Illumination Invariant Foreground Object Segmentation using ForeGANs

Maryam Sultana, Soon Ki Jung
School of Computer Science and Engineering
Kyungpook National University
Daegu, South Korea
Email:maryam@vr.knu.ac.kr, skjung@knu.ac.kr

Abstract—The foreground segmentation algorithms suffer performance degradation in the presence of various challenges such as dynamic backgrounds, and various illumination conditions. To handle these challenges, we present a foreground segmentation method, based on generative adversarial network (GAN). We aim to segment foreground objects in the presence of two aforementioned major challenges in background scenes in real environments. To address this problem, our presented GAN model is trained on background image samples with dynamic changes, after that for testing the GAN model has to generate the same background sample as test sample with similar conditions via back-propagation technique. The generated background sample is then subtracted from the given test sample to segment foreground objects. The comparison of our proposed method with five state-of-the-art methods highlights the strength of our algorithm for foreground segmentation in the presence of challenging dynamic background scenario.

Index Terms—Background subtraction, Foreground Segmentation, Generative Adversarial Networks.

I. INTRODUCTION

The fundamental steps in many computer vision and artificial intelligence applications involves background subtraction and foreground segmentation for the tasks of moving object detection. Foreground segmentation has further applications such as visual object tracking, video surveillance and salient motion detection. Background modeling is a crucial process, which describes the scene without the presence of any foreground objects. However, foreground segmentation is a process for extracting moving objects with prior knowledge of background scene. Foreground segmentation is significantly affected by various challenges in background scene information, for instance, camera jitters, dynamic background, and sudden illumination variations. Despite that occlusion caused by foreground objects also, effects background model. Over the few decades, many techniques have been proposed in the literature to address problems of challenging background scenes for the tasks of foreground segmentation and evaluation [1], [2].

In this study, our primary focus is foreground segmentation in the presence of two major challenges in background scenes in real environments.

II. RELATED WORK

Background subtraction leads to foreground segmentation; it is a chicken-egg problem so many inclusive studies have been conducted to address this problem [3]–[5]. A very famous and well-known method for background subtraction and foreground segmentation is Gaussian Mixture Model (GMM) [6]. The basic idea of GMM is to use probability density functions based on mixture of Gaussians to model intensity variations in color at pixel level. Another very efficient and well-known technique for foreground segmentation along with background modeling is Robust Principal Component Analysis (RPCA). Until now many techniques have been proposed based on RPCA method [7]–[12] for background subtraction and foreground segmentation. However, RPCA based techniques are mostly offline methods with high computational complexity and global optimization, which is a great challenge in these techniques.

III. PROPOSED METHOD

In this section, we describe each step of our proposed algorithm in detail, which we call "ForeGAN". Our proposed method is adopted from [13]. The proposed foreground segmentation technique has two phases. Phase 1.) Training of the ForeGAN model with background video sequences containing various illumination variations. Phase 2.) Testing of ForeGAN model with video sequences containing illumination variations including foreground objects.

A. ForeGAN Model

A GAN model has two neural networks, a discriminator $D$ and a generator $G$. The objective of generator $G$ is to learn a distribution $p_{gen}$ over input data $X_t$ via mapping of $z$ samples through $G(z)$. This mapping facilitates the 1D vectors of input noise which is uniformly distributed and sampled from latent space $Z$ to the 2D image representation. In a GAN model discriminator, $D$ is a CNN model that maps a 2D image representation to a single value $D(\cdot)$. This single value $D(\cdot)$ of discriminator’s output is considered as a probability that whether the input given to the discriminator $D$ was a fake image generated $G(z)$ by the generator $G$ or a real image $X$ sampled from training data $X_t$. The discriminator and the generator are simultaneously optimized via cross entropy loss functions in a following two-player minimax game with $\Gamma(D, G)$ as a value function:

$$\min_D \max_G \Gamma_1(G, D) = \mathbb{E}_{x \sim p_{data}(x)}[\log(D(x))] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))]$$  \quad (1)
The discriminator in GAN model is a decision maker entity which is trained to maximize the probability of assigning real training sample to actual input data and samples from \(P_{gen}\) to the fake generated data. During the training process, the generator tries to improve itself by generating realistic images and the discriminator tries to identify the real and fake generated images.

B. Back-propagation of The Testing Data

During the phase:1 training, the generator learns the mapping from latent space representations \(z\) to more realistic images, \(G(z) = z \mapsto x\). However inverse mapping \(\theta(x) = x \mapsto z\) in GAN is not a straightforward process; instead we need a different mechanism for this purpose. To achieve inverse mapping, a back-propagation method is applied to input data. It is the same back-propagation method which has also been used to understand and visualize neural network’s learned features by inverting the network by updating gradients at input layer \([15]\). The loss functions to achieve back-propagation is discussed in detail in the next two sections.

1) Foreground Segmentation Loss: Given a test image \(x\) we aim to find that particular random noise \(z\) in the latent space that was mapped to generate image \(G(z)\) via back-propagation method. In order to find that specific \(z\), we have to select an initial random sample \(z_o\), from the latent space and reinforce it to the trained generator network to generate \(G(z_o)\). The loss functions are designed from the generated image \(G(z_o)\), which provides significant updating information of the coefficients of \(z_o\) which results in our back-propagation step to be shifted from \(z_o\) to \(z_1\) in the latent space distribution. The most similar generated image \(G(z_β)\) can be found by various back-propagation steps \(β = 0, 1, 2, ..., φ\) by minimizing the following loss function:

\[
Υ_F(z_β) = \sum |x - G(z_β)|. \tag{2}
\]

2) Back-propagation Image Generation Loss: Now the next issue is how to generate those images \(G(z_β)\) which makes the best match with test image \(x\). To solve this problem, we need another loss term which helps the generator to generate similar images as a test image and minimize the loss function in equation \((2)\). In order to improve the inverse mapping of the test image, \(x\), to that specific random noise \(z\), feature matching technique \([13]\) is designed to force the generator to generate the data with similar statistics as test data. This loss function is defined on discriminator in such a way that the intermediate feature layer is feed with the generated image \(G(z_β)\):

\[
Υ_M(z_β) = \sum |l(x) - l(G(z_β))|, \tag{3}
\]

where \(l(\cdot)\) represents the output of the intermediate layer of the discriminator, which describes the test image \(x\). Based on this loss function, the discriminator is now used as a feature extractor rather than a decision maker for real or fake images representations. The overall loss functions are represented as the weighted sum of both loss terms defined in equation \((4)\) and \((3)\):

\[
Υ(z_β) = (1 - η)Υ_F(z_β) + ηΥ_M(z_β). \tag{4}
\]

The back-propagation method is only applied on the coefficients of \(z\), other hyper parameters of the trained GAN model remains unchanged in Phase 2: Testing.

IV. EXPERIMENTS

We presented results on two benchmark datasets Wallflower \([16]\) and I2R \([17]\) for foreground segmentation. The testing of the proposed model is performed individually on all three datasets. All the testing samples are resized to \(64 \times 64\) and given as input to the all three models individually for validation. We set back-propagation steps to be 2000 on all three datasets which are evaluated by using following 5 metrics:

\[
A = \frac{T_p + T_n}{T_p + F_p + F_n + T_n}, \tag{5}
\]

\[
F = \frac{2(Pre \times Re)}{Pre + Re}. \tag{6}
\]

\[
Pre = \frac{T_p}{T_p + F_p}. \tag{7}
\]

\[
Re = \frac{T_n}{T_n + F_n}. \tag{8}
\]

\[
Sp = \frac{T_n}{T_n + F_p}. \tag{9}
\]

where \(T_p\) is True positives, \(T_n\) is True negatives, \(F_p\) is False positives, \(F_n\) is False negatives, \(A\) is Accuracy, \(F\) is F-Measure score, \(Pre\) is Precision, \(Re\) is Recall and \(Sp\) is Specificity. For better foreground segmentation the aim of the metrics (defined in equations \((5)-(9)\)) is to achieve maximum values in all of 5 metrics.

1) Evaluation of ForeGAN model on I2R and Wallflower dataset: Despite the self-comparison of our proposed method, we have also presented the comparison of ForeGAN model with 5 state-of-the-art methods in context to foreground segmentation. By using original implementations of the authors, we have compared our proposed method with GRASTA \([7]\), DECOLOR \([8]\), 3TD \([9]\), RAMAR \([10]\) and TVRPCA \([11]\) and evaluated the results of our proposed method on two benchmark datasets. Wallflower dataset and I2R dataset has challenges like dynamic background changes, illuminations conditions, and camouflage objects. It can be seen in Table I that our ForeGAN model on average has achieved the highest F-measure score in both datasets.

| Dataset       | GRASTA | DECOLOR | 3TD | RAMAR | TVRPCA | ForeGAN |
|---------------|--------|---------|-----|-------|--------|---------|
| IDR            | 0.5489 | 0.7401  | 0.7251 | 0.7906 | 0.6954 | 0.792   |
| Wallflower     |        |         |      |       |        |         |
| Datasets GRAST A \([7]\) DECOLOR \([8]\) 3TD \([9]\) RAMAR \([10]\) TVRPCA \([11\]) ForeGAN | | | | | | |
V. Conclusion

In this study, we present the foreground segmentation algorithm based on Generative Adversarial Network (GAN). Our goal is to segment foreground objects in the presence of two major challenges in background scenes in real environments. The two challenges are dynamic background and camouflage conditions. For this problem, we have presented a solution based on GAN working on the principle of generating background image samples with specific conditions. The comparison of our proposed method with five state-of-the-art methods highlights the strength of our algorithm for foreground segmentation in the presence of challenging illumination conditions and dynamic background scenario.

VI. Acknowledgements

This research was supported by Development project of leading technology for future vehicle of the business of Daegu metropolitan city (No. 20171105).

References

[1] T. Bouwmans, L. Maddalena, and A. Petrosino, “Scene background initialization: a taxonomy,” Pattern Recognition Letters, vol. 96, pp. 3–11, 2017.
[2] 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.
[3] M. Sultana, A. Mahmood, S. Javed, and S. K. Jung, “Unsupervised deep context prediction for background estimation and foreground segmentation,” Machine Vision and Applications, pp. 1–21, 2018.
[4] 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, vol. 28, no. 6, pp. 1315–1329, 2018.
[5] S. Javed, S. K. Jung, A. Mahmood, and T. Bouwmans, “Motion-aware graph regularized rpca for background modeling of complex scenes,” in Pattern Recognition (ICPR), 2016 23rd International Conference on. IEEE, 2016, pp. 120–125.
[6] C. Stauffer and W. E. L. Grimson, “Adaptive background mixture models for real-time tracking,” in Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on., vol. 2. IEEE, 1999, pp. 246–252.
[7] 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. IEEE, 2012, pp. 1568–1575.
[8] X. Zhou, C. Yang, and W. Yu, “Moving object detection by detecting contiguous outliers in the low-rank representation,” IEEE T-PAMI, vol. 35, no. 3, pp. 597–610, 2013.
[9] O. Oreifej, X. Li, and M. Shah, “Simultaneous video stabilization and moving object detection in turbulence,” IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 2, pp. 450–462, 2013.
[10] X. Ye, J. Yang, X. Sun, K. Li, C. Hou, and Y. Wang, “Foreground–background separation from video clips via motion-assisted matrix restoration,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 25, no. 11, pp. 1721–1734, 2015.
[11] X. Cao, L. Yang, and X. Guo, “Total variation regularized rpca for irregularly moving object detection under dynamic background,” IEEE transactions on cybernetics, vol. 46, no. 4, pp. 1014–1027, 2016.
[12] S. Javed, T. Bouwmans, M. Sultana, and S. K. Jung, “Moving object detection on rgbd videos using graph regularized spatiotemporal rpca,” in International Conference on Image Analysis and Processing. Springer, 2017, pp. 230–241.
[13] M. Sultana, A. Mahmood, S. Javed, and S. K. Jung, “Unsupervised rgbd video object segmentation using gans,” CoRR, vol. abs/1811.01526, 2018.
[14] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” arXiv preprint arXiv:1511.06434, 2015.