Comparative analysis of background subtraction algorithms in the task of recognizing and identifying people in a video stream

I A Idimechev¹, A S Mikhaley¹, A A Goponenko¹, S S Stavenko¹, V V Kukartsev¹,² and V S Tynchenko¹,²

¹ Siberian Federal University, 79 Svobodny pr., Krasnoyarsk, 660041, Russian Federation
² Reshetnev Siberian State University of Science and Technology, 31 Krasnoyarsky Rabochy Av., Krasnoyarsk, 660037, Russian Federation

E-mail: vadimond@mail.ru

Abstract. The article is devoted to one of the most urgent tasks for today: preprocessing video material in the process of recognizing and identifying people in a video stream. The authors of this article conducted a study, which consists in comparing the results of several algorithms for subtracting the background on various input video materials (from the office and the minimum concentration of people, to the dining room in which there are up to ten or more people). The data obtained during the analysis of the algorithms is compared (such data are the number of frames processed per second, and the average load of the computer’s central processor). The most efficient algorithms that coped with the task most quickly, accurately and with the least expenditure of resources, were identified, as well as those that coped with the task the worst.

1. Introduction
Today, in the century of a rapid development of information technologies related to artificial intelligence [1], data analysis [2] and automation of technological processes [3], the task of identifying and tracking objects in a video stream is of particular relevance.

Machine vision technologies are widely used in security systems, production control, as well as in other areas where the use of human labor is economically unreasonable or leads to high-risk indicators [4]. However, with such an application of intelligent video processing algorithms, it is necessary to make high demands on their reliability, accuracy and efficiency. If these requirements are successfully fulfilled, the system becomes protected from the risks associated with the “human factor”.

Thus, the task of detecting the movement of objects in the video stream is relevant in the construction of video surveillance systems [5, 6].

One of the methods currently used to detect a person is to find his figure in a two-dimensional image using a predetermined model of the human body. This paper provides a comparative analysis of the algorithms for recognizing people in a video based on background subtraction. Among the considered algorithms, the most efficient one was selected, which most accurately, quickly and with less expenditure of resources coped with the task.
2. Literature review
The bulk of the research literature related to the identification and identification of people in images and videos contains a description and analysis of various algorithms for recognizing and classifying people by gender, age and race [7-9]. The tasks of identifying a person by silhouette and contour, by gait and specific patterns of behavior remain unresolved due to difficult factors (clothes and accessories that mask a person’s silhouette, circumstances that visually change gait, for example, heavy objects. Several works describe various approaches to solving these problems (using neural networks, wearable sensors, tracking the trajectories of silhouette points, etc.), the advantages of these identification methods [10, 11]. In the present work, binary silhouettes of people obtained as a result of the background subtraction process are demonstrated [12].

3. Background subtraction problem
The main stage of video processing for solving the problem of recognition, tracking and identification of people is the separation of moving foreground objects from the background. On how correctly this problem is solved, all subsequent stages of solving the task depend.

Background subtraction is a method used to segment moving areas in sequences of images taken from a static camera by comparing each new frame with a background model of the scene. The complexity of the implementation of this approach is due to the influence of a large number of various factors, such as the camera’s own noise, uneven lighting and scene movement. In this regard, the task of choosing the most robust approach becomes relevant. In this paper, the following background subtraction algorithms were chosen for the study:

- Fuzzy Sugeno Integral (with Adaptive-Selective Update).
- Pixel-Based Adaptive Segmented (PBAS).
- Gaussian Mixture Model.
- Type-2 Fuzzy GMM-UV.
- VuMeter.
- Eigenbackground / SL-PCA.

The following metrics were chosen to compare the studied algorithms:

- FPS (frame rate) of the algorithm - an estimate of the speed of the algorithm.
- Average CPU utilization in percent.

4. Background subtraction algorithms
Fuzzy Sugeno Integral (with Adaptive-Selective Update) [6]. This method is based on using the fuzzy Sugeno integral to combine texture and color features to subtract the background. It is usually used to process scenes with slight movement of background objects. It does not require large computational costs.

Pixel-Based Adaptive Segmented (PBAS) [13]. This approach is based on the paradigm of nonparametric modeling of the background. The background is modeled based on a history of recently observed pixel values. The definition of the foreground depends on the decision threshold. The background update is based on a learning option. Both of these parameters are expanded to dynamic variables describing the state of the pixel, and dynamic controllers are presented for each of them. In addition, both controllers are driven by background dynamics.

Gaussian Mixture Model [14]. The method simulates each pixel as a mixture of Gaussians and also uses real-time approximation to update the model. Gaussian distributions are evaluated to determine which ones are most likely as a background. Each pixel is classified based on whether the Gaussian distribution representing it is part of the background model. The classification is based on the constancy and dispersion of each of the mixture of Gaussians. Pixel values that do not match the background distribution are considered the foreground until a Gaussian is found that includes them.
This algorithm works stably in real time when lighting changes, it also copes with repetitive movements and long-term scene changes.

Type-2 Fuzzy GMM-UV with MRF [5]. This algorithm is based on two other algorithms: Type-2 Fuzzy Gaussian Mixture Model (T2-FGMM) and Markov Random Field (MRF). T2-FGMM does not take into account space-time constraints, this leads to poor performance for dynamic scenes. MRF helps to cope with this problem. Type-2 Fuzzy Algorithm GMM-UV with MRF is a new background modeling technique for detecting motion in dynamic scenes. The key idea of the proposed approach is the successful introduction of space-time constraints in T2-FGMM using the Bayesian structure.

VuMeter. This method belongs to the class of nonparametric. The mathematical model is based on a discrete estimate of the probability distribution. The key aspects of this method are the probabilistic model and temporary updating. This is a probabilistic approach to determining the background image model.

Eigenbackground / SL-PCA [4]. In this method, moving objects are detected using adaptively adjusting their own space, modeling the background. This model of intrinsic space describes the range of changes in appearance (for example, changes in lighting during the day, weather changes, etc.) that have been noticed.

5. Description of computing experiment
We used the BGSLibrary library written in the C++ programming language to study the algorithms considered in this work.

Computer features:

- Processor: AMD A10-9620P RADEON R5, 10 COMPUTE CORES 4C + 6G 2.50 GHz.
- Installed memory (RAM): 6.00 GB.
- System type: Windows 10, 64-bit OS, x64 processor.

Video materials with various shooting conditions were found to compare the selected algorithms: changing lighting; shooting from various angles.

These conditions allow more full evaluation of the algorithms, because depending on the lighting and shooting angle, additional noises appear due to the shadow falling from objects and the quality of the room’s illumination. The results of the algorithms are presented in figures 1-6.

![Figure 1. Fuzzy Sugeno Integral.](image)
Figure 1 shows the result of the Fuzzy Sugeno Integral algorithm. In the case of shooting in a poorly lit room on the frames obtained, the silhouette of a person is visually almost indistinguishable. This is due to the fact that the contrast of the background and the moving object in low light is very low. In frames with good lighting, the silhouette of a person is distinguishable, but in addition to the person, the algorithm highlights light flares on the glossy surface of the floor. Also, when processing frames where several people in dark clothes are on a dark background, the algorithm inaccurately distinguishes the silhouette of people.

Figure 2. Pixel-Based Adaptive Segmenter.

The result of the operation of the Pixel-Based Adaptive Segmenter algorithm is shown in figure 2. In low light conditions, the algorithm distinguishes a moving person; its silhouette is visually clearly distinguishable. Under normal lighting, in addition to a moving person, a shadow stands out, which falls from him. People silhouettes are visually recognizable. If a person remains motionless for some time, he becomes invisible to the algorithm.

Figure 3. Gaussian Mixture Model.
Figure 3 shows the result of the Gaussian Mixture Model algorithm. For the case of shooting with poor lighting from the above frames, we can conclude that the silhouette of a person in motion is quite clearly distinguishable. But in addition to the moving person, the algorithm highlights its shadow. In well-lit rooms, the algorithm also highlights the distinguishable silhouette of a moving person and the shadow falling from it. At the same time, the algorithm does not see people who are not moving. In the bottom set of screenshots, we can see that the algorithm emits a highlight on the floor surface.

![Figure 3](image1.png)

**Figure 3.** The result of the Gaussian Mixture Model algorithm.

Frames that reflect the result of the Type-2 Fuzzy GMM-UV with MRF algorithm are shown in figure 4. This algorithm does not provide an exact selection of a moving object. In poor lighting, the silhouette of a moving person is not always distinguished, in some cases; the algorithm almost does not see a moving object. In the case of normal lighting, the silhouettes of people stand out inaccurately. It is almost impossible to understand that a person is highlighted in the frame. In some cases, moving people remain completely invisible to the algorithm.

![Figure 4](image2.png)

**Figure 4.** Type-2 Fuzzy GMM-UV with MRF.

Frames that reflect the result of the Type-2 Fuzzy GMM-UV with MRF algorithm are shown in figure 5. This algorithm does not provide an exact selection of a moving object. In poor lighting, the silhouette of a moving person is not always distinguished, in some cases; the algorithm almost does not see a moving object. In the case of normal lighting, the silhouettes of people stand out inaccurately. It is almost impossible to understand that a person is highlighted in the frame. In some cases, moving people remain completely invisible to the algorithm.

![Figure 5](image3.png)

**Figure 5.** Eigenbackground / SL-PCA (corridor, turn off the lights).
For the Eigenbackground / SL-PCA algorithm, the result of the work is shown in Figure 6. In the case when the lighting is poor in the room, the silhouette of a person highlighted by this algorithm is easily distinguishable. But at the moment of lighting change, the algorithm stops working correctly. He highlights almost a whole room as a moving object. This is a drawback, since a change in lighting will greatly affect the quality of this algorithm. For shooting under normal lighting, in addition to a moving subject, this algorithm emits shadows that fall from objects in the room, as well as glare of light on the floor.

An analysis of the results allows concluding that the Pixel-Based Adaptive Segmenter algorithm works most efficiently.

The following is a comparison of the algorithms for the selected metrics. The obtained metric values are presented in figure 6.

![Figure 6. Diagrams of obtained metric values: a) Average CPU usage when running the algorithm, % CPU; b) Average number of frames processed per second, FPS.](image)

**6. Conclusion**

The PBAS algorithm, which is more efficient in unstable shooting, is the most resource-intensive. The algorithms of the Gaussian Mixture Model and Type-2 Fuzzy GMM-UV with MRF in terms of resource consumption differ quite strongly (more than two times in terms of FPS and almost twice in terms of CPU load), but both of them have average indicators in comparison with other algorithms.

That is, these algorithms are more suitable than others for use in security systems, since with average efficiency in terms of resource cost they give the best result in terms of accuracy.

**References**

[1] Milov A V, Tynchenko V S, Kukartsev V V, Tynchenko V V and Bukhtoyarov V V 2018 Use of artificial neural networks to correct non-standard errors of measuring instruments when creating integral joints Journal of Physics: Conference Series **1118**(1) 012037

[2] Bukhtoyarov V V, Tynchenko V S, Petrovsky E A, Kukartsev V V and Kuklina A I 2018 Evolutionary method for automated design of models of vortex flowmeters transformation function Journal of Physics: Conference Series **1118**(1) 012041

[3] Tynchenko, V., Milov, A., Tynchenko, V., Bukhtoyarov, V., Antamoshkin, O. 2018 Application of artificial neural networks for identification of non-normative errors in measuring instruments for controlling the induction soldering process In International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management **18**(2.1) 117-24
[4] Oliver N, Rosario B and Pentland A 1999 A Bayesian computer vision system for modeling human interactions In *International Conference on Computer Vision Systems* pp. 255-72
[5] Zhao Z, Bouwmans T, Zhang X and Fang Y 2012 A fuzzy background modeling approach for motion detection in dynamic backgrounds In *International Conference on Multimedia and Signal Processing* pp. 177-85
[6] Baf F E, Bouwmans T and Vachon B 2008 Fuzzy integral for moving object detection In *2008 IEEE International Conference on Fuzzy Systems* 1729-36
[7] Saparkhojayev N and Akhmetov Y 2014 Human identification in video streaming based on some facial classification parameters *Middle-East J. Sci. Res.* 21(7) 1153-6
[8] Akhmetzhanov K R, Sazanov V I and Lipin Yu N 2017 Human face recognition software package *Scient. J. of Ufa State Aviation Technical University* 3(21) 79-86
[9] Trofimenko Ya M, Vinogradova L N and Yershov E V 2019 Algorithms for image preprocessing in the system of face identification in the video stream *Bulletin of Cherepovets state University* 4(91) 21-9
[10] Sokolova A and Konushin A 2019 Human identification by gait from event-based camera In *2019 16th International Conference on Machine Vision Applications* pp. 1-6
[11] Kazantseva A G 2013 Human gait identification using wearable sensors *Research review* 2(28) 117-21
[12] Piccardi M 2004 Background subtraction techniques: a review In: *Int. Conf. on systems, man and cybernetics* pp. 3099–104
[13] Hofmann M, Tiefenbacher P, and Rigoll G 2012 Background segmentation with feedback: The Pixel-Based Adaptive Segmenter In *Computer Society Conf. on Computer Vision and Pattern Recognition Workshops* pp. 38–43
[14] Stauffer C and Grimson W E L 1999 Adaptive background mixture models for real-time tracking In *1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2 246-52