination changes, while foreground detection corrupts the intensity of the background model. In addition to a non-parametric background model, these authors used a foreground model based on small spatial neighborhood. The hypothesis test and the Markov random field improved a spatial coherence of the detections. Such approach can be combined with non-parametric kernel or mixture of Gaussians.

In [22], a texture-based background model was proposed using Local Binary Patterns (LBPs). In spite of LBPs are invariant to illumination changes, they are not robust to noise. For example, if the central pixel value in LBP is affected by noise, then the corresponding LBP histogram provides the increased number of false positive or false negative errors.

A short literature survey shows that the interest to the background extraction algorithms remains stable. These algorithms are developed in various directions, such as accuracy of object detection, computational costs, and robustness to various

factors. Also, a goal of solving task determines a choice of approach.

### Mixture of Gaussians models

In Gaussian mixture model, each pixel's intensity is determined by a mixture of  $K$  Gaussian distributions, where  $K$  is a small number ranging between 3 and 5. Each Gaussian distribution is associated with its contributing weight. The mean  $\mu_k$ , the variance  $\sigma_k^2$ , and a weight  $w_k$  are the main parameters of GMMs. Evaluation of these parameters can be implemented using an Expectation Maximization (EM) algorithm with recently observed data. The EM algorithm has high computational cost, instead of which a recursive algorithm that updates the GMM parameters at each time instance is commonly used.

Thus, in general GMM each pixel is considered as a mixture of  $K$  Gaussian distributions, which probability  $P(X_t)$  is evaluated by equation

$$P(X_t) = \sum_{j=1}^K w_{j,t} * \eta(X_t, \mu_{j,t}, \Sigma_{j,t}), \quad (3)$$

where  $X_t$  is the pixel value at time  $t$ ;  $K$  is the number of Gaussian distributions;  $w_{j,t}$  is the weight value;  $\mu_{j,t}$  is the mean value, and  $\Sigma_{j,t}$  is the covariance matrix of the  $j^{\text{th}}$  Gaussian at time  $t$ , respectively;  $\eta$  is the Gaussian PDF. The Gaussian PDF  $\eta$  is defined by equation

$$\eta(X_t, \mu_{j,t}, \Sigma_{j,t}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_{j,t}|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - \mu_{j,t})^T (\Sigma_{j,t})^{-1} (X_t - \mu_{j,t})}, \quad (4)$$

where  $n$  is the dimension of  $X_t$ .

For simplicity, the covariance matrix  $\Sigma_{j,t}$  is defined as  $\sigma_{j,t}^2 \mathbf{I}$  for the  $j^{\text{th}}$  component, where  $\mathbf{I}$  is the identity matrix, under assumption that the components of  $X_t$  (Red, Green, and Blue) are independent and have the same variances.

The background distributions have higher probabilities and smaller variances due to the probable background colors stay longer than the foreground colored objects. This observation makes the GMM an updating model. New coming pixel is checked respect to the existing model components. If the pixel value is within 2.5 standard deviations of some weighted Gaussian distribution, then this distribution is updated. In the opposite case, a distribution with minimum weight is replaced by a new distri-

bution using the current mean value. This new distribution obtains the high initial variance and low prior weight. Then the  $K$  distributions are sorted according to value  $w_{j,t}/\sigma_{j,t}$ , where  $\sigma_{j,t}$  is the 1D variance of the  $j^{\text{th}}$  Gaussian in the mixture at time  $t$ . The first  $B$  distributions are selected as the background model using equation

$$B = \arg \min_b \left( \sum_{k=1}^b w_{k,t} > T_B \right), \quad (5)$$

where  $b$  is the number of selected Gaussian distributions;  $T_B$  is the predefined threshold (it represents the minimal quantity of the data that ought to be considered as the background model and usually is set to close to 90 %).

When the matching process of the incoming pixel is completed, the prior weights of  $K$  Gaussian distributions are changed by equation

$$w_{k,t} = (1 - \alpha)w_{k,t-1} + \alpha(M_{k,t}), \quad (6)$$

where  $\alpha$  is the learning rate;  $M_{k,t}$  equals 1 for the matched distribution and equals 0 for the unmatched distribution.

The weights of distributions are renormalized by updating the values of mean and variance applying equation

$$\begin{aligned} \mu_t &= (1 - \rho)\mu_{t-1} + \rho X_t; \\ \sigma_t^2 &= (1 - \rho)\sigma_{t-1}^2 + \rho(X_t - \mu_t)^T (X_t - \mu_t), \end{aligned} \quad (7)$$

where

$$\rho = \alpha \eta(X_t | \mu_k, \sigma_k).$$

The multiple modifications of GMM follow the main idea to support three consecutive stages: background initialization, background estimation, and background update.

### Proposed method for object detection

Consider a scene, where an object (animal, bird, or rarely human) is periodically appeared and its appearance is captured by a camera trap. The location of camera trap is chosen by foresters based on long-time observations of a territory. A camera trap captures any motion in scene in any time and as a result provides a series of images through 3–5 s (this means an obtaining of 6–8 images with a relatively good visibility) or a movie with duration 8–10 s. Suppose that we have a set of image series taken in different seasons.

A scene remains the same with different chromaticity due to a season. This means that we can elab-



orate a background initialization in detail building the GMMs for all seasons. The following two stages: background estimation and background update, are executed when a new image series is incoming. Note that each image contains information about time, day, and temperature, and each GMM is also associated with time and season information. Thus, we ought to find a correspondence between the current image series and GMM by query.

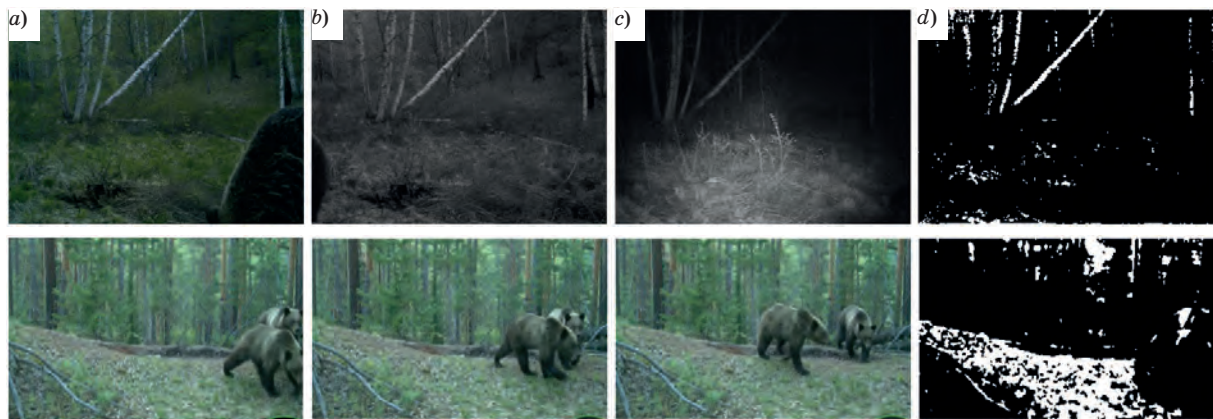
Consider the consecutive stages of the proposed method.

### Preliminary image segmentation

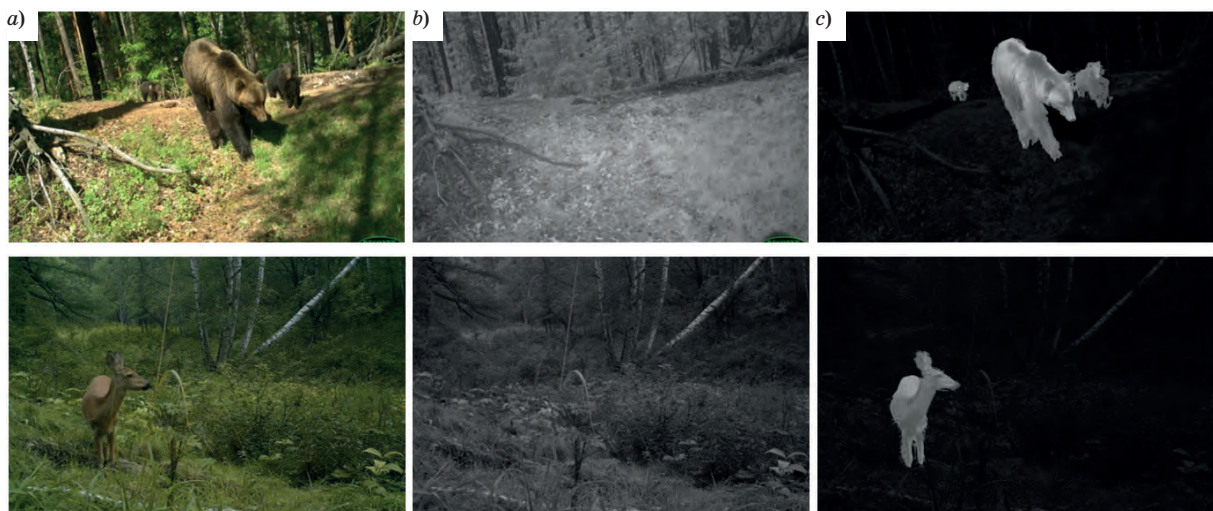
The best selection for preliminary image segmentation is a winter season, when a background has a restricted palette of colors with prevailing white, brown, and black colors and their corresponding color shades. Let us roughly consider a scene as a combination of the structural elements (trunks of growing and/or fallen trees), which posi-

tions change rarely, other textured regions depends from season and unknown moving objects.

First, the dark colored masks (in the case of boreal forests) are extracted from a series of winter images and combined in order to create a common winter mask. During mask creation, only extensional dark regions are marked as the candidates for structure elements. Second, this procedure is applied for the series of spring, summer, and autumn images. As a result, the common spring, summer, and autumn masks are obtained. Note that the sizes of structural elements are the biggest in the winter mask, while in other seasons can be overlapped by foliage. Third, the generalized masks with structural elements are created by imposing the common masks. Only the common parts of all masks are considered as the reliable landmarks in a scene. The building of season masks is depicted in Fig. 1, *a-d*, while their combination and obtaining the reliable landmarks is given in Fig. 2, *a-c*.



■ Fig. 1. Examples of background model building: *a, b, c* — original images (the top row — at night time, the bottom row — at daily time); *d* — detected structural elements of the corresponding images



■ Fig. 2. Background model building using reliable landmarks: *a* — images with animals; *b* — reference background images; *c* — results of animal detection

The detected structural elements may be useful for alignment of following incoming images.

Then the distributions of corresponding textured regions (trunks, snow, foliage, grass, sky, lake, river, and Earth surface for boreal forests) are built using the rich experimental material stored by 5 previous years. In other word, a description of each texture transmitted from RGB- to YUV-color space is a feature vector, which includes statistical parameters of distribution (mean value  $M$ , variances  $\sigma$ , homogeneity  $U$ , smoothness  $R$ , and entropy  $E$ ) [23]. The corresponding formulae are pointed in Table 1.

Also, two modified texture features — relative smoothness  $R_{md}$ , and normalized entropy  $E_{nr}$  — can be calculated using equations

$$R_{md} = \begin{cases} -\log R & \text{if } R > 0; \\ 10 & \text{if } R = 0, \end{cases} \quad (8)$$

$$E_{nr} = E / \log_2 L, \quad (9)$$

where  $L$  is a number of brightness levels,  $L > 1$ .

If parameter  $R = 0$ , then we forcibly maintain a relative smoothness  $R_{md} = 10$  (small empirical value differing from 0). Normalized entropy  $E_{nr}$  indicates some equalization effect in dark and bright areas of frame.

The main parameters are the mean value and variance. The remained parameters, such as the

■ **Table 1.** Statistical texture features

Caption	Equation
Central moment by order $k$	$\mu_k(z) = \sum_{i=0}^{L-1} (z_i - M)^k p(z_i)$
Mean value $M$	$M = \sum_{i=0}^{L-1} z_i p(z_i)$
Variance $\sigma$	$\sigma^2 = \mu_2(z) = \sum_{i=0}^{L-1} (z_i - m)^2 p(z_i)$
Homogeneity $U$	$U = \sum_{i=0}^{L-1} p^2(z_i)$
Smoothness $R$	$R = 1 - \frac{1}{1 + \sigma^2(z) / (L-1)^2}$
Entropy $E$	$E = - \sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i)$

homogeneity, smoothness, and entropy, serve as the additional parameters in order to decrease a number of background clusters.

### Background initialization (training stage)

Suppose that the training set includes several dozen of images captured by a single camera trap in all seasons. Each image is divided into non-overlapping blocks of sizes of  $n \times n$  pixels. Each block is characterized by two parameters: the normalized mean value  $M$  calculated as follows:

$$M = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n pv_{ij}, \quad (10)$$

where  $pv_{ij}$  is the pixel value in the position  $(i, j)$  of the block, and binary bitmap  $BM$  similar to modified LBPs:

$$bv_{ij} = \begin{cases} 1, & \text{if } pv_{ij} \geq M; \\ 0, & \text{otherwise,} \end{cases} \quad (11)$$

where  $bv_{ij}$  means the bit in the position  $(i, j)$  of a  $BM$ .

During experiments, we used  $n = 4$  and  $n = 8$  depending an image resolution.

To avoid a impact of sunny weather that stimulates the deviation of hue and shadows' presence, we replace two low bits of pixels' values in textured regions by zeros.

The forest background is such that small number of clusters describes a background model. Initially,  $K$  different binary bitmaps are randomly generated for  $\{BM_1, BM_2, \dots, BM_K\}$  blocks with the weights  $1/K$ . Each binary bitmap is assigned a weight  $w_k$  between 0 and 1 and the sum of  $K$  weights equals 1. A new block  $BM_{new}$  is compared with the  $K$  bitmaps using the Hamming distance  $HD$  provided by following Equation, where  $k$  is in the range of  $[1, K]$ ,  $\oplus$  is the summation of mod 2:

$$HD(BM_{new}, BM_k) = \sum_{i=1}^n \sum_{j=1}^n (\{bv_{ij}\}_{new} \oplus \{bv_{ij}\}_k). \quad (12)$$

The block  $BM_{new}$  matches to the block  $BM_k$  if inequality (13) is satisfied, where  $TH$  is the predefined threshold, and, otherwise, the block  $BM_{new}$  is regarded as a new background cluster:

$$\min HD(BM_{new}, BM_k) < TH. \quad (13)$$

Then, the weight of each block is updated by:

$$w'_k = \alpha W_k + (1 - \alpha) w_k, \quad (14)$$

where  $\alpha$  is the learning rate;  $W_k$  is the coefficient, which for the best-matched bitmap equals 1 and for remaining bitmaps equals 0.

The GMM, which short description one can find in Section “Mixture of Gaussians models”, helps to create a generalized background model, reliable in statistical meaning. For simplicity, we analyze the variances instead of using a covariance matrix. According to Equations (8)–(14), we create several GMMs for all seasons in daily/night time for Y-channel in YUV-color space. The requirements to GMM can be weaken and computation becomes simpler through use of texture descriptor in a view of a binary bitmap.

The proposed method provides the background models of a single scene with low computational loads during the working stage because the comparison is implemented on the level of binary values.

### Background estimation and background update (working stage)

The stored images through a half of a year are processed in the package mode. First, they ought to be sorted manually or automatically respect to camera traps and seasons. Second, a procedure of background estimation and update is executed based on the initial corresponding GMM. The working stage does not principally differ from the training stage. The calculations use Equations (10)–(14) but the goal of the working stage is to find the position of an object of interest. Note that camera trap captures an image, when a movement in a scene is detected. The algorithm finds the region in an image, which distribution differs from the background distributions. The definition of animal or bird type is outside of this article.

The described scheme works well under good meteorological conditions. However, meteorological conditions impact significantly on the quality on an image and, consequently, a potential ability for animal/bird detection. At this sense, the algorithm ought to detect the type of meteorological impact, estimate a degree of distortion, and restore an image if a degree of distortion is minor.

Fog can be detected by analysis of color ranges. If the color ranges are restricted and deposed to the higher values, then the effect of “whitened color” has a high probability. This is a simple procedure of histogram analysis in RGB-color space. The

threshold of decision making is determined empirically.

Rain and snowfall are simulated as a noticeable noise with specific structure. For example, rain remains the short line segments of white color, which have the identical directions. Snowfall keeps the white spots of different size and shape. Thus, the algorithm searches these structural elements uniform distributed on a whole image. Also, the decision marking is based on empirical observations.

The most interest cases appear, when the distortions are small and the algorithm tries to restore the damaged images. One can read about possible restoration techniques in previous publications of the authors [24]. Sometimes the complex methods including morphological closing of visual objects ought to be applied. Example of reconstructed image is depicted in Fig. 3, *a–d*.

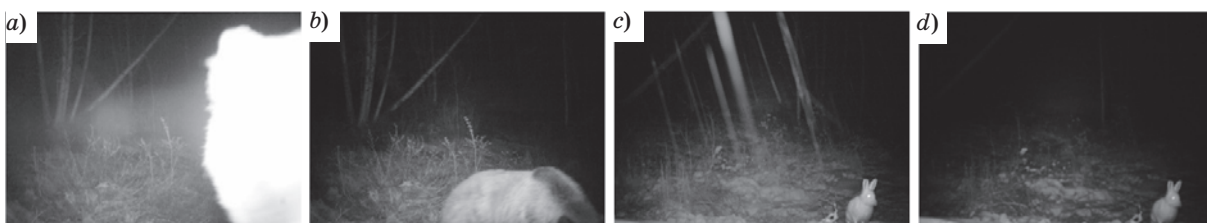
However, when meteorological conditions are too bad during shooting, the object detection is impossible ever by a human vision.

### Experimental Results

Experiments were conducted using the dataset of images captured in the territory of wildlife sanctuary Ergaki, Krasnoyarsky Kray, Russia. This dataset includes more than 38,000 images of animals captured by camera traps in different weather conditions and different seasons. The most number of images have the complex structure, various artifacts, and noises. Near for 1,000 images, there were built the masks with the localized animals or birds manually (Fig. 4, *a, b*).

During experiments, the automatic detection of localization of animal or bird was implemented using the marked volume of dataset. The designed algorithm includes such main steps as a background modelling, saliency detection, and localization of animal or bird in an image. Some results are depicted in Fig. 5, *a–e*.

Animal localization using saliency detection procedure shows good results if an animal is situated in the middle area of an image and also if an animal differs by color or intensity compared the



■ Fig. 3. Examples of complex cases: *a* — high illumination at night; *b* — foreground object close to the camera; *c* — weather impact (small fog); *d* — example of a reconstructed image





■ Fig. 4. Examples of ground truth images from Ergaki 2018 dataset: *a* — original images; *b* — ground truth masks



■ Fig. 5. Examples of segmentation and animal localization using Ergaki 2018 dataset: *a* — original images 2012\_bear.jpg, 2017\_IMG5044.jpg, and 2013\_PICT1696.jpg; *b* — ground truth image segmentation; *c* — saliency detection; *d* — masks obtained from background and saliency estimation; *e* — results of localization





■ Fig. 6. Examples of saliency detection using Ergaki 2018 dataset: a — successful examples; b — poor examples

■ Table 2. Comparative results of detection of animals and birds using a coverage measure

Method	80 % Coverage	90 % Coverage	Best Coverage (Maximum %)
Selective Search [25]	2829.7	5903.5	13 882 (99.8)
GOP [26]	2489.1	3984.6	9874 (98.2)
MOP [27]	335.8	482.3	891.7 (96.7)
FCOP [28]	132.8	384.2	<b>393.1 (90.4)</b>
SORPPV [29]	<b>95.4</b>	237.3	626.9 (93.1)
Proposed method	127.2	<b>226.1</b>	689.1 (94.2)

background (Fig. 6, a, b). In some cases, an animal localization in an image is even difficult even for a human (Fig. 6, b).

For efficiency evaluation, the *F*-measure was applied:

$$F = \frac{2TP}{2TP + FN + FP}, \quad (15)$$

where *TP* is the number of true positive; *FP* is the number of false positive; *FN* is the number of false negatives. A region is considered as a true positive if it has more than 50 % intersection with a ground-truth bounding mask.

Also many algorithms are evaluated using a term of coverage as a numerical measure of the corresponding detected and ground truth pixels [25–29]. The comparative results using a coverage measure and common comparative results using *F*-measure for detection of animals and birds in the images, which are involved in Ergaki 2018 dataset, are represented in Tables 2 and 3, respectively.

■ Table 3. Common comparative results of detection of animals and birds using *F*-measure

Method	<i>F</i> -measure
EC-Best [30]	0.7703
YOLO [31]	0.7515
Fast-RCNN [32]	0.7937
SORPPV [29]	<b>0.8398</b>
Proposed method	0.7812

The best resulted in Tables 2 and 3 are marked by Bold.

The efficiency of detection of animals and birds in images achieves 70–80 % depending an image quality and weather conditions. The use of saliency detection algorithm allows us to increase this parameter on 3–8 %.

### Conclusions

In this article, a background extraction method for automatic detection of animals and birds in the wild using camera trap images was developed. The experiments were conducted using rich dataset of natural images obtained on the territory of wildlife sanctuary Ergaki, Krasnoyarsky Kray, Russia. The proposed method provides the detection of animals and birds on a level 70–80 % due to the multiple challenges caused by shooting and weather conditions.

### Acknowledgments

The reported study was funded by Russian Foundation for Basic Research, Government of Krasnoyarsk Territory, Krasnoyarsk Regional Fund of Science to the research project No. 18-47-240001.

## References

1. Bouwmans T. Recent advanced statistical background modeling for foreground detection — a systematic survey. *Recent Patents on Computer Science*, 2011, vol. 4, no. 3, pp. 147–176.
2. Bouwmans T. Traditional and recent approaches in background modeling for foreground detection: an overview. *Computer Science Review*, 2014, vol. 11–12, pp. 31–66.
3. Hung M. H., Pan J. S., Hsieh C. H. Speed up temporal median filter for background subtraction. *International Conference on Pervasive Computing Signal Processing & Applications*, 2010, pp. 297–300.
4. Yi Z., Fan L. Moving object detection based on running average background and temporal difference. *International Conference on Intelligent Systems and Knowledge Engineering*, 2010, pp. 270–272.
5. Stauffer C., Grimson W. E. L. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, vol. 22, no. 8, pp. 747–757.
6. Zhang Y., Liang Z., Hou Z., Wang H., Tan M. An adaptive mixture Gaussian background model with online background reconstruction and adjustable foreground merge time for motion segmentation. *IEEE International Conference on Industrial Technology*, 2005, pp. 23–27.
7. Kim H., Sakamoto R., Kitahara I., Toriyama T., Kogure K. Robust foreground extraction technique using Gaussian family model and multiple thresholds. *Asian Conference on Computer Vision*, 2007, pp. 758–768.
8. Bouwmans T., El Baf F. Modeling of dynamic backgrounds by type-2 fuzzy Gaussians mixture models. *MASAUM Journal of Basic and Applied Sciences*, 2010, vol. 1, no. 2, pp. 265–276.
9. Shah M., Deng J., Woodford B. Illumination invariant background model using mixture of Gaussians and SURF features. *International Workshop on Background Models Challenge, Asian Conference on Computer Vision*, 2012, pp. 308–314.
10. Elguebaly T., Bouguila N. Background subtraction using finite mixtures of asymmetric Gaussian distributions and shadow detection. *Machine Vision and Applications*, 2014, vol. 25, no. 5, pp. 1145–1162.
11. Alvar M., Rodriguez-Calvo A., Sanchez-Miralles A., Arranz A. Mixture of merged Gaussian algorithm using RTDENN. *Machine Vision and Applications*, 2014, vol. 25, no. 5, pp. 1133–1144.
12. Elgammal A., Harwood D., Davis L. Non-Parametric Model for background subtraction. *The 6th European Conference on Computer Vision*, 2000, part II, LNCS, vol. 1843, pp. 751–767.
13. Zhang R., Gong W., Yaworski A., Greenspan M. Non-parametric on-line background generation for surveillance video. *The 21st International Conference on Pattern Recognition*, 2012, pp. 1177–1180.
14. Liu Z., Huang K., Tan T. Foreground object detection using top-down information based on EM framework. *IEEE Transactions on Image Processing*, 2012, vol. 21, no. 9, pp. 4204–4217.
15. Kim K., Chalidabhongse T. H., Harwood D., Davis L. Real-time foreground–background segmentation using codebook model. *Real-Time Imaging*, 2005, vol. 11, no. 3, pp. 172–185.
16. Guo J. M., Hsia C. H., Liu Y. F., Shih M. H. Fast background subtraction based on a multilayer codebook model for moving object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 2013, vol. 23, no. 10, pp. 1809–1821.
17. Barnich O., Van Droogenbroeck M. ViBe: a universal background subtraction algorithm for video sequences. *IEEE Transactions on Image Processing*, 2011, vol. 20, no. 6, pp. 1709–1724.
18. Guang H., Wang J., Xi C. Improved visual background extractor using an adaptive distance threshold. *Journal of Electronic Imaging*, 2014, vol. 23, no. 6, pp. 063005-1–063005-12.
19. Huynh-The T., Banos O., Lee S., Kang B. H., Kim E. S., Le-Tien T. NIC: a robust background extraction algorithm for foreground detection in dynamic scenes. *IEEE Transactions on Circuits and Systems for Video Technology*, 2017, vol. 27, no. 7, pp. 1478–1490.
20. Ge W., Guo Z., Dong Y., Chen Y. Dynamic background estimation and complementary learning for pixel-wise foreground/background segmentation. *Pattern Recognition*, 2016, vol. 59, pp. 112–125.
21. McHugh J. M., Konrad J., Saligrama V., Jodoin P. M. Foreground-adaptive background subtraction. *IEEE Signal Processing Letters*, 2009, vol. 16, no. 5, pp. 390–393.
22. Heikkilä M., Pietikäinen M. A Texture-based method for modeling the background and detecting moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, vol. 28, no. 4, pp. 657–662.
23. Favorskaya M. N., Damov M. V., Zotin A. G. Intelligent method of texture reconstruction in video sequences based on neural networks. *International Journal of Reasoning-based Intelligent Systems*, 2013, vol. 5, no. 4, pp. 223–236.
24. Favorskaya M., Jain L. C., Bolgov A. Image inpainting based on self-organizing maps by using multi-agent implementation. *Procedia Computer Science*, 2014, vol. 35, pp. 861–870.
25. Uijlings J. R., van de Sande K. E., Gevers T., Smeulders A. W. Selective search for object recognition. *International Journal Computer Vision*, 2013, vol. 104, no. 2, pp. 154–171.
26. Krahenbuhl P., Koltun V. Geodesic object proposals. *Proceeding European Conference Computer Vision*, 2014, pp. 725–739.
27. Fragkiadaki K., Arbelaez P., Felsen P., Malik J. Learning to segment moving objects in videos. *Pro-*

- ceeding *IEEE Conference Computer Vision Pattern Recognition*, 2015, pp. 4083–4090.
28. Perazzi F., Wang O., Gross M., Sorkine-Hornung A. Fully connected object proposals for video segmentation. *Proceeding IEEE International Conference Computer Vision*, 2015, pp. 3227–3234.
29. Zhang Z., He Z., Cao G., & Cao W. Animal detection from highly cluttered natural scenes using spatio-temporal object region proposals and patch verification. *IEEE Transactions on Multimedia*, 2016, vol. 18, no. 10, pp. 2079–2092.
30. Redmon J., Divvala S., Girshick R., Farhadi A. You only look once: Unified, real-time object detection CoRR. 2015. Available at: <http://arxiv.org/abs/1506.02640> (accessed 5 August 2013).
31. Girshick R. Fast r-CNN. *Proceeding International Conference Computer Vision*, 2015, pp. 1440–1448.
32. Shaoqing Ren K. H., Ross Girshick J. S. Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.

УДК 004.932

doi:10.31799/1684-8853-2018-6-35-45

**Метод извлечения фона в естественных изображениях, полученных от фотоловушек**М. Н. Фаворская<sup>а</sup>, доктор техн. наук, профессор, [orcid.org/0000-0002-2181-0454](https://orcid.org/0000-0002-2181-0454), [favorskaya@sibsau.ru](mailto:favorskaya@sibsau.ru)В. В. Буряченко<sup>а</sup>, канд. техн. наук, доцент, [orcid.org/0000-0003-1151-1159](https://orcid.org/0000-0003-1151-1159)<sup>а</sup>Сибирский государственный университет науки и технологий им. академика М. Ф. Решетнёва, Красноярский рабочий пр., 31, Красноярск, 660037, РФ

**Постановка проблемы:** автоматическое обнаружение животных и птиц в природе на изображениях, полученных от фотоловушек, остается нерешенной проблемой из-за условий съемки и погодных факторов. В результате таких наблюдений формируется большой объем изображений, тысячи или миллионы, которые невозможно анализировать вручную. Обычно в заповедниках и национальных парках используются бюджетные фотоловушки. Поэтому низкокачественные изображения, полученные с их помощью, требуют тщательной многократной обработки перед тем, как распознавать виды животных или птиц. **Цель:** разработка метода извлечения фона на основе модели смеси гауссианов для обнаружения объекта интереса при любых временных/сезонных/метеорологических условиях. **Результаты:** предложен метод извлечения фона на основе модифицированной модели смеси гауссианов. Модификация заключается в усечении значений пикселей (младшие разряды) для уменьшения зависимости от изменений освещенности и наличия теней с последующим созданием и обработкой бинарных масок вместо реальных значений интенсивностей. Предлагаемый метод предназначен для оценки фона естественных сцен в заповедниках и национальных парках. Структурные элементы (стволы растущих и (или) упавших деревьев) считаются регионами, медленно изменяющимися в течение сезонов, в то время как другие текстурированные области моделируются текстурными шаблонами, соответствующими текущему сезону. Такой подход обеспечивает компактную модель фона сцены. Помимо этого, мы рассматриваем влияние временных/сезонных/метеорологических атрибутов сцены относительно возможности ее восстановления. Метод был протестирован с использованием богатого набора данных естественных изображений, полученных на территории заповедника «Ергаки», Красноярский край, Россия. **Практическая значимость:** применение модифицированной модели смеси гауссианов показывает точность распознавания объектов 79–83 % в дневное время и 60–69 % в ночное время суток при нормальных метеорологических условиях. При этом точность восстановленных изображений, полученных при плохих метеорологических условиях, снижается на 5–8 %.

**Ключевые слова** — вычитание фона, естественная сцена, модель смеси гауссианов, обнаружение животного, модель фона.

**Цитирование:** Favorskaya M. N., Buryachenko V. V. Background extraction method for analysis of natural images captured by camera traps. *Информационно-управляющие системы*, 2018, № 6, с. 35–45. doi:10.31799/1684-8853-2018-6-35-45

**Citation:** Favorskaya M. N., Buryachenko V. V. Background extraction method for analysis of natural images captured by camera traps. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2018, no. 6, pp. 35–45. doi:10.31799/1684-8853-2018-6-35-45