







Visual Data Processing Framework for a Skin-Based Human Detection

Valery Myrzin¹, Tatyana Tsoy¹ , Yang Bai² , Mikhail Svinin² ,
and Evgeni Magid¹ 

¹ Laboratory of Intelligent Robotics Systems (LIRS), Intelligent Robotics
Department, Institute of Information Technology and Intelligent Systems,
Kazan Federal University, Kazan, Russia
{tt,magid}@it.kfu.ru

² Information Science and Engineering Department, College of Information Science
and Engineering, Ritsumeikan University, 1-1-1 Noji-higashi, Kusatsu, Shiga
525-8577, Japan
{yangbai,svinin}@fc.ritsumei.ac.jp

<https://kpfu.ru/eng/itis/research/laboratory-of-intelligent-robotic-systems>,
<http://en.ritsumei.ac.jp/academics/college-of-information-science-and-engineering/>

Abstract. For a large variety of tasks autonomous robots require a robust visual data processing system. This paper presents a new human detection framework that combines rotation-invariant histogram of oriented gradients (RIHOG) features and binarized normed gradients (BING) pre-processing and skin segmentation. For experimental evaluation a new *Human body dataset* of over 60000 images was constructed using the Human-Parts dataset, the Simulated disaster victim dataset, and the Servosila Engineer robot dataset. Random, Linear SVM, Quadratic SVM, AdaBoost, and Random Forest approaches were compared using the Human body dataset. Experimental evaluation demonstrated an average precision of 90.4% for the Quadratic SVM model and showed the efficiency of RIHOG features as a descriptor for human detection tasks.

Keywords: Visual data processing · Skin segmentation · Feature extraction · Image classification · Mobile robot

1 Introduction

In the recent decades a consistent growth of interest to a human-robot interaction (HRI) field is being indicated. Such interactions have many application areas along with challenges to overcome, and a significant number of issues is tied to a natural difference in human and robot perception of an environment [9]. There are several areas where robust perception of human beings is crucial for a robot in order to effectively perform its tasks [26]. Industrial robots that work in a tight collaboration with a human need a human visual perception system to decrease a potential risk of harming people, which work within the same workspace [13].

Social robots heavily rely on visual and voice recognition to act according to their corresponding roles such as a teacher [38], a nurse [2], a personal assistant [6] etc. Autonomous surgical robots require a strong and precise visual perception of a human body in order to perform complicated operations successfully [20].

Another field that heavily rely on robot perception capabilities is Urban search and rescue (USAR) [25]. UAVs and UGVs can be used in cooperation with USAR teams to detect victims allowing to reduce safety risk for the teams at a disaster site [24, 34]. SLAM-based navigational algorithms [31] and properly implemented communication protocols [28] allow rescue robots to explore the disaster site without endangering the rescuers life. Victims identification system with a built-in robot sensor as an integral part of such rescue robots is crucial for increasing a number of survivors. Detecting disaster victims is a challenging task due to articulated nature of human body, cluttered background environment, large area covered by dust layer [1]. Considering such obstacles only a part of human body may be seen from the robot camera which makes victim detection even more challenging.

This paper takes a closer look at various image processing methods and aggregate them into a single framework suitable for a human body detection by a Servosila Engineer rescue robot camera [30, 36]. For each frame we used binarized normed gradients (BING, [7]) approach for a windows generation proposal. Next, extracted RIHOG [23] features were classified and combined with a skin segmentation method in order to achieve a robust human body identification system.

The rest of the paper is structured in the following way. An overview of related work is introduced in Sect. 2. Section 3 describes image processing algorithms typically used for a human detection. Experimental results are presented in Sect. 4. Finally, we conclude in Sect. 5.

2 Related Work

A large number of research works are devoted to a selection of a so-called visual saliency - an area of an image that visually stands out from its surroundings. Itti et al. [18] were the first to introduce the visual saliency approach, which had been inspired by a visual system of early primates. The authors used a color, an intensity, and an orientation of an image to create saliency maps, which are then used for a region of interest identification. Harsen et al. [14] proposed another approach of finding saliency based on a graph calculation. The Graph-Based Visual Saliency (GBVS) approach was modified using Local Entropy Feature specially for a disaster victim detection in [16]. Fast, accurate, and size-aware salient object detection (FASA) algorithm [40] used a combination of saliency probability and global contrast maps to calculate a final saliency map. Cheng et al. [7] developed a new method for finding saliency using a normalized gradient magnitude of low-resolution images to create a fast algorithm for image pre-processing. Bharath et al. [5] used saliency calculation as an integral part of a framework for classification and scene understanding in an image.

Pre-processing algorithms only reduce a region of interest (ROI) for an estimated object and are unable to recognize an object itself. To recognize the object it is required to distinguish it using classification, i.e., to determine if the object corresponds to a certain class. Classification can be done by various machine learning techniques, which use feature descriptors extracted from an image. The descriptors are vectors with a high enough discriminative power allowing to distinguish one object from another. A number of research works were devoted to an effective descriptor development, e.g., Dalal and Triggs [11] proposed a method of human recognition based on a feature descriptor - a histogram of orientated gradients (HOG), which successfully described a local shape of an object and kept invariance to various conditions of illumination, shadowing, noise, etc.

A significant number of scientists adopted HOG for various classification tasks including disaster victims detection [4, 32]. Soni and Sowmya [35] developed a framework for a HRI in search and rescue operations and proposed methods for human body parts recognition. Davis and Sahin [12] combined HOG with RGB, Thermal, and Depth images to create a robust multi-layer classifier. Huang and Chu [17] used HOG with Support Vector Machines (SVM) learning to detect a person and developed a specific system to determine if a victim was immobilized by a heavy object. Liu et al. [23] proposed a modification of HOG adding invariance to an object orientation with a help of Fourier analysis and data translation to a polar system. Other researchers used a rotation invariant histogram of oriented gradients (RIHOG) to supplement a Bag of Visual Words classification method for the task of disaster victims identification [22].

Other works related to determination of a victim presence in an image included various skin segmentation techniques [10, 19], e.g., Hoshino and Niimura used Optical Flow combined with Convolution Neural Network to achieve a real-time human detection [15]. Perera et al. [29] applied a video stream stabilization and signal processing to track human chest movements in order to detect a process of the human breathing, and then localized the human in an image based on the breathing region.

3 Image Data Processing

The proposed algorithm workflow is presented in Fig. 1. It consists of several steps including pre-processing, skin segmentation, and feature extraction. For the pre-processing step we adopted BING method, which generates a reliable window proposal. Next, the skin segmentation is used to rearrange the window proposal in order to use the windows with skin entries first. The feature extraction is performed in the final window proposal, which is then classified to accomplish the detection process.

3.1 Pre-processing

In the classic approach of determining a desired object location within an image the entire image is scanned with a fixed size window and a predefined step.

Afterwards a region of the image in all windows is classified for a presence of an object of interest. Such process is called a sliding window paradigm. The problem of this method is a consumption of a significant amount of computational resources, which reduces the efficiency of the detector and makes it unsuitable for real-time detection tasks due to slow computations. Many works were devoted to various image pre-processing algorithms that attempted to solve the sliding window problem and reduce the problem by reducing a search area of the object thus increasing a speed of the image processing.

According to Cheng et al. [7], Binarized Normed Gradients were designed to speed up the sliding window paradigm. The authors discovered that generic objects with a well-defined closed boundary could be discriminated by looking at a norm of gradients magnitude with a suitable resizing of their corresponding image windows into a small fixed size. Taking this feature as a basis in their algorithm, the size of all windows in the image was changed to 8×8 pixels. Thus, processing of the entire image resulted in a 64-dimensional gradient magnitude vector. The support vector machine method was used for training the 64-dimensional descriptors to differentiate generic objects on the image and evaluate objectness. With a help of SVM model, the 64-dimensional vector was translated to its binary analog, thus reducing the time of image processing to 0.003s (300 FPS) and providing a windows that cover an area with the required object with a high probability of 96%.

This method fits well for the original idea of a human detector, which requires to quickly determine an expected location of an object and perform a fast classification based on data from a mobile robot camera. The use of BING significantly reduced a computational cost of image processing, as it allowed to use generated probability regions as a classification zone and thus not to use a sliding window on the entire image. For our method we used BING on 1280×720 input images from the robot camera to generate a set of bounding boxes. Next, the bounding boxes were ordered according to objectness scores [37] and then used together with the skin segmentation step to rearrange them based on the skin presence inside the bounding box.

3.2 Skin Segmentation

According to [10], for skin segmentation we used a pregenerated transformation matrix, which is referred as Color Attention Vectors (CAVs):

$$CAV = \begin{pmatrix} 0.5902 & -0.7543 & -0.1467 \\ -0.7543 & 1.2729 & -0.4443 \\ 0.1467 & -0.4443 & 0.2783. \end{pmatrix}. \quad (1)$$

Skin attention map (SAM) for an image was obtained by:

$$SAM(x, y) = CAV \cdot I(x, y), \quad (2)$$

where I denotes the image reshaped in $3 \times (I_{width} \cdot I_{height})$ dimension. SAM showed the affinity of a pixel to the skin region. We manually thresholded SAM to obtain skin segmentation map.

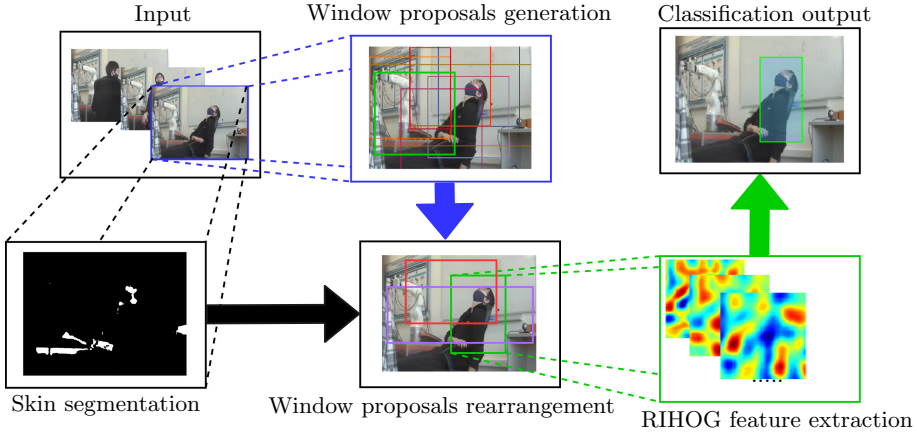


Fig. 1. Framework of the proposed image processing algorithm.

BING bounding boxes rearrangement with the help of SAM was performed in the following way: the bounding boxes that contain skin pixels did not change their order inside a queue, while other bounding boxes were transferred to a bottom of the queue. When the queue construction was completed, the bounding boxes were sent for the feature extraction algorithm.

3.3 Feature Extraction

We used square shape centroids of the rearranged bounding boxes and resized them into 64×64 cell blocks, which were then used as an input for RIHOG feature extraction process. RIHOG features encoded gradient information in rotation-invariant manner. Taking a gradient of a pixel as g , we defined an orientation distribution function at each pixel as a delta function:

$$P(\varphi) = \|g\| \delta(\varphi - \psi(g)), \quad (3)$$

where $\|g\|$ was a gradient magnitude and $\psi(g)$ represented the gradient orientation. This function was projected into Fourier space. Afterwards Fourier coefficients for a pixel were defined as follows:

$$f_m = \frac{1}{2\pi} P(\varphi) e^{-im\varphi} d\varphi = \|g\| e^{-im\psi(g)}, \quad (4)$$

where m denotes Fourier basis degree. From Eq. 4 above the gradient information could be represented by a sequence of Fourier coefficients. f_m applied to each point of the image generated a Fourier field F_m , which already described the entire image as a matrix with Fourier coefficient:

$$F_m = \frac{1}{2\pi} \|G\| e^{-im\psi(G)}, \quad (5)$$

where G denotes the gradient of the image patch.

For the next step we computed basis functions for regional descriptors using an angular part of a circular ring as Fourier basis. Thus functions for computing regional descriptors were defined as follows:

$$U_{j,k}(r, \varphi) = \wedge(r - r_j, \sigma)e^{-ik\varphi}, \quad (6)$$

where \wedge is a triangular function with a width of 2σ and defined as $\wedge(x, \sigma) = \max(\frac{\sigma - |x|}{\sigma}, 0)$, k is a degree of Fourier basis, r_j is a radius of j circular ring from a point (r, φ) .

As it was shown in the above Eq. 6, each pixel can be described by a particular number of rings that contained regional information and varied function parameters; this way an image of any size could be described at the cost of a computational loss.

Using the $U_{j,k} * F_m$ convolution operation, we created the rotation-invariant regional descriptor that described HOG values in the region covered by $U_{j,k}$. Then, to obtain vectors with values invariant to rotations, we performed the following computation:

$$\overline{(U_{j_1, k_1} * F_{m_1})(U_{j_2, k_2} * F_{m_2})}, \forall k_1 - m_1 = k_2 - m_2. \quad (7)$$

The regional descriptors created using the multiplication from the above Eq. 7 is an effective way to create many invariant features applying different parameters instead of only taking the magnitude of expansion coefficients; this allowed to increase the final feature vector. Thus, using different combinations of k and m we received a final RIHOG descriptor that was rotation-invariant and at the same time had all the HOG based descriptors advantages.

4 Experiments

4.1 Datasets

In our experiments for human detection task we used the following datasets:

1. Human-Parts dataset [39] includes 14,962 images and 106,879 annotations, which includes human body, head and hands as annotated classes.
2. MPII Human Pose dataset [3] includes around 25000 images containing over 40000 people with annotated body joints.
3. Simulated disaster victim dataset [10] consists of two parts SDV1 and SDV2. SDV1 contains 128 images and ground truth binary images with skin segmentation. SDV2 contains 15 video clips, a total of 6315 frames and 557 ground truth skin segmentation for particular frames.
4. Servosila Engineer robot dataset. The constructed by LIRS¹ dataset contains frames recorded from Servosila Engineer camera as shown on Fig. 2.

¹ Laboratory of Intelligent Robotic Systems, <https://kpfu.ru/eng/itis/research/laboratory-of-intelligent-robotic-systems>.

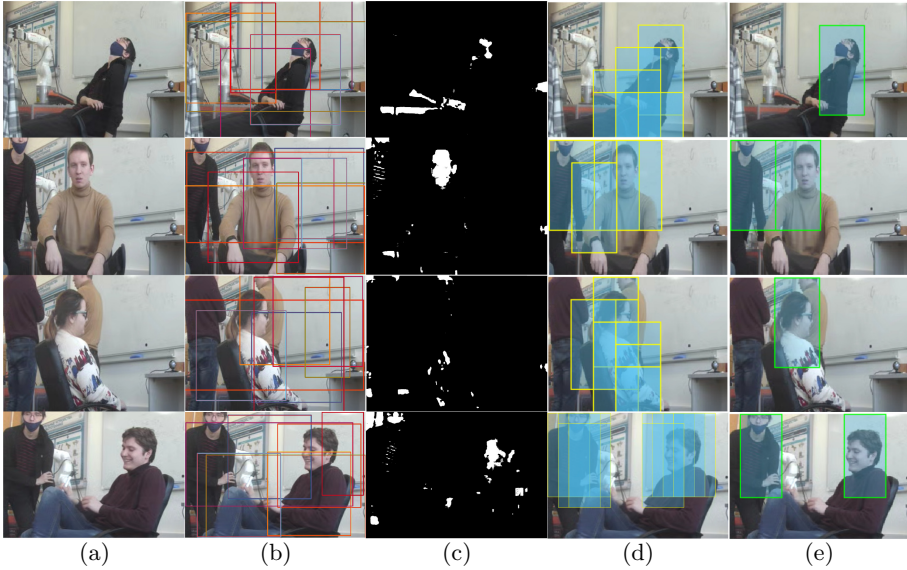


Fig. 2. Detection algorithm results with Linear SVM classifier using Servosila Engineer camera images: (a) input image, (b) BING window proposals, (c) Dadwhal et al. skin segmentation [10], (d) Positive classified bounding boxes, (e) Non-maximum suppression bounding boxes.

The datasets were used to create a new *Human body dataset*, which consist of over 60000 64×64 resolution images for machine learning algorithms. The positive part of the dataset mainly contained an upper part of a human body. The negative part focused on cluttered background segments of the images surrounding a human body.

4.2 Experiment Setup

The RIHOG extraction algorithm implementation was adapted to 64×64 size images and human detection task, some parameters were adjusted in order to meet new requirements. First five degrees of Fourier features were used $F_m : m \in [0; 4]$, for $U_{j,k}(r, \varphi)$ the following parameters used: $k \in [-4; 4]$, $\sigma = 9$, $r_j \in \{0, 9, 18, 27\}$. By combining m and k parameters in Eq. 7 on each individual ring j and coupling the values between the rings j_a and $j_b, \forall a \neq b$, we obtained a 233-dimensional final real vector that described a circular area around a single pixel.

For description of the above mentioned 64×64 images from The Human body dataset, we decided to take only 9 central pixels of those images in order to reduce the training time of various machine learning algorithms, which finally provided $233 \times 9 = 2097$ - dimensional feature vector, and since a diameter of the largest local ring was 54 pixels, which was enough for the classification purposes.

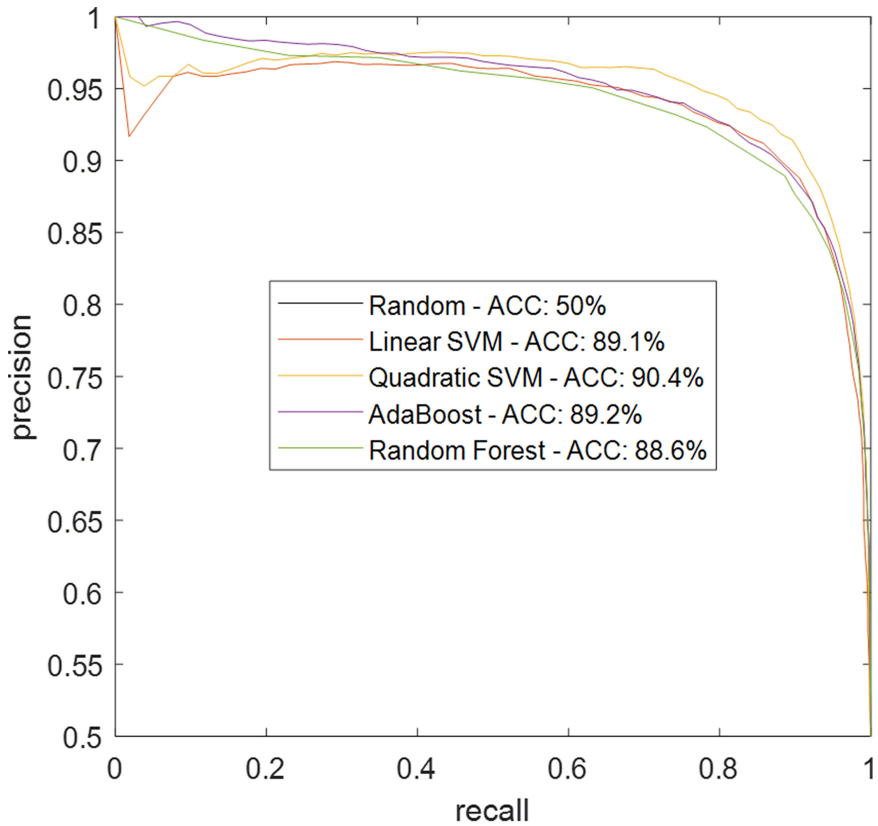


Fig. 3. PR curves and average accuracy values for the proposed algorithm.

4.3 Classification Evaluation

We trained Linear and Quadratic SVM [8], Random Forest [21] and AdaBoost [33] classifiers for performance evaluation of RIHOG features on the new data. For training we used 2097-dimensional feature vectors extracted from the Human body dataset. Those features, based on description of a human body or a background, were divided into a positive set and a negative set respectively. We used 5-fold cross-validation to evaluate detection performance of our framework, precision-recall curves and average accuracy that are shown in Fig. 3. Qualitative image examples for each step of our detection framework were captured with Servosila Engineer robot [27] camera; they are demonstrated in Fig. 2.

Validation performance showed around 88% average precision for Linear SVM, Random Forest and AdaBoost models, and reached 90.4% for the Quadratic SVM model. The results demonstrated a strong discriminative power of RIHOG features and their effectiveness particularly for a human classification with an articulated body position.

5 Conclusions

This paper presented a new human detection framework that could be integrated into a control system of a mobile robot or a stationary computer vision system. The framework uses Rotation-Invariant Histogram of Oriented Gradients (RIHOG) features to achieve independence from illumination changes, geometric transformations and orientation changes, along with BING pre-processing and skin segmentation steps to speed up the overall detection process. BING method was used for instant objectness prediction for an image to generate appropriate region proposals ordered by objectness scores. Essentially, the BING output tells the framework to which region of the image should be paid more attention. In the suggested framework the order of the region proposals was rearranged based on skin entries' intersections and used as an input to a pre-trained classifier.

For experimental evaluation a new *Human body dataset* was constructed using the Human-Parts dataset, the Simulated disaster victim dataset, and the Servosila Engineer robot dataset. The resulting Human body dataset consists of over 60000 images of 64×64 resolution that could be applied for machine learning algorithms evaluation. Random, Linear SVM, Quadratic SVM, AdaBoost, and Random Forest approaches were compared using the Human body dataset. Experimental analysis demonstrated an average precision of 90.4% for the Quadratic SVM model, while other approaches showed 88% average precision. Experimental evaluation showed the efficiency of RIHOG features as a descriptor for human detection tasks.

Acknowledgments. This work was supported by the Russian Foundation for Basic Research (RFBR), project ID 19-58-70002. The third and the fourth authors acknowledge the support of the Japan Science and Technology Agency, the JST Strategic International Collaborative Research Program, Project No. 18065977. This work is part of Kazan Federal University Strategic Academic Leadership Program.

References

1. Abbyasov, B., Lavrenov, R., Zakiev, A., Yakovlev, K., Svinin, M., Magid, E.: Automatic tool for gazebo world construction: from a grayscale image to a 3d solid model. In: 2020 IEEE International Conference on Robotics and Automation (ICRA), pp. 7226–7232. IEEE (2020)
2. Alvarez, J., Campos, G., Enríquez, V., Miranda, A., Rodriguez, F., Ponce, H.: Nurse-bot: a robot system applied to medical assistance. In: 2018 International Conference on Mechatronics, Electronics and Automotive Engineering (ICMEAE), pp. 56–59. IEEE (2018)
3. Andriluka, M., Pishchulin, L., Gehler, P., Schiele, B.: 2D human pose estimation: new benchmark and state of the art analysis. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2014)
4. Andriluka, M., et al.: Vision based victim detection from unmanned aerial vehicles. In: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1740–1747. IEEE (2010)

5. Bharath, R., Nicholas, L.Z.J., Cheng, X.: Scalable scene understanding using saliency-guided object localization. In: 2013 10th IEEE International Conference on Control and Automation (ICCA), pp. 1503–1508. IEEE (2013)
6. Chebotareva, E., Safin, R., Hsia, K.-H., Carballo, A., Magid, E.: Person-following algorithm based on laser range finder and monocular camera data fusion for a wheeled autonomous mobile robot. In: Ronzhin, A., Rigoll, G., Meshcheryakov, R. (eds.) ICR 2020. LNCS (LNAI), vol. 12336, pp. 21–33. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-60337-3_3
7. Cheng, M.M., Zhang, Z., Lin, W.Y., Torr, P.: Bing: binarized normed gradients for objectness estimation at 300fps. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3286–3293 (2014)
8. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995)
9. Crick, C., Osentoski, S., Jay, G., Jenkins, O.C.: Human and robot perception in large-scale learning from demonstration. In: Proceedings of the 6th International Conference on Human-robot Interaction, pp. 339–346 (2011)
10. Dadwhal, Y.S., Kumar, S., Sardana, H.: Data-driven skin detection in cluttered search and rescue environments. *IEEE Sens. J.* **20**(7), 3697–3708 (2019)
11. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), vol. 1, pp. 886–893. IEEE (2005)
12. Davis, M., Sahin, F.: Hog feature human detection system. In: 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 002878–002883. IEEE (2016)
13. Galin, R., Meshcheryakov, R.: Automation and robotics in the context of industry 4.0: the shift to collaborative robots. In: IOP Conference Series: Materials Science and Engineering, vol. 537, p. 032073. IOP Publishing (2019)
14. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In: Advances in Neural Information Processing Systems, pp. 545–552 (2007)
15. Hoshino, S., Niimura, K.: Robot vision system for real-time human detection and action recognition. In: Strand, M., Dillmann, R., Menegatti, E., Ghidoni, S. (eds.) IAS 2018. AISC, vol. 867, pp. 507–519. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-01370-7_40
16. Htwe, K.Y., Thein, T.L.L.: Region of interest detection based on local entropy feature for disaster victim detection system. In: 2018 IEEE 7th Global Conference on Consumer Electronics (GCCE), pp. 390–391. IEEE (2018)
17. Huang, P.R., Chu, E.T.H.: Indoor trapped-victim detection system. In: 2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI), pp. 1–6. IEEE (2017)
18. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(11), 1254–1259 (1998)
19. Kawulok, M., Kawulok, J., Nalepa, J., Smolka, B.: Self-adaptive algorithm for segmenting skin regions. *EURASIP J. Adv. Signal Process.* **2014**(1), 1–22 (2014). <https://doi.org/10.1186/1687-6180-2014-170>
20. Li, Y., et al.: Super: a surgical perception framework for endoscopic tissue manipulation with surgical robotics. *IEEE Robot. Autom. Lett.* **5**(2), 2294–2301 (2020)
21. Liaw, A., Wiener, M., et al.: Classification and regression by randomforest. *R News* **2**(3), 18–22 (2002)

22. Liu, B., Wu, H., Zhang, Y., Xu, W., Mu, K., Yao, Y.: Rihog-bovws for rotation-invariant human detection. In: IOP Conference, vol. 428, p. 10 (2018)
23. Liu, K., et al.: Rotation-invariant hog descriptors using fourier analysis in polar and spherical coordinates. *Int. J. Comput. Vis.* **106**(3), 342–364 (2014)
24. Magid, E., et al.: Artificial intelligence based framework for robotic search and rescue operations conducted jointly by international teams. In: Ronzhin A., Shishlakov V. (eds) *Proceedings of 14th International Conference on Electromechanics and Robotics “Zavalishin’s Readings”*. Smart Innovation, Systems and Technologies, vol. 154, pp. 15–26. Springer, Singapore (2020). https://doi.org/10.1007/978-981-13-9267-2_2
25. Magid, E., Tsubouchi, T.: Static balance for rescue robot navigation: discretizing rotational motion within random step environment. In: Ando, N., Balakirsky, S., Hemker, T., Reggiani, M., von Stryk, O. (eds.) *SIMPAR 2010*. LNCS (LNAI), vol. 6472, pp. 423–435. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-17319-6_39
26. Magid, E., Zakiev, A., Tsoy, T., Lavrenov, R., Rizvanov, A.: Automating pandemic mitigation. *Adv. Robot.* **35**(9), 572–589 (2021)
27. Moskvina, I., Lavrenov, R.: Modeling tracks and controller for Servosila engineer robot. In: Ronzhin, A., Shishlakov, V. (eds.) *Proceedings of 14th International Conference on Electromechanics and Robotics “Zavalishin’s Readings”*. SIST, vol. 154, pp. 411–422. Springer, Singapore (2020). https://doi.org/10.1007/978-981-13-9267-2_33
28. Pashkin, A., Lavrenov, R., Zakiev, A., Svinin, M.: Pilot communication protocols for group of mobile robots in USAR scenarios. In: 2019 12th International Conference on Developments in eSystems Engineering (DeSE), pp. 37–41. IEEE (2019)
29. Perera, A.G., Khanam, F.T.Z., Al-Naji, A., Chahl, J., et al.: Detection and localisation of life signs from the air using image registration and spatio-temporal filtering. *Remote Sens.* **12**(3), 577 (2020)
30. Ramil, S., Lavrenov, R., Tsoy, T., Svinin, M., Magid, E.: Real-time video server implementation for a mobile robot. In: 2018 11th International Conference on Developments in eSystems Engineering (DeSE), pp. 180–185. IEEE (2018)
31. Safin, R., Lavrenov, R., Martínez-García, E.A.: Evaluation of visual SLAM methods in USAR applications using ROS/Gazebo simulation. In: Ronzhin, A., Shishlakov, V. (eds.) *Proceedings of 15th International Conference on Electromechanics and Robotics “Zavalishin’s Readings”*. SIST, vol. 187, pp. 371–382. Springer, Singapore (2021). https://doi.org/10.1007/978-981-15-5580-0_30
32. Salfikar, I., Sulistijono, I.A., Basuki, A.: Automatic samples selection using histogram of oriented gradients (hog) feature distance. *EMITTER Int. J. Eng. Technol.* **5**(2), 234–254 (2017)
33. Schapire, R.E.: Explaining adaboost. In: Schölkopf, B., Luo, Z., Vovk, V. (eds.) *Empirical Inference*, pp. 37–52. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-41136-6_5
34. Simakov, N., Lavrenov, R., Zakiev, A., Safin, R., Martínez-García, E.A.: Modeling USAR maps for the collection of information on the state of the environment. In: 2019 12th International Conference on Developments in eSystems Engineering (DeSE), pp. 918–923. IEEE (2019)
35. Soni, B., Sowmya, A.: Victim detection and localisation in an urban disaster site. In: 2013 IEEE international conference on robotics and biomimetics (ROBIO), pp. 2142–2147. IEEE (2013)

36. Tsoy, T., Zakiev, A., Shabalina, K., Safin, R., Magid, E., Saha, S.K.: Validation of fiducial marker systems performance with rescue robot servosila engineer onboard camera in laboratory environment. In: 2019 12th International Conference on Developments in eSystems Engineering (DeSE), pp. 495–499. IEEE (2019)
37. Valdenegro-Toro, M.: Objectness scoring and detection proposals in forward-looking sonar images with convolutional neural networks. In: Schwenker, F., Abbas, H.M., El Gayar, N., Trentin, E. (eds.) ANNPR 2016. LNCS (LNAI), vol. 9896, pp. 209–219. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46182-3_18
38. Verner, I.M., Polishuk, A., Krayner, N.: Science class with robothespian: using a robot teacher to make science fun and engage students. IEEE Robot. Autom. Mag. **23**(2), 74–80 (2016)
39. Xiaojie Li, L.y., Qing Song, F.Z.: Detector-in-detector: multi-level analysis for human-parts. arXiv preprint arXiv:**** (2019)
40. Yildirim, G., Süssstrunk, S.: FASA: fast, accurate, and size-aware salient object detection. In: Cremers, D., Reid, I., Saito, H., Yang, M.-H. (eds.) ACCV 2014. LNCS, vol. 9005, pp. 514–528. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16811-1_34