Complete Moving Object Detection in the Context of Robust Subspace Learning

Maryam Sultana[1], Arif Mahmood[2], Thierry Bouwmans[3], Soon Ki Jung[1]
[1]School of Computer Science and Engineering, Kyungpook National University, Republic of Korea.
[2] Department of Computer Science, Information Technology University (ITU), Lahore, Pakistan.
[3] Laboratoire MIA, LaRochelle, Universite deLaRochelle, France.
[1]{maryam,skjung}@knu.ac.kr, [2]arif.mahmood@itu.edu.pk, [3]thierry.bouwmans@univ-lr.fr

Abstract

Complete moving object detection plays a vital role in many applications of computer vision. For instance, depth estimation, scene understanding, object interaction, semantic segmentation, accident detection and avoidance in case of moving vehicles on a highway. However, it becomes challenging in the presence of dynamic backgrounds, camouflage, bootstrapping, varying illumination conditions, and noise. Over the past decade, robust subspace learning based methods addressed the moving objects detection problem with excellent performance. However, the moving objects detected by these methods are incomplete, unable to generate the occluded parts. Indeed, complete or occlusion-free moving object detection is still challenging for these methods. In the current work, we address this challenge by proposing a conditional Generative Adversarial Network (cGAN) conditioned on non-occluded moving object pixels during training. It therefore learns the subspace spanned by the moving objects covering all the dynamic variations and semantic information. While testing, our proposed Complete cGAN (CcGAN) is able to generate complete occlusion free moving objects in challenging conditions. The experimental evaluations of our proposed method are performed on SABS benchmark dataset and compared with 14 state-of-the-art methods, including both robust subspace and deep learning based methods. Our experiments demonstrate the superiority of our proposed model over both types of existing methods.

1. Introduction

Complete moving object detection is a fundamental step in many computer vision based applications such as human activity analysis [38], smart cities traffic monitoring [2, 19], surveillance, and security [41, 34]. However, the aspect which makes moving object detection diverse is the presence of many challenging conditions like occlusion, bootstrapping, camouflage, illumination variations, and dynamic background. To address these problems, state-of-the-art algorithms based on Robust Principal Components Analysis (RPCA) have been applied which decompose a data matrix into a low-rank component of background and sparse component of moving objects. These algorithms have shown good performance over the past decade [16, 14, 3]. RPCA methods decompose an input matrix into a low-rank component, which corresponds to the complete background, from partial observations. While the sparse component constitutes irregular behavior of moving objects. Many RPCA based algorithms have been successfully employed for moving object detection [4, 9, 17, 22, 42]. Though, problems associated with the sparse component need to be handled such as each element of moving object is considered independently which results in incomplete moving object regions. Also, in the case of dynamic backgrounds, the sparse component tends to be scattered. RPCA based methods cannot effectively handle cluttered background or very small sized moving objects. Moreover, for camouflaged moving objects, existing methods show degraded perfor-

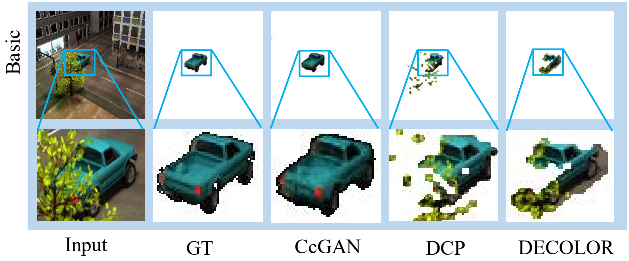

Figure 1. Visual illustration of CcGAN for complete moving object detection on SABS benchmark dataset [5] in comparison with ground truth information and two state-of-the-art methods a deep learning based : DCP [32] and a robust subspace learning based: DECOLOR [42]. It can be seen that only our proposed CcGAN has successfully detected moving object even if its occluded by background information however, other two compared methods have shown performance degradation.

mance. These limitations of RPCA based methods can be solved by exploiting deep neural networks because robust deep autoencoders are capable of achieving occlusion free complete moving object detection [36, 25, 24].

In order to estimate the occlusion free complete moving object we need a discriminative model which should have an efficient encoding properties along with a generative model to reconstruct the occluded missing information in the moving objects. Therefore, we present a solution based on conditional Generative Adversarial Network (cGAN) [11] with an architecture of robust deep autoencoder trained with occlusion free moving object detection scenario. Our proposed model *Complete cGAN (CcGAN)* is trained in the presence of different challenging environments. The occlusion free training of CcGAN aims to learn the spanned subspace by covering all the semantic information of the scene with dynamic background information and complete moving object detection which has some similarities with RPCA mechanism [7].

It can be seen in Figure 1 that there is a moving object in the scene, a car on the road, as shown by the groundtruth information. Everything else in the scene belong to the background information including dynamic movements of the tree as specified by SABS benchmark dataset [5]. It poses challenge to many state-of-the-art robust subspace learning methods. The reason behind this fact is that, these methods [42, 6] work on the assumptions of RPCA to handle outliers in the input data matrix with a constraint that background should be static and foreground/moving objects should be dynamic. In many practical cases this assumptions gets violated resulting in degraded performance. Figure 1 shows that the moving object is occluded by the tree belonging to background information. In contrast to the existing methods, our proposed CcGAN is able to reconstruct complete moving object.

2. Related Work

Over the past decade many research studies have been conducted to address the problem of moving object detection with different challenging conditions in complex scenes [8, 26, 4, 31, 28, 30]. The most classical and state-of-the-art algorithms for moving object detection are based on robust subspace learning [6, 23, 27, 22, 21]. For instance, recently, Javed *et al.* [15] proposed a method called MSCL based on the RPCA methodology by adding constraints to handle Spatio-temporal information of dynamic as well as static background and moving objects information. Although RPCA based methods show good performance, the limitation of offline data processing and high computational cost makes them unsuitable for real-time applications. Wipf *et al.* [7] presented an interesting study relating classical robust subspace learning approaches with deep learning methods. It demonstrates that Variational autoencoders can be understood as the natural evolution of RPCA, which has the capability of learning nonlinear manifolds of unknown dimension cloak by entire data corruptions.

However, recently, supervised as well as unsupervised robust deep learning based methods [1, 37, 10], have shown significantly high-performance, including GANs in many applications. For instance, Wang *et al.* [35] presented novel defensive mechanism in GAN framework. It works by modeling the adversarial noise via a generative network, trained jointly with the discriminative network for classification as a minimax game. This technique presents the robustness of neural networks in GAN against black-box attacks. Although GAN is originally unsupervised learning based algorithm, nonetheless their supervised version, conditional GANs have shown momentous performance over the past few years. Kaneko *et al.* has recently proposed A robust cGAN in which a noise transition model is incorporated so that it can learn a noise-free/clean labeled conditional generative distribution even if training labels are noisy [18].

Most of the existing moving object detection algorithms try to label input image pixels as moving objects or background. In the current manuscript we try to estimate complete moving objects despite occlusions. For this purpose we propose a conditional GAN as described in the following section.

3. Proposed Methodology

Our proposed method is based on conditional Generative Adversarial Networks (cGAN) for complete moving object detection, as shown in figure 2. The details of the training and testing of Complete cGAN (CcGAN) is discussed in the following sections.

3.1. CcGAN Model

In a classical GAN, there are two neural networks, a discriminator D and a generator G, competing with each other. The purpose of the generator network is to map a random noise sample z to an output y by generating a $2d$ image sample $G(z)$, i.e., $G(z) : z \mapsto y$. Whereas the discriminator network maps the input given to it into a single value $D(\cdot)$ which is considered as a probability of whether it belongs to training data or generated data by generator network. In conditional GAN, the generator learns mapping from random noise sample z and observed image sample x to an output image y i.e., $G(z) : \{z, x\} \mapsto y$. Similar to GAN, cGAN is also trained by using a two-player minimax game:

$$\begin{aligned}\min_G \max_D \ \Phi_1(G, D) = \mathbb{E}_{x,y\sim p_{data}(x,y)}[log(D(x, y))] \\ +\mathbb{E}_{x\sim p_{data}(x),z\sim p_z(z)}[log(1 - D(x, G(x, z)))].\end{aligned} \tag{1}$$

During training, the discriminator network aims to minimize the probability of classifying generated data as real data. Nonetheless, the objective of the generator network is to fool the discriminator by generating as similar as possible to training samples. Previously proposed approaches [13, 32] have suggested that cGAN training can be improved by updating it with a more traditional loss such as L_1 or L_2 loss. Since we aim to improve the output of generator, therefore, using L_1 distance rather than L_2 along with the adversarial loss encourages less blurring:

$$\Phi_2(G) = \mathbb{E}_{x,y\sim p_{data}(x,y),z\sim p_z(z)}[||y - G(x,z)||_1]. \quad (2)$$

In this way discriminator network purpose does not change but generator network has to generate samples similar to training data by minimizing not only its adversarial loss but also the L_1 loss between its generated samples and actual training data. The final objective function is formulated as:

$$F = arg\min_{G}\max_{D}\ \Phi_1(G,D) + \delta\Phi_2(G), \quad (3)$$

where δ is weight term for L_1 loss working with conventional adversarial loss of cGAN.

3.2. CcGAN Discriminator Network Architecture

The discriminator network of our proposed CcGAN takes input of size $256 \times 256 \times 3$ pixels and has four convolutional downsampling layers in which the 1st layer generates 64 feature maps, and the 4th layer generates 512 feature maps with size 30×30. Randomly initialized weights of 3×3 spatial filters are applied in all convolutions by a stride of 2. Furthermore, the convolutions are followed by leakyReLU activation function. The last layer of CcGAN is a fully connected layer which transforms the features map to a $1d$ vector followed by a Sigmoid function.

3.3. CcGAN Generator Network Architecture

The generator network of our proposed CcGAN is an Unet encoder-decoder network with skip connections [29] comprising of four downsampling layers each followed by a convolutional operation. The decoder has four upsampling layers followed by a deconvolutional operation to reconstruct an output image sample with the same size as the input image. Similar to [13], the generator network of our proposed CcGAN has overall eight convolutional layers in which the last layer of encoder generate 512 feature maps, and the last layer of decoder generates 64 feature maps. The decoder architecture is structured similar to the encoder except deconvolutional operation in reverse order for upsampling. The weights are randomly initialized in all the layers with ReLU activation functions except the last deconvolution layer that generates our occlusion free moving objects by using `Tanh` activation function.

4. Implementation of CcGAN

Our proposed method CcGAN is a conditional GAN composed of two different neural networks with architectures explained in sections 3.2 and 3.3. The training data of CcGAN is fixed at dimension $256\times256\times3$ with input image containing the whole scene conditioned on its occlusion free corresponding moving objects segmentation. The model is optimized via Adam [20] with a momentum $\beta = 0.5$ and a learning rate of 0.0002 for 200 epochs. The empirical value of δ mentioned in eq (3) is set to be 100. The random flipping of training images also does data augmentation during training. For testing occluded data is given to the trained model so that it can detect occlusion free complete moving objects.

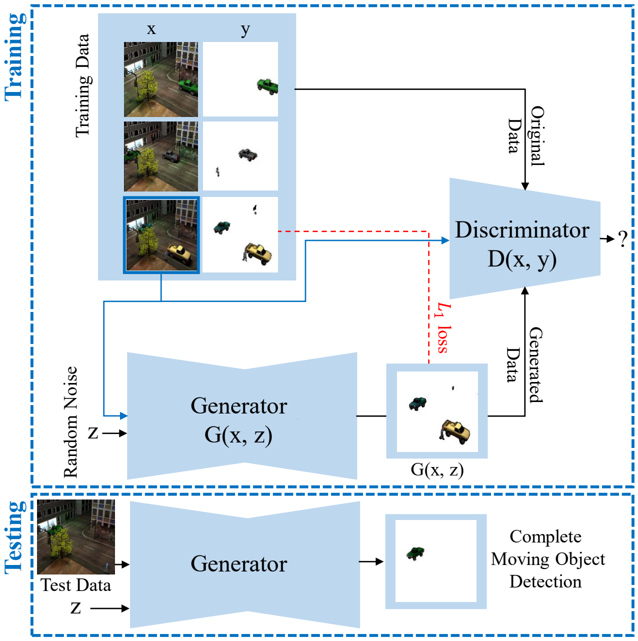

Figure 2. Visual representation of training and testing of or proposed CcGAN.

5. Experiments

We have evaluated our proposed CcGAN on a synthetically created dataset SABS[1] [5] with seven video sequences containing occluded moving objects in different challenging conditions: 'Basic', 'Noisy Night', 'Light Switch', 'Darkening', 'Camouflage', 'No Camouflage' and 'Bootstrap'. For the training of the CcGAN model, we have used 70% of SABS dataset from each category and 30% for testing in the evaluation of the scene-specific model. We have compared our method with 14 state-of-the-art-methods in which 9 methods are robust subspace learning including MSCL [15], DECOLOR [42], LSD [21], GRASTA [12], TVRPCA [6], 3TD [22], BMTDL [31], MoG-RPCA [39] and BRTF [40], 4 methods are deep learning including ForeGAN [33],

[1] http://www.vis.uni-stuttgart.de/hoeferbn/bse/

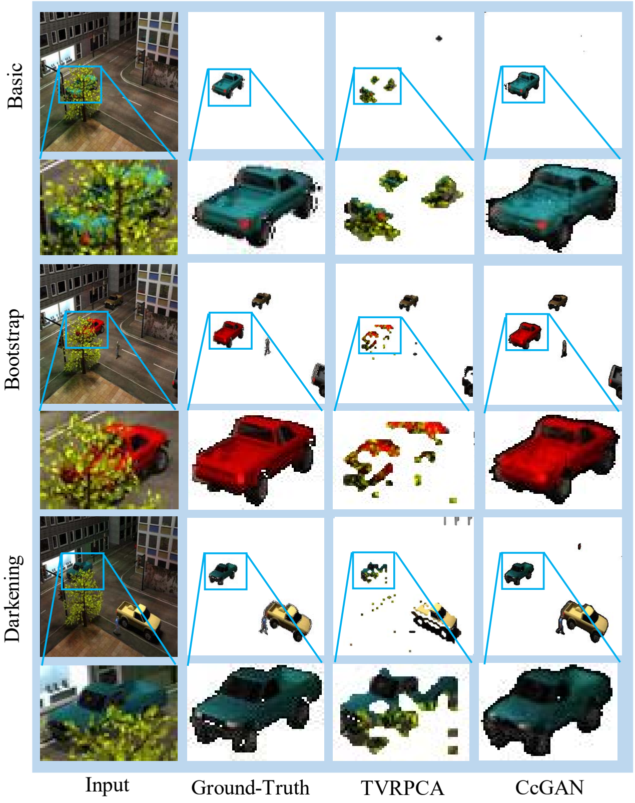

Figure 3. Visual illustration of CcGAN for complete moving object detection.

DCP [32], DeepBS [1] and CNN [37] and one method is color model, LRGB [28]. Note that, since we only used 30% of SABS dataset per category for testing, in order to do a fair comparison with other state-of-the-art methods, we have used the same 30% dataset to evaluate all methods. The evaluation is done by using F_1 measure:

$$P_r = \frac{T_P}{T_P + F_P}, \quad R_e = \frac{T_P}{T_P + F_N}, \tag{4}$$

$$F_1 = \frac{2(P_r \cdot R_e)}{P_r + R_e}. \tag{5}$$

The quantitative measure we used to estimate the occlusion free results is Jaccard similarity coefficient also known as Intersection over Union (IOU), calculated by:

$$IOU = \frac{T_P}{T_P + F_P + F_N}, \tag{6}$$

where F_N is False Negatives, F_P is False Positives, T_P is True Positives, P_r is precision and R_e is Recall. Another metric we used to estimate complete moving object detection is ‘Mean Sum of Square Error or MSSE‘ which is calculated by following:

$$MSSE = \frac{1}{mnf}\sum_{k=1}^{f}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[G_T(i,j,k) - O(i,j,k)]^2, \tag{7}$$

where $G_T \in \mathcal{R}^{m\times n}$ is the ground truth information, $O \in \mathcal{R}^{m\times n}$ is the estimated output which is complete moving object detection and f is the total number of frames in a test sequence. For more efficient complete moving object detection empirical value of MSSE should be low as shown in table 3. The details of the results evaluation per category is discussed as follows:

Basic category in SABS dataset, is defined with an artificially created road scene where the cars are considered to be the only moving objects while the rest of the information is considered as background as shown in figure 3. F_1 measure evaluated on category basic as presented in table 1, which shows that our proposed CcGAN has achieved F_1 measure (0.92) which is a significantly high score of as compared to all state-of-the-art methods. While table 2 also shows that our proposed CcGAN has achieved IOU score (0.86), highest among all the compared methods. Visual results presented in figure 3 and 4 shows that even if some background information occludes the moving objects, our proposed method can detect complete moving objects.

Noisy Night category in SABS dataset is also defined with an artificially created road scene, as shown in figure 4 in a noisy environment with low illumination. The state-of-the-art robust subspace learning based method DECOLOR [42] has achieved F_1 measure (0.84), highest in this category as shown in table 1, and our proposed CcGAN have achieved F_1 measure (0.82), which is second-best score but with a minimal difference. However, for the case of occlusion free complete moving object detection, our proposed CcGAN has achieved best, IOU score (0.73) and MSSE score (2.20) as compared to all methods.

Light Switch category in SABS dataset is also defined with same artificially created road scene, as shown in figure 4 with dynamic conditions in a low illumination environment. Our proposed CcGAN have achieved best F_1 measure (0.80) and IOU score (0.71) while robust subspace learning based method MSCL [16] and TVRPCA [6] has achieved the second best score.

Darkening category in SABS dataset, is also defined in a similar way as artificially created road scene as shown in figure 4 with dynamic high to low illumination environment including occluded moving objects. Our proposed CcGAN have also achieved the best F_1 measure (0.85) while ForeGAN [33] has achieved the second best score. Nonetheless, our proposed method has also achieved the best IOU score and MSSE in this category too, thus achieving complete moving object detection, as shown in table 2 and 3. A visual illustration is presented in figure 3 of occlusion free

Categories	CcGAN	ForeGAN [33]	DCP [32]	MSCL [15]	LRGB [28]	DECOLOR [42]	LSD [21]	GRASTA [12]	TVRPCA [6]	3TD [22]	BMTDL [31]	MoG-RPCA [39]	BRTF [40]	CNN [37]	DeepBS [1]
Basic	0.92	0.79	0.78	0.82	0.79	0.78	0.73	0.76	0.77	0.75	0.78	0.80	0.41	0.87	0.90
Noisy Night	0.82	0.65	0.59	0.75	0.75	0.84	0.72	0.51	0.59	0.76	0.59	0.79	0.54	0.71	0.80
Light Switch	0.80	0.59	0.42	0.79	0.24	0.66	0.62	0.36	0.41	0.53	0.32	0.57	0.56	0.64	0.66
Darkening	0.85	0.80	0.65	0.82	0.77	0.51	0.70	0.67	0.61	0.59	0.67	0.68	0.60	0.63	0.70
Camouflage	0.81	0.78	0.71	0.80	0.76	0.60	0.75	0.59	0.80	0.64	0.78	0.75	0.30	0.80	0.81
No Camouflage	0.84	0.74	0.72	0.79	0.81	0.84	0.80	0.63	0.76	0.70	0.79	0.82	0.60	0.75	0.80
Bootstrap	0.84	0.77	0.71	0.84	0.74	0.59	0.70	0.66	0.71	0.65	0.67	0.70	0.38	0.65	0.59
Average	0.84	0.73	0.65	0.80	0.69	0.69	0.71	0.59	0.66	0.66	0.66	0.73	0.49	0.72	0.75

Table 1. Quantitative comparison of CcGAN on SABS dataset with twelve state-of-the-art methods using F_1-measure. The first and second best performing methods are shown in red and blue colors, respectively.

Categories	CcGAN	ForeGAN [33]	DCP [32]	DECOLOR [42]	TVRPCA [6]	3TD [22]
Basic	0.86	0.69	0.68	0.65	0.59	0.59
Noisy Night	0.73	0.62	0.54	0.42	0.49	0.48
Light Switch	0.71	0.50	0.37	0.45	0.51	0.50
Darkening	0.77	0.64	0.55	0.67	0.55	0.57
Camouflage	0.72	0.63	0.60	0.62	0.51	0.58
No Camouflage	0.76	0.68	0.60	0.61	0.58	0.55
Bootstrap	0.74	0.62	0.57	0.54	0.53	0.51
Average	0.75	0.62	0.55	0.56	0.53	0.54

Table 2. Quantitative comparison of CcGAN on SABS dataset with five state-of-the-art methods using IOU. The first and second best performing methods are shown in red and blue colors, respectively.

Categories	CcGAN	ForeGAN [33]	DCP [32]	DECOLOR [42]	TVRPCA [6]	3TD [22]
Basic	1.91	2.89	3.11	2.91	5.72	3.01
Noisy Night	2.20	2.93	4.06	3.10	8.27	3.27
Light Switch	1.24	6.33	5.10	3.50	6.71	3.44
Darkening	1.75	1.93	3.17	2.77	2.00	4.15
Camouflage	0.96	1.03	3.44	4.12	2.95	3.91
No Camouflage	0.78	0.80	4.60	5.82	1.19	8.90
Bootstrap	1.46	1.87	3.47	5.17	1.85	5.62
Average	1.47	2.54	3.85	3.91	4.09	4.61

Table 3. Quantitative comparison of CcGAN on SABS dataset with five state-of-the-art methods using MSSE. The first and second best performing methods are shown in red and blue colors, respectively.

complete moving object detection by our proposed CcGAN.

Camouflage category in SABS dataset is also same presented with the scene as shown in figure 4 with dynamic background environment, including occluded camouflage moving objects. Our proposed CcGAN have also achieved best F_1 measure (0.81), IOU score (0.72) and MSSE score (0.96) while robust subspace learning based method TVRPCA [6] have achieved a second-best score with a minimal difference.

No Camouflage category in SABS dataset is the same as previous category (figure 4) with similar dynamic background environment including occluded moving objects. Table 1 shows that our proposed CcGAN and robust subspace learning based method DECOLOR [42] have achieved equal best F_1 measure (0.84) as compared to all methods. However, for occlusion free complete moving object detection, our proposed CcGAN has achieved best IOU score (0.76) and MSSE score (0.78) as presented in table 2.

Bootstrap category in SABS dataset is also same (figure 4 and 3) with similar dynamic background including occluded moving objects but in bootstrapped scenario. Note that the bootstrap is only related to background information, so it means that it does not affect our proposed CcGAN. In this category also our proposed CcGAN and MSCL [16] have achieved equal best F_1 measure (0.84) as compared to all state-of-the-art methods. While in case of IOU our proposed CcGAN has achieved the highest score as compared to all methods.

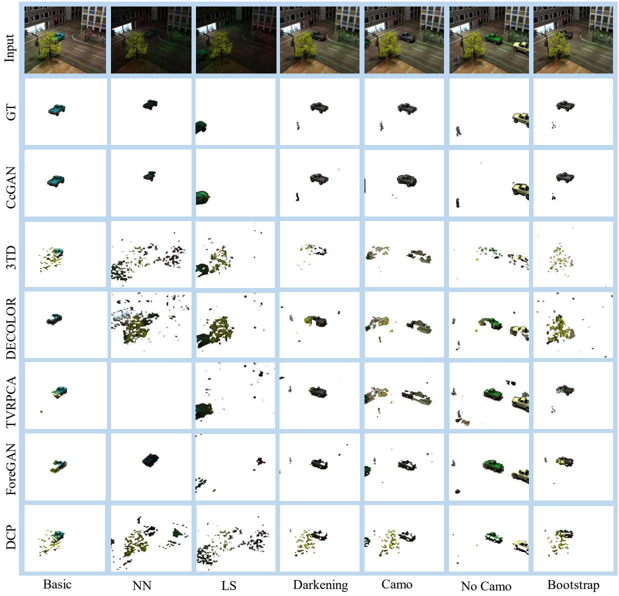

Figure 4. Visual comparison of CcGAN with five state-of-the-art methods and ground truth information on seven categories of SABS dataset, where NN is Noisy Night, LS is Light Switch, Camo is Camouflage, and No Camo is No Camouflage.

5.1. Comparative Analysis of CcGAN with Robust Subspace Learning

As mentioned in table 1 that on average, our proposed CcGAN has achieved best F_1 measure (0.84), while a subspace learning based method MSCL [16] has achieved second-best score as compared to all state-of-the-art methods. It is due to the fact that MSCL has incorporated the spatial and temporal sparse robust subspace clustering into the basic framework of RPCA. Thus, these spatio-temporal constraints embedded with the RPCA objective function enforces the background model to be spatially and temporally consistent, on linear as well as nonlinear manifolds. Therefore, it ensures that any dynamic background information should not be included in the sparse components of the input data matrix, which are the moving objects. However, due to unavailability of MSCL code, we could not compute its IOU and MSSE scores. Furthermore, table 2 shows that our proposed CcGAN has achieved best IOU score, and TVRPCA [6] has achieved a second-best score in a challenging category 'Light Switch' containing varying illumination conditions. The reason behind this fact is TVRPCA works on the assumption that the dynamic background is sparser than the moving objects with a smooth trajectory. As video sequence in TVRPCA is decomposed into low-rank static background information, a sparse and smooth moving objects information, and sparser dynamic background information. It can deal with dynamic background variations and perform moving object detection, as shown in figure 4; however, occlusion free moving object is still a challenge for TVRPCA. On the other hand, our proposed CcGAN method can not only detect moving objects successfully but also detect occlusion free complete moving objects. Subsequently, in the case of MSSE, our proposed CcGAN has also achieved the best score, and TVRPCA has again achieved the second best score in category 'Bootstrap' for the same reason mentioned previously. Moreover, for category 'Light Switch', CcGAN has again achieved the best score of MSSE and another robust subspace learning based method 3TD [22] has achieved the second best MSSE score. Since 3TD is designed to address the problem of simultaneous turbulence mitigation and moving object detection, so it decomposes the data matrix into a three-term low-rank matrix with components: the object, the background, and the turbulence. Therefore, occlusion free moving object detection is also challenging for 3TD. For instance, in category 'Bootstrap' as shown in figure 3 the moving object, the red car in the scene, is occluded by the tree, which is part of background information. The test input contains missing or occluded regions of moving objects, but CcGAN can generate the occluded regions thus achieving occlusion free complete moving object detection, on the other hand, it poses a significant challenge to robust subspace learning based method TVRPCA as shown in figure 3 and 4.

6. Conclusion

In this study, we addressed the problem of recovering complete moving objects in the presence of occlusions in complex scenes. Robust subspace learning based methods have also been used for moving object detection, though these methods cannot recover the occluded parts of these objects. For this purpose, we proposed CcGAN, which is a generative adversarial network conditioned on the complete objects during training. Therefore, the proposed GAN not only detect the moving objects but also estimates the occluded part of these objects. It is because, during training, CcGAN learns the subspace spanned by the moving objects covering all the dynamic changes and semantic information. It can, therefore, successfully generate complete moving objects at test time. The evaluation of the proposed CcGAN is done on SABS benchmark dataset for occlusion-free complete moving object detection and compared with state-of-the-art methods. The proposed CcGAN has shown excellent performance compared to the existing robust subspace learning and deep learning based methods.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2019R1A2C1010786).

References

[1] M. Babaee, D. T. Dinh, and G. Rigoll. A deep convolutional neural network for video sequence background subtraction. *Pattern Recognition*, 76:635–649, 2018.

[2] C. Baber, N. S. Morar, and F. McCabe. Ecological interface design, the proximity compatibility principle, and automation reliability in road traffic management. *IEEE Transactions on Human-Machine Systems*, 2019.

[3] T. Bouwmans, S. Javed, M. Sultana, and S. K. Jung. Deep neural network concepts for background subtraction: A systematic review and comparative evaluation. *arXiv preprint arXiv:1811.05255*, 2018.

[4] T. Bouwmans and E. H. Zahzah. Robust pca via principal component pursuit: A review for a comparative evaluation in video surveillance. *Computer Vision and Image Understanding*, 122:22–34, 2014.

[5] S. Brutzer, B. Höferlin, and G. Heidemann. Evaluation of background subtraction techniques for video surveillance. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1937–1944. IEEE, 2011.

[6] X. Cao, L. Yang, and X. Guo. Total variation regularized rpca for irregularly moving object detection under dynamic background. *IEEE transactions on cybernetics*, 46(4):1014–1027, 2016.

[7] B. Dai, Y. Wang, J. Aston, G. Hua, and D. Wipf. Connections with robust pca and the role of emergent sparsity in variational autoencoder models. *The Journal of Machine Learning Research*, 19(1):1573–1614, 2018.

[8] A. Farnoosh, B. Rezaei, and S. Ostadabbas. Deeppbm: Deep probabilistic background model estimation from video sequences. *arXiv preprint arXiv:1902.00820*, 2019.

[9] Z. Gao, L.-F. Cheong, and Y.-X. Wang. Block-sparse rpca for salient motion detection. *IEEE transactions on pattern analysis and machine intelligence*, 36(10):1975–1987, 2014.

[10] J. García-González, J. M. Ortiz-de Lazcano-Lobato, R. M. Luque-Baena, M. A. Molina-Cabello, and E. López-Rubio. Background modeling for video sequences by stacked denoising autoencoders. In *Conference of the Spanish Association for Artificial Intelligence*, pages 341–350. Springer, 2018.

[11] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[12] J. He, L. Balzano, and A. Szlam. Incremental gradient on the grassmannian for online foreground and background separation in subsampled video. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1568–1575. IEEE, 2012.

[13] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *arxiv*, 2016.

[14] S. Javed, A. Mahmood, S. Al-Maadeed, T. Bouwmans, and S. K. Jung. Moving object detection in complex scene using spatiotemporal structured-sparse rpca. *IEEE Transactions on Image Processing*, 28(2):1007–1022, 2018.

[15] S. Javed, A. Mahmood, T. Bouwmans, and S. K. Jung. Background–foreground modeling based on spatiotemporal sparse subspace clustering. *IEEE Transactions on Image Processing*, 26(12):5840–5854, 2017.

[16] S. Javed, A. Mahmood, T. Bouwmans, and S. K. Jung. Spatiotemporal low-rank modeling for complex scene background initialization. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(6):1315–1329, 2018.

[17] S. Javed, S. H. Oh, T. Bouwmans, and S.-K. Jung. Robust background subtraction to global illumination changes via multiple features-based online robust principal components analysis with markov random field. *Journal of Electronic Imaging*, 24(4):043011, 2015.

[18] T. Kaneko, Y. Ushiku, and T. Harada. Label-noise robust generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2467–2476, 2019.

[19] N. Kim and N. Bodunkov. Automated decision making in road traffic monitoring by on-board unmanned aerial vehicle system. In *Computer Vision in Control Systems-3*, pages 149–175. Springer, 2018.

[20] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[21] X. Liu, G. Zhao, J. Yao, and C. Qi. Background subtraction based on low-rank and structured sparse decomposition. *IEEE Transactions on Image Processing*, 24(8):2502–2514, 2015.

[22] O. Oreifej, X. Li, and M. Shah. Simultaneous video stabilization and moving object detection in turbulence. *IEEE transactions on pattern analysis and machine intelligence*, 35(2):450–462, 2013.

[23] P. Pan, Y. Wang, M. Zhou, Z. Sun, and G. He. Background recovery via motion-based robust principal component analysis with matrix factorization. *Journal of Electronic Imaging*, 27(2):023034, 2018.

[24] E. Plaut. From principal subspaces to principal components with linear autoencoders. *arXiv preprint arXiv:1804.10253*, 2018.

[25] J. Pu, Y. Panagakis, and M. Pantic. Learning low rank and sparse models via robust autoencoders. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3192–3196. IEEE, 2019.

[26] B. Rezaei and S. Ostadabbas. Background subtraction via fast robust matrix completion. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1871–1879, 2017.

[27] B. Rezaei and S. Ostadabbas. Moving object detection through robust matrix completion augmented with objectness. *IEEE Journal of Selected Topics in Signal Processing*, 12(6):1313–1323, 2018.

[28] J. D. Romero, M. J. Lado, and A. J. Mendez. A background modeling and foreground detection algorithm using scaling coefficients defined with a color model called lightness-red-green-blue. *IEEE Transactions on Image Processing*, 27(3):1243–1258, 2017.

[29] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[30] P.-L. St-Charles, G.-A. Bilodeau, and R. Bergevin. Subsense: A universal change detection method with local adaptive sensitivity. *IEEE Transactions on Image Processing*, 24(1):359–373, 2014.

[31] A. Stagliano, N. Noceti, A. Verri, and F. Odone. Online space-variant background modeling with sparse coding. *IEEE Transactions on Image Processing*, 24(8):2415–2428, 2015.

[32] M. Sultana, A. Mahmood, S. Javed, and S. K. Jung. Unsupervised deep context prediction for background estimation and foreground segmentation. *Machine Vision and Applications*, Nov 2018.

[33] M. Sultana, A. Mahmood, S. Javed, and S. K. Jung. Unsupervised rgbd video object segmentation using gans. In *Asian Conference on Computer Vision*, 2018.

[34] W. Sultani, C. Chen, and M. Shah. Real-world anomaly detection in surveillance videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6479–6488, 2018.

[35] H. Wang and C.-N. Yu. A direct approach to robust deep learning using adversarial networks. *arXiv preprint arXiv:1905.09591*, 2019.

[36] Y. Wang, B. Dai, G. Hua, J. Aston, and D. Wipf. Recurrent variational autoencoders for learning nonlinear generative models in the presence of outliers. *IEEE Journal of Selected Topics in Signal Processing*, 12(6):1615–1627, 2018.

[37] Y. Wang, Z. Luo, and P.-M. Jodoin. Interactive deep learning method for segmenting moving objects. *Pattern Recognition Letters*, 96:66–75, 2017.

[38] S. Yeung, O. Russakovsky, N. Jin, M. Andriluka, G. Mori, and L. Fei-Fei. Every moment counts: Dense detailed labeling of actions in complex videos. *International Journal of Computer Vision*, 126(2-4):375–389, 2018.

[39] Q. Zhao, D. Meng, Z. Xu, W. Zuo, and L. Zhang. Robust principal component analysis with complex noise. In *International conference on machine learning*, pages 55–63, 2014.

[40] Q. Zhao, G. Zhou, L. Zhang, A. Cichocki, and S.-I. Amari. Bayesian robust tensor factorization for incomplete multiway data. *IEEE transactions on neural networks and learning systems*, 27(4):736–748, 2016.

[41] J. T. Zhou, J. Du, H. Zhu, X. Peng, Y. Liu, and R. S. M. Goh. Anomalynet: An anomaly detection network for video surveillance. *IEEE Transactions on Information Forensics and Security*, 2019.

[42] X. Zhou, C. Yang, and W. Yu. Moving object detection by detecting contiguous outliers in the low-rank representation. *IEEE T-PAMI*, 35(3):597–610, 2013.