International Journal of Informatics and Applied Mathematics
e-ISSN:2667-6990 Vol. 2, No. 1, 73-84

Background subtraction based on a Self-Adjusting MoG

Wafa Nebili, Samir Hallaci, and Brahim Farou

LabSTIC Laboratory ,Computer Science department, 8 May 1945 Guelma University, BP 401 Guelma, Algeria
nebili.wafa@univ-guelma.dz
hallaci.samir.hamid@univ-guelma.dz
farou.brahim@univ-guelma.dz

Abstract. The diversity in background scenes such as, illumination changes, dynamics of the background, camouflage effect, shadow, etc. is a big deal for moving objects detection methods makes it impossible to manage the multimodality of scenes in video surveillance systems. In this paper we present a new method that allows better detection of moving objects. This method combine the robustness of the Artificial Immune Recognition System (AIRS) with respect to the local variations and the power of Gaussian mixtures (MoG) to model changes at the pixel level. The task of the AIRS is to generate several MoG models for each pixel. This models are filtred through two mecanism: the competition for resources and the development of a candidate memory cell. The best model is merged with the exesting MoG according to the Memory cell introduction process. Obtained results on the Wallflower dataset proved the performance of our system compared to other state-of-the-art methods.

Keywords: Background Subtraction· MoG· AIRS · Video Surveillance · Pixel Classification · Foreground Segmentation.

1 Introduction

Background subtraction (BS) represented a key step for applications related to automatic processing of video data, since it is necessary to detect moving and static objects before doing more complex operations such as tracking, event analysis, etc. During the background subtraction process, each pixel of a video sequence is labeled [6], for example: pixels of moving objects (foreground) take the value 255, on the other hand the value zero is given to pixels of static objects.

In the recent years, many methods and techniques have been proposed to effectively separate the foreground from the background. The most intuitive method is to calculate the absolute difference (Δ_t) either between two successive frames [6], or between the current frame and a reference background frame IR. To define pixel nature, a binary mask is applied according to a predefined threshold on the output frame pixels [10].

Another way to subtract the background is to describe the history of the last n pixel values by a Gaussian probability distribution [30]. However, modeling using a single Gaussian is sensitive to fast pixel variations. Indeed, a single Gaussian cannot memorize the old states of the pixel. This requires migration to a robust and multimodal approach. Authors in [11] proposed the first model which describes the variance of the recent values of each pixel by a mixture of the Gaussians. In this model, the Expectation Maximization (EM) algorithm is used to initialize and estimate the parameters of each Gaussian. An improvement of this version with efficient update equations was proposed by [23]. Several works and contributions have been proposed to improve the quality of MoG. Some of them focused on improving the model adaptation speed such as: [21] [16]. While others are interested on hybrid models such as: MoG and K-means [5], MoG and fuzzy logic [8], MoG and adaptive background [7], Markov Random Fields [22], MoG and Block matching [10], MoG with PSO [29] and MoG with correlation coefficient [26] to overcome MoG problems.

Authors in [13] provided a nonparametric estimation of the background pattern. They used the concept of a visual dictionary words to model the pixels of the background. Indeed, each pixel of the frame is represented by a set of three values (visual word) which describes its current state. These values are initially estimated during the learning phase and are updated regularly over time to build a robust modeling. In the same context, Elgammal et al. [9] used Kernel Density Estimator (KDE) of the N recent values of each pixel to estimate the background model.

Several works have taken spatial information into consideration. The first technique in this context was proposed by Oliver et al. [20], this letter are used the principal component analyses (PCA) to create a robust model of background. To determinate, the foreground pixels, an absolute difference is calculated between the current frame and the reconstructed frame from its projection in the reduced dimension space. Tsai and Lai provided in [25] a quick schema (SL-ICA) for background subtraction with Independent Component Analysis (ICA). Another work of [4] used an incremental non-negative matrix factorization (INMF) to decompose video content.

In addition to deep learning methods, some approaches are interested on the selection and the combination of several features (colors, texture, edges) to improve the results quality of background subtraction, among these methods: [1] [15] [14] [17] [32].

In the last years, many methods has been introduced deep learning to separate static and dynamic objects, among them we cite: FgSegNet_S (FPM) [19], Cascade CNN [27], DeepBS [2], Deep background subtraction with scene-specific convolutional neural networks [3]. However, deep leaning methods require a large number of simples and needs more time for training.

Works done in [31] show that the MoG offers a good compromise between quality and execution time compared to other methods. However, this method is sensitive to illumination changes and camouflaged areas. These problems are related to the nature of MoG model. Our work consists to describing a new approach for modeling the background using another mechanism to update MoG model in the system.
Initially, the system begins with a single MoG in the learning phase. Then, for each background pixel we created several MoGs through the process of Memory cell identification and ARB generation of the AIRS algorithm [28]. Created models are filtered according to the Competition for resources and the development of a candidate memory cell of the AIRS. This mechanism allows to choose only the best models which will be used to select a single candidate model. This model can be merged with the old Gaussian Mixture Model with the Memory cell introduction process, this mechanism can participate in model diversity and can generate a strong model to pixel classification.

2 Proposition

Recently, MoG approach has achieved considerable success in moving objects detection for video surveillance systems. However, this method has some drawbacks due to the nature of the model used in background subtraction. Indeed, the old MoG model is not enough to describe pixel variations. From this principle, we have proposed a new mechanism based on AIRS algorithm to update the MoG model for each pixel according to the environment changes. Indeed, AIRS mechanism allows us to add and create new MoG models that can describe and predict states that can take a pixel.

Firstly, our system is initialized by a single MoG model for each pixel. The latter is updated during the learning phase like that indicate in the standard MoG (see [23]). We used H component of HSV color space to characterize each pixel. The choice of HSV space was based on the capacity of this space compared to the RGB space since it allows to channel the light into a single component (V), therefore, the brightness affects only on the element V and not on the component H, which allows to reduce variations related to light. Furthermore, this model is the closest model of human perception [12].

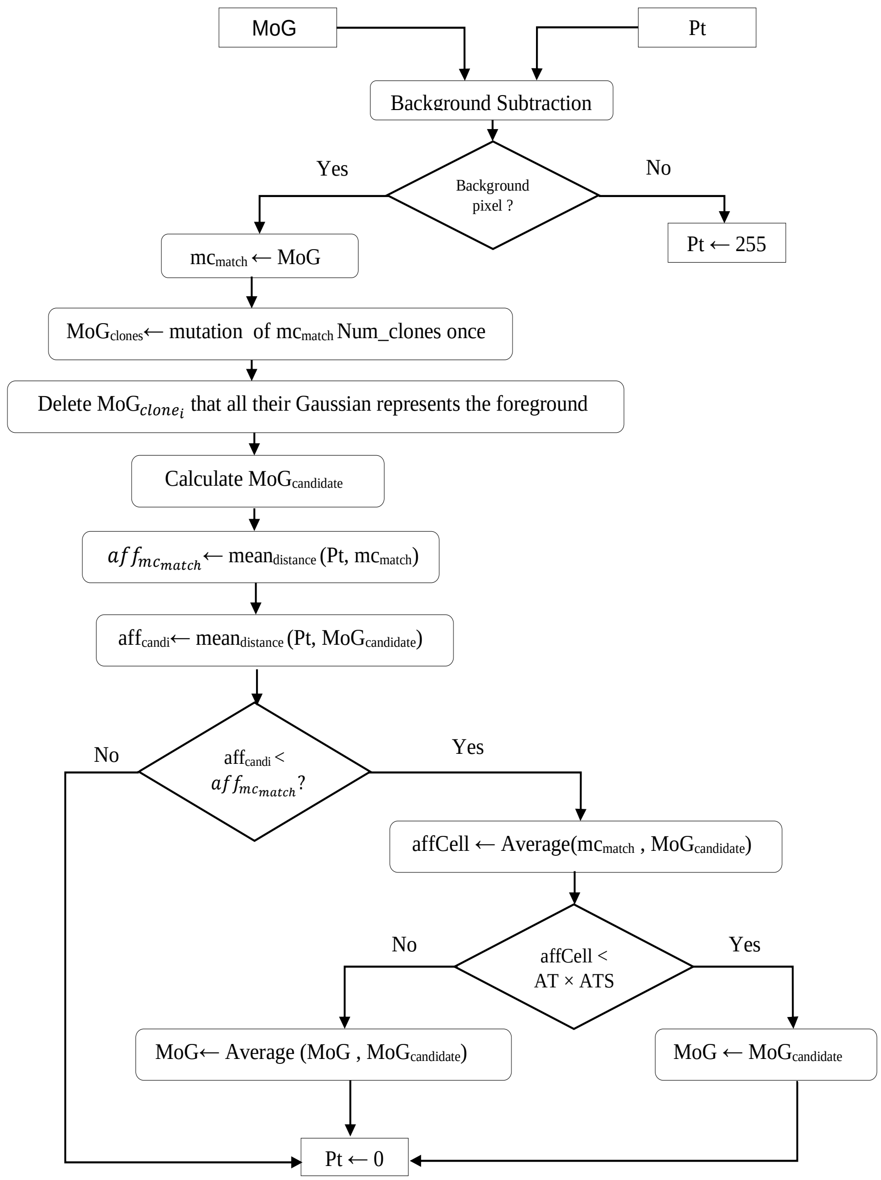

Fig. 1: Flowchart of the proposed system

Each gaussian g_i in MoG model is represented by: the pixel value P_t, the average u_i, the variance σ_i and the weight w_i.

$$g_i = \{P_t, u_i, \sigma_i, w_i\} \tag{1}$$

Algorithm 1 Proposed algorithm

Require: P_tPixel value , MoG : Gaussian mixture model
Ensure: 0 or 255
  state $\leftarrow$applied MoG approach to determine the nature of P_t
  if $state == 255$ **then**
    $P_t \leftarrow state$
  else
    $mc_{match} \leftarrow MoG$
    $affmc_{match} \leftarrow mean_{g_i \in MoG}\left(\frac{P_t - u_{g_i}}{\sigma_{g_i}} < 2.5\right)$
    $Num\_clones \leftarrow CR \times HCR \times affmc_{match}$
    $MoG_{clones} \leftarrow \emptyset$
    while $|MoG_{clones}| < Num\_clones$ **do**
      $tr \leftarrow False$
      $mc_{clone} \leftarrow mc_{match}$
      $tr, mc_{clone} \leftarrow Mutation(mc_{clone}, MR, tr)$
      if $tr == True$ **then**
        $MoG_{clones} \leftarrow MoG_{clones} \cup mc_{clone}$
      end if
    end while
    for each $MoG_{clone_i} \in MoG_{clones}$ **do**
      if $All_{g_j \in MoG_{clone_i}}\left(\frac{P_t - u_{g_j}}{\sigma_{g_j}} > 2.5\right)$ **then**
        $MoG_{clones} \leftarrow MoG_{clones} - MoG_{clone_i}$
      end if
    end for
    $MoG_{candidate} = argmax_{MoG_{clone_i} \in MoG_{clones}}\left(\frac{P_t - u_{g_j}}{\sigma_{g_j}} < 2.5\right)$
    $aff_{candi} \leftarrow mean_{g_i \in MoG_{candidate}}\left(\frac{P_t - u_{g_i}}{\sigma_{g_i}} < 2.5\right)$
    if $affmc_{match} < aff_{candi}$ **then**
      $affCell \leftarrow \frac{mc_{match} + MoG_{cadidate}}{2}$
      if $affCell < AT \times ATS$ **then**
        $MoG \leftarrow MoG_{candidate}$
      else
        $MoG \leftarrow avrage(MoG, MoG_{candidate})$
      end if
    end if
    $P_t \leftarrow 0$
  end if

After creating the background model, our system begins pixel classification phase. To classify the pixels into background or foreground, the Gaussians of the MoG model will be ordered according to the value of $w_{k,t}/\sigma_{k,t}$. The Gaussians that represent the state of P_t is the B first distribution that satisfies the equation 2.

$$\beta = argmin(\sum_{k=1}^{b} w_{k,t} > B) \tag{2}$$

Where :
B : Determines the minimum part of the data corresponding to the background.
b : The number of Gaussian in the MoG model.
If the pixel represente the background, we take their MoG model as the memory cell mc_{match}.
The mc_{match} will be mutated by a Mutation Rate (MR) in the ARB generation phase of the AIRS algorithm. The mutation is applied at the Gaussians g_i that satisfies the equation sited in the model of [23])):

$$\frac{P_t - u_i}{\sigma_i} < 2.5 \tag{3}$$

At the end of this phase, a set of MoG models is created (MoG_{clones}).

$$MoG_{clones} = \{MoG_{clone_1}, .., MoG_{clone_{Num\_clones}}\} \tag{4}$$

With :

$$MoG_{clone_i} = Mutation(mc_{match}) \tag{5}$$

The number of clones is calculated by the following equation :

$$Num\_clones = CR \times HCR \times distance(P_t, mc_{match}) \tag{6}$$

Clonal Rate (CR) and Hyper Clonal Rate (HCR) are two integer values chosen by the user.
Note that the distance between P_t and mc_{match} is the average of the distances between the pixel P_t and the Gaussians that satisfies equation 3.
All new clones (MoG_{clones}) will be filtered by Competition for resources and development of a candidate memory cell process, keeping only the best MoG_{clone_i} in whose Gaussian mutated g_j remains satisfies equation 3.
After this step, we will choose from the remains of MoG_{clones} set a single MoG the most similar and the most closest to the current pixel P_t according to equation 7.

$$MoG_{candidate} = argmax_{MoG_{clone_i} \in MoG_{clones}} \left(\frac{P_t - u_{g_j}}{\sigma_{g_j}} < 2.5 \right) \tag{7}$$

The last step in our process is to introduce a new MoG model using Memory cell introduction process of the AIRS algorithm. This step consists of adding the $MoG_{candidate}$ to all background models. The $MoG_{candidate}$ is accepted as a new model if it verifies the following equation:

$$mean_{distance}(P_t, MoG_{candidate}) < mean_{distance}(P_t, mc_{match}) \tag{8}$$

The mean distance is calculated between the pixel P_t and the Gaussians that satisfies equation 3.
If equation 8 is satisfied, the average of the two previous distances is compared with the value of the affinity threshold AT multiplied by affinity threshold scalar ATS.

$$Average(mc_{match}, MoG_{candidate}) < AT \times ATS \tag{9}$$

With :
AT : The average distance of all background models generated in the learning phase.
ATS : A value between 0 and 1 chosen by the user.
If equation 9 is satisfied the mc_{match} will be removed from all background models. If equation 8 is not satisfied, mc_{match} will be merged with the existing MoG, such that the new MoG is the avrage of the existing MoGs and mc_{match}.

3 Tests and results

The system presented in this paper is implemented in *Python* on a computer with an Intel Core $i7$ and a $8GB$ memory capacity.
This section presents experimental results obtained by our method on some videos from the Wallflower dataset [24]. Our results are compared to the obtained results of other methods cited in literature works.
Wallflower is a public dataset containing 7 videos (Moved Object (MO), Time of Day (TD), Light Switch (LS), Waving Trees (WT), Camouflage (Ca), Bootstrap (Bo), Foreground Aperture (FA)) with a resolution of 160×120. To ensure the stability of our system during the test phase, the values of (learinig rate α, the minimum part of the data corresponding to the background B, Number of Gaussians in a MoG model b, Hyper Mutation Rate HMR, Clonal Rate CR, ATS, Mutation Rate MR) are fixed, after several empirical tests, respectively to (0.001, 0.3, 5, 10, 2, 0.2, 0.1). Qualitative results do not allow to get a in depth conclusions on system performance. For this, we calculated the number of errors (false positive, false negative) in each video.

Table 1: Description of the Wallflower datasets.

Video name	Number of frames	Resolution	Image evaluated
Moved Object	1745	160×120	00985
Time of Day	5890	160×120	01850
Light Switch	2715	160×120	01865
Waving Trees	243	160×120	00247
Camouflage	281	160×120	00251
Bootstrap	3055	160×120	00299
Foreground Aperture	2113	160×120	00489

With :

- False negative (FN): The result is negative(0), but the ground truth is positive (255).
- False positive (FP): The result is positive(255), but the ground truth is negative(0).

Table 2: Qualitative results on Wallflower dataset

In addition to false negative and false positive, we also used three other metrics to evaluate the performance of our method. This metrics are calculated using the following formulas:

1. Recall (Re) : $\frac{TP}{TP+FN}$
2. Precision (Pre) : $\frac{TP}{TP+FP}$
3. F_measure : $\frac{2\times Precision\times Recall}{Precision+Recall}$

Our method archived good results compared to other state of the art methods occupying the first place with a total error of 9055. However, the proposed system has failed to solve the problems related to camouflaged areas due to the

Table 3: Quantitative results on Wallflower dataset.

	Error	MO	TD	LS	WT	Ca	Bo	FA	Total errors
SG [30]	FN	0	949	1857	3110	4101	2215	3464	35133
	FP	0	535	15123	357	2040	92	1290	
MOG [23]	FN	0	1008	1633	1323	398	1874	2442	27053
	FP	0	20	14169	341	3098	217	530	
KDE [9]	FN	0	1298	760	170	238	1755	2413	26450
	FP	0	125	14153	589	3392	933	624	
SL-ICA [25]	FN	0	1199	1557	3372	3054	2560	2721	15308
	FP	0	0	210	148	43	16	428	
SL-INMF [4]	FN	0	724	1593	3317	6626	1401	3412	19098
	FP	0	481	303	652	234	190	165	
SL-IRT [18]	FN	0	1282	2822	4525	1491	1734	2438	17053
	FP	0	159	389	7	114	2080	12	
Proposed	FN	0	1024	950	438	2164	1115	336	9055
	FP	0	1204	370	45	2	997	410	

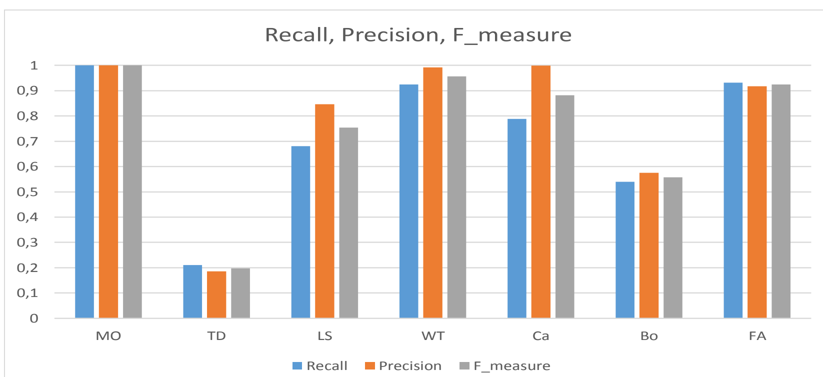

Fig. 2: Recall, Precision, F_measure of the proposed system on Wallflower dataset

nature of the features vector used.

The system can achieve more efficient results by adding other features. One feature remains insufficient for background modeling. We have only used the H component, since our objective in this work is to propose a new method of background subtraction and not selecting the good discriminator features.

The observable results clearly show that our system has obtained a good detection rate, since it has detected all moving objects with some false negative in the Bootstrap and Time of Day videos, this is due to the nature of videos. Indeed,

Table 4: Recall, Precision and F_measure of background subtraction methods on Wallflower dataset

Algorithm	Performance criteria	MO	TD	LS	WT	Ca	Bo	FA
SG	Recall	1.000	0.949	0.545	0.835	0.761	0.884	0.807
	Precision	1.000	0.971	0.128	0.978	0.865	0.995	0.918
	F_measure	1.000	0.960	0.207	0.901	0.810	0.936	0.859
MOG	Recall	1,000	0,947	0,675	0,930	0,975	0,901	0,869
	Precision	1,000	0,999	0,193	0,981	0,835	0,987	0,968
	F_measure	1,000	0,972	0,301	0,955	0,900	0,942	0,916
KDE	Recall	1.000	0.932	0.849	0.991	0.985	0.904	0.870
	Precision	1.000	0.993	0.232	0.969	0.821	0.947	0.963
	F_measure	1.000	0.962	0.365	0.980	0.896	0.925	0.914
SL-ICA	Recall	1.000	0.938	0.918	0.823	0.841	0.867	0.855
	Precision	1.000	1.000	0.988	0.991	0.997	0.999	0.974
	F_measure	1.000	0.968	0.952	0.899	0.912	0.928	0.911
SL-INMF	Recall	1.000	0.961	0.916	0.821	0.651	0.926	0.821
	Precision	1.000	0.974	0.983	0.959	0.981	0.989	0.990
	F_measure	1.000	0.968	0.948	0.885	0.782	0.957	0.897
SL-IRT	Recall	1.000	0.933	0.850	0.764	0.922	0.899	0.993
	Precision	1.000	0.991	0.976	1.000	0.994	0.881	0.880
	F_measure	1.000	0.961	0.909	0.866	0.956	0.890	0.933
Proposed	Recall	1.000	0.211	0.681	0.925	0.789	0.540	0.932
	Precision	1.000	0.185	0.846	0.992	0.999	0.575	0.918
	F_measure	1.000	0.197	0.754	0.957	0.882	0.557	0.925

Boostrap video does not contain a sufficient number of samples for learning the system.

4 Conclusion

In this paper, we have presented a new approach for background subtraction. The idea is to update the MoG model by the AIRS algorithm instead of updating only like it indicate in the basic MoG.
The obtained results on Wallflower public dataset showed the effectiveness of our approach in videos with small variations of the background. It should also be noted that this method has allowed us to treat MoG problems in scenes where the change in brightness is very fast.
As future work and to overcome the drawbacks of this system, we will focus our study on selecting better features and applying this method to other datasets.

References

1. Azab, M.M., Shedeed, H.A., Hussein, A.S.: A new technique for background modeling and subtraction for motion detection in real-time videos. In: Image Processing (ICIP), 2010 17th IEEE International Conference on. pp. 3453–3456. IEEE (2010)
2. Babaee, M., Dinh, D.T., Rigoll, G.: A deep convolutional neural network for video sequence background subtraction. Pattern Recognition **76**, 635–649 (2018)
3. Braham, M., Van Droogenbroeck, M.: Deep background subtraction with scene-specific convolutional neural networks. In: 2016 international conference on systems, signals and image processing (IWSSIP). pp. 1–4. IEEE (2016)
4. Bucak, S.S., Gunsel, B.: Video content representation by incremental non-negative matrix factorization. In: Image Processing, 2007. ICIP 2007. IEEE International Conference on. vol. 2, pp. II–113. IEEE (2007)
5. Charoenpong, T., Supasuteekul, A., Nuthong, C.: Adaptive background modeling from an image sequence by using k-means clustering. In: Electrical Engineering/Electronics Computer Telecommunications and Information Technology (ECTI-CON), 2010 International Conference on. pp. 880–883. IEEE (2010)
6. Collins, R.T., Lipton, A.J., Kanade, T., Fujiyoshi, H., Duggins, D., Tsin, Y., Tolliver, D., Enomoto, N., Hasegawa, O., Burt, P., et al.: A system for video surveillance and monitoring. VSAM final report pp. 1–68 (2000)
7. Doulamis, A., Kalisperakis, I., Stentoumis, C., Matsatsinis, N.: Self adaptive background modeling for identifying persons' falls. In: Semantic Media Adaptation and Personalization (SMAP), 2010 5th International Workshop on. pp. 57–63. IEEE (2010)
8. El Baf, F., Bouwmans, T., Vachon, B.: Fuzzy statistical modeling of dynamic backgrounds for moving object detection in infrared videos. In: Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on. pp. 60–65. IEEE (2009)
9. Elgammal, A., Harwood, D., Davis, L.: Non-parametric model for background subtraction. In: European conference on computer vision. pp. 751–767. Springer (2000)
10. Farou, B., Kouahla, M.N., Seridi, H., Akdag, H.: Efficient local monitoring approach for the task of background subtraction. Engineering Applications of Artificial Intelligence **64**, 1–12 (2017)
11. Friedman, N., Russell, S.: Image segmentation in video sequences: A probabilistic approach. In: Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence. pp. 175–181. Morgan Kaufmann Publishers Inc. (1997)
12. Haq, A.u., Gondal, I., Murshed, M.: Automated multi-sensor color video fusion for nighttime video surveillance. In: Computers and Communications (ISCC), 2010 IEEE Symposium on. pp. 529–534. IEEE (2010)
13. Haritaoglu, I., Harwood, D., Davis, L.S.: W4: Real-time surveillance of people and their activities. IEEE Transactions on Pattern Analysis & Machine Intelligence (8), 809–830 (2000)
14. Jain, V., Kimia, B.B., Mundy, J.L.: Background modeling based on subpixel edges. In: Image Processing, 2007. ICIP 2007. IEEE International Conference on. vol. 6, pp. VI–321. IEEE (2007)
15. Jian, X., Xiao-qing, D., Sheng-jin, W., You-shou, W.: Background subtraction based on a combination of texture, color and intensity. In: Signal Processing, 2008. ICSP 2008. 9th International Conference on. pp. 1400–1405. IEEE (2008)
16. KaewTraKulPong, P., Bowden, R.: An improved adaptive background mixture model for real-time tracking with shadow detection. In: Video-based surveillance systems, pp. 135–144. Springer (2002)

17. Kristensen, F., Nilsson, P., Öwall, V.: Background segmentation beyond rgb. In: Asian COnference on Computer Vision. pp. 602–612. Springer (2006)
18. Li, X., Hu, W., Zhang, Z., Zhang, X.: Robust foreground segmentation based on two effective background models. In: Proceedings of the 1st ACM international conference on Multimedia information retrieval. pp. 223–228. ACM (2008)
19. Lim, L.A., Keles, H.Y.: Foreground segmentation using a triplet convolutional neural network for multiscale feature encoding. arXiv preprint arXiv:1801.02225 (2018)
20. Oliver, N.M., Rosario, B., Pentland, A.P.: A bayesian computer vision system for modeling human interactions. IEEE transactions on pattern analysis and machine intelligence **22**(8), 831–843 (2000)
21. Power, P.W., Schoonees, J.A.: Understanding background mixture models for foreground segmentation. In: Proceedings image and vision computing New Zealand. vol. 2002 (2002)
22. Schindler, K., Wang, H.: Smooth foreground-background segmentation for video processing. In: Asian Conference on Computer Vision. pp. 581–590. Springer (2006)
23. Stauffer, C., Grimson, W.E.L.: Adaptive background mixture models for real-time tracking. In: cvpr. p. 2246. IEEE (1999)
24. Toyama, K., Krumm, J., Brumitt, B., Meyers, B.: Wallflower: Principles and practice of background maintenance. In: Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on. vol. 1, pp. 255–261. IEEE (1999)
25. Tsai, D.M., Lai, S.C.: Independent component analysis-based background subtraction for indoor surveillance. IEEE Transactions on image processing **18**(1), 158–167 (2009)
26. Wafa, N., Hamid, S., Nadjib, K.M.: A new process for selecting the best background representatives based on gmm. International Journal of Informatics and Applied Mathematics **1**(1), 35–46
27. Wang, Y., Luo, Z., Jodoin, P.M.: Interactive deep learning method for segmenting moving objects. Pattern Recognition Letters **96**, 66–75 (2017)
28. Watkins, A., Boggess, L.: A new classifier based on resource limited artificial immune systems. In: Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No. 02TH8600). vol. 2, pp. 1546–1551. IEEE (2002)
29. White, B., Shah, M.: Automatically tuning background subtraction parameters using particle swarm optimization. In: Multimedia and Expo, 2007 IEEE International Conference on. pp. 1826–1829. IEEE (2007)
30. Wren, C.R., Azarbayejani, A., Darrell, T., Pentland, A.P.: Pfinder: Real-time tracking of the human body. IEEE Transactions on pattern analysis and machine intelligence **19**(7), 780–785 (1997)
31. Yu, J., Zhou, X., Qian, F.: Object kinematic model: A novel approach of adaptive background mixture models for video segmentation. In: Intelligent Control and Automation (WCICA), 2010 8th World Congress on. pp. 6225–6228. IEEE (2010)
32. Zhang, H., Xu, D.: Fusing color and texture features for background model. In: Fuzzy Systems and Knowledge Discovery: Third International Conference, FSKD 2006, Xian, China, September 24-28, 2006. Proceedings 3. pp. 887–893. Springer (2006)