Single image de-raining algorithm based on semi-supervised learning

Qiu Yongru(23020221154174), Huang Tongxin(23320221154295), Peng Zhaopeng(23020221154104), Chen Han(23020221154071), Huang Guanjie(23020221154089)

Abstract

Abstract: Rain streaks particularly in heavy rain can cause severe occlusion on the background scene, which will degrade the visual quality of images captured outdoor and affect adversely the performance of many computer vision algorithms. Recently, numerous de-raining algorithms based on deep learning were proposed and achieved great success, most of which trained the model on the fully labeled synthetic rainy images due to real-world rainy images lacking paired label images. However, there is a performance bias between synthetic and real-world rainy images due to the distribution discrepancy of features between these two types of images. Thus, most algorithms generalize poorly to real-world images. To alleviate this bias, a semi-supervised learning was proposed to train the network both on the labeled synthetic and unlabeled real-world rainy images, which reduces the distribution discrepancy by minimizing the difference of the first-order and second-order statistic information. Meanwhile, in view of the complex and diverse characteristics of rain patterns, a multi-scale network was introduced to obtain richer image features and improve the performance of the model. Experimental results show that the proposed algorithm improves Peak Signal-to-Noise Ratio (PSNR) and Structural SIMilarity (SSIM) by at least 0.66dB and 0.01, respectively, in the Rain100H dataset compared with other algorithms, such as Joint Network (JDNet), Synthetic-to-Real Transfer Learning (Syn2Real) and so on. More importantly, with the reduction of distribution discrepancy, the proposed algorithm achieves a clear performance gain on the real-world rainy images.

1 Introduction

In the age of autonomous driving, how to deal with noise in the rain is becoming more and more important. The presence of raindrops in the picture will lead to the problems of low visibility, low contrast, blurring and color offset in the whole scene, and then lead to the low recognition accuracy of other objects, which greatly increases the possibility of traffic accidents due to recognition errors, resulting in serious consequences. Therefore, removing the raindrop noise from the image to obtain a clear image has a great positive impact on the whole field of autonomous driving and even computer vision. Rain removal, in fact, is a classification process that considers an image as two layers: a rainless layer and a rain layer, and then separates the rain layer from the original image, leaving a rainless map. In recent years, single image rain removal algorithms can be roughly divided into model-driven and data-driven. Before 2017, the typical image rain removal algorithm was a model-driven algorithm (non-deep learning algorithm) influenced by image decomposition, sparse coding and prior based Gaussian mixture model. Since 2017, influenced by deep convolutional network, generative adversarial network and semi-supervised or unsupervised algorithm, image rain removal algorithm has entered the period of data-driven algorithm (deep learning algorithm).

Despite the advance of the data-driven method, there is an obvious limitation. As typical deep learning, the de-raining network requires a large-scale labeled dataset for training. Due to the difficulty of the annotation for the real-world rainy images, many synthetic datasets were built, in which the rain-free background images are referred to as the pixel-wise supervised labels and the corresponding rainy images are synthesized by image procession software or code. However, the rain streaks in synthesized images differ from those in real-word images. Thus, there would be a performance drop when the network trained on the synthetic rainy images is tested on the real-world rainy images. To resolve this problem, we propose a semi-supervised learning for single image de-raining, which trains the deep network on the labeled synthetic and unlabeled real-world rainy images simultaneously.

In order to solve the problems of the above single-image rain removal algorithm, this paper proposes a new single-image rain removal algorithm based on semi-supervised learning, which inputs the labeled synthetic rain images and the unlabeled real-word rain images into the network before the rain pattern removal operation, accurately models the first-order and second-order statistical information of the two feature maps, and makes the feature distribution of the synthetic rain images and the real rain images consistent by minimizing the Euclidean distance of the mean vector and covariance matrix of the two, so as to improve the rain removal performance of the rain removal model in the real-word rain images.

2 Related Work

Deep learning based methods have shown dramatic improvements in image rain removal by using large-scale paired data. Since to obtain paired images of rain and rain-

Figure 1: Rain leads to low object recognition accuracy

free images is intractable, many of deep-learning deraining methods train the networks in a fully supervised. However, there are domain gaps between synthetic rain and real rain images, which make the deraining performance not optimum. To solve the problem, semi-supervised methods that exploit real rain images is introduced.

2.1 Supervised Learning Methods

(Yang et al. 2017a) firstly developed a multitask architecture that can detect rain locations by predicting the binary rain mask and take a recurrent framework to remove rain streaks and clear up rain accumulation progressively. (Fu et al. 2017) explored a three-layer convolutional network to predict clean image high-frequency component from its rain-contaminated counterpart. Motivated by deep residual network (ResNet), (Fu et al. 2017) further extended it to a 20-layer CNN structure (DDN) to reduce the mapping range from input to output and then to make the learning process significantly easier. Instead of relying on image decomposition framework, (Zhang and Patel 2018) proposed a conditional generative adversarial networks (GAN) for single image derainin and rs further presented a density-aware image deraining method using a multistream dense network (DID-MDN) to adaptively determine the rain-density information by integrating a residual-aware classifier process. (Li et al. 2018a) proposed a stage-by-stage recurrent squeeze-andexcitation based context aggregation network(RESCAN) to remove rain streaks in multiple stages. (Ye et al. 2021) argue that the rain generation and removal should be tightly coupled and propose to jointly learn real rain generation and removal procedure within a unified disentangled image translation framework. Extensive experiments on synthetic and real-world rain datasets show the superiority of proposed method compared to state-of-the-arts.

2.2 Semi Learning Methods

In order to improve the generality and scalability of the rain removal model, semi-supervised and unsupervised learning methods make an attempt to learn directly from real rain data. (Wei et al. 2018) adopted DDN as the backbone (supervised part) and regularized rain layer with GMM to feed unsupervised rainy images. What's more, presented a GP-based SSL framework and estimate the pseudo-GT that is used to supervise for the unlabeled samples at the latent space by jointly modeling the labeled and unlabeled latent space vectors using the GP. (Wang et al. 2018) propose a memory-oriented semi-supervised (MOSS) method which enables the network to explore and exploit the properties of rain streaks from both synthetic and real data. The key aspect of this method is designing an encoder-decoder neural network that is augmented with a self-supervised memory module, where items in the memory record the prototypical patterns of rain degradations and are updated in a self-supervised way. (Yue et al. 2021) proposes a new semi-supervised video deraining method, in which a dynamic rain generator is employed to fit the rain layer, expecting to better depict its insightful characteristics. Specifically, such dynamic generator consists of one emission model and one transition model to simultaneously encode the spatially physical structure and temporally continuous changes of rain streaks, respectively, which both are parameterized as deep neural networks (DNNs).

3 Proposed Solution

As shown in Figure 2, the overall idea of the single image rain removal algorithm based on semi-supervised learning is as follows:

1. In the feature transformation space, the input rain image is converted into feature vectors;
2. In the semi-supervised matching network, the difference between the first-order and second-order statistical information of the synthetic rain map and the real rain map feature vectors is minimized so that the two distributions are consistent;
3. In the multi-scale rain removal network, the rain pattern is removed by multiple multi-scale rain removal units to achieve to obtain a clean background feature map;
4. In the image conversion space, the clean background feature vector after rain removal is converted to an image representation.

3.1 Semi-supervised matching networ

To solve the problem of poor generalization of purely supervised learning rain removal model in real-world rain images,(Yasarla, Sindagi, and Patel 2020) proposed a semi-supervised rain removal algorithm based on rain removal,

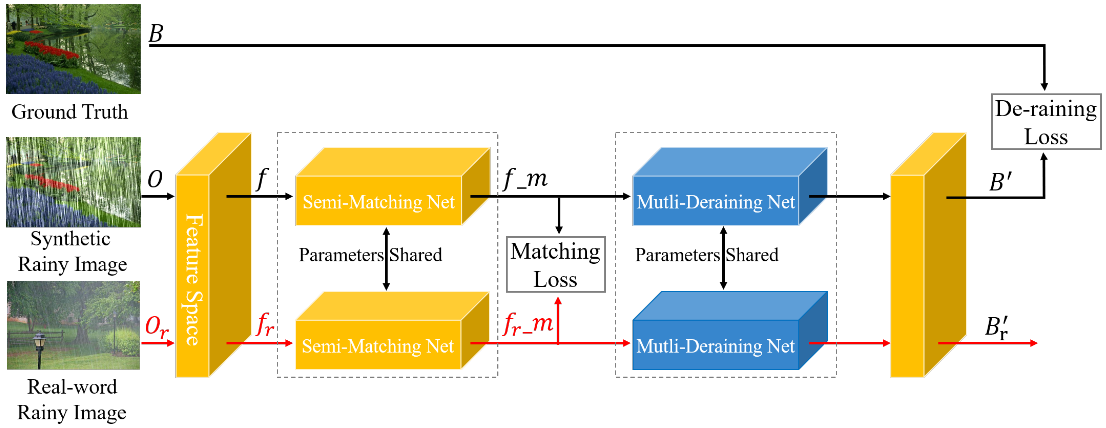


Figure 2: Multi-scale network structure based on semi-supervised learning

but in fact the rain removal algorithm did not improve the rain removal effect in real rain images. The reason may be that the algorithm is to minimize the difference between the synthetic rain map rain pattern and the real rain map rain pattern after the rain removal, which makes the model deviate from the target during the learning process. Therefore, the semi-supervised learning algorithm proposed in this paper will minimize the difference between the first-order and second-order statistical information of the synthetic rain map and the real rain map in the semi-supervised matching network before the de-rain step, so that the distribution of the synthetic rain map rain pattern features is consistent with the real rain pattern and improve the model's ability to de-rain in real-world rain images.

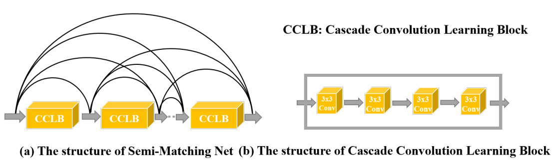


Figure 3: Detailed structure of Semi-supervised matching network

It can be observed that there are huge differences between synthetic rain patterns and real rain patterns in terms of color, brightness, and shape. To minimize the differences between synthetic and real rain images, this paper designs the semi-supervised matching network structure in Figure 3 and proposes the loss value function based on the probability distribution distance and the second-order statistical feature distance information of synthetic and real rain images from two perspectives of data distribution and feature alignment as

$$L_{match} = L_{mean} + L_{covariance}$$

where L_{mean} is the first-order information loss function and $L_{covariance}$ is the second-order information loss function. According to the theory of Maximum Mean Discrepancy (MMD), if the mean difference between two samples is equal to zero, the two samples are equally distributed.In this paper, the loss function is proposed to minimize the distance between the first-order information mean vectors of synthetic and real rain maps

$$L_{mean} = ||M_s - M_r||_F$$

where M_s is the mean vector of the synthetic rain map, M_r is the mean vector of the real rain map, and $||\cdot||_F$ is the F-parametrization of the matrix. In addition, this paper also considers the information of the second-order covariance matrix of the synthetic rain map and the real rain map, by first calculating the covariance matrix of both, and then calculating the Euclidean distance of the two covariance matrices, and finally improving the difference between the synthetic rain map and the real rain map by minimizing the value of the loss function , $L_{covariance}$ and the effect is shown in Figure 4. The $L_{covariance}$ calculation is as follows:

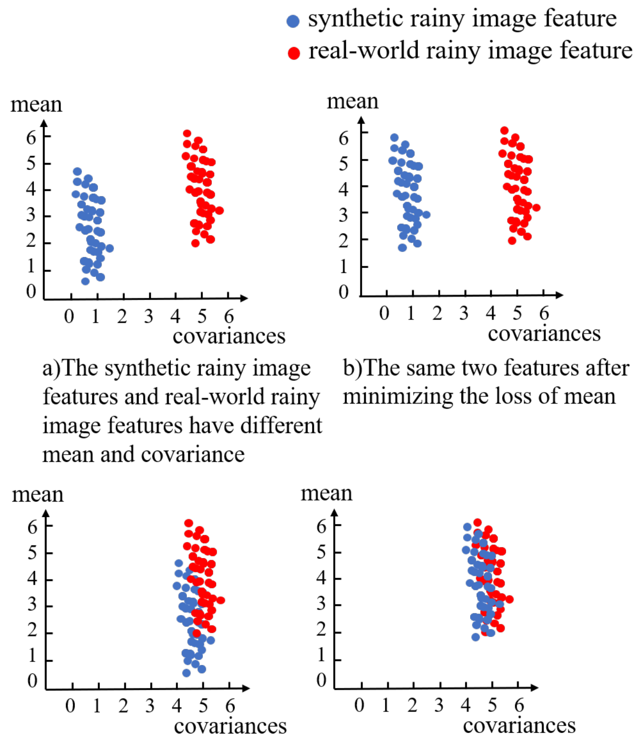
$$L_{covariance} = \frac{1}{d}||(F_s - M_s)(F_s - M_s)^T - (F_r - M_r)(F_r - M_r)^T||_F$$

where d is the dimension of the feature vector, F_s is the feature vector of the synthetic rain map, and F_r is the feature vector of the real rain map.

3.2 Multi-scale De-raining network

Complementary information can be obtained during the training process of networks of different scales, in which low-resolution networks can capture the appearance details of a given rainy day image, and high-resolution networks can retain the semantic information of a given rainy day image, so this paper introduces a multi-scale rain removal network, and the network structure is shown in Fig.5.

The multi-scale de-raining network contains multiple multi-scale de-raining units, and residual connections (He et al. 2016) and dense connections (Huang et al. 2017) are used between each multi-scale de-raining unit, in which the residual connection can more effectively transmit the feature information to the deeper network information, and the dense connection can connect any two layers in the network,



a)The synthetic rainy image features and real-world rainy image features have different mean and covariance

b)The same two features after minimizing the loss of mean

c)The same two features after minimizing the loss of covariances

d)The same two features after minimizing the loss of mean and covariances

Figure 4: Detailed structure of Semi-supervised matching network

so that each layer in the network can receive the feature input of all the previous layers, maximizing the network information reception while effectively suppressing the gradient dissipation problem that will occur in the training process. In the multi-scale rain removal unit, multi-scale features are first obtained by pooling operation downsampling with different kernels and step sizes:

$$S_i = Pooling_i(F), \quad i = 1, 2, ..., n$$

where F represents the input of each multiscale raining unit, $Pooling_i$ is the maximum pooling operation with $2^{i-1} \times 2^{i-1}$ kernel and $2^{i-1} \times 2^{i-1}$ step size maximum pooling operation, S_i is the output of the i-th scale maximum pooling operation in the multiscale structure, and n is the number of scales. Then use the convolutional cascade module on each scale to extract features for each scale and implement rain removal:

$$Z_i = CCLB_i(S_i), \quad i = 1, 2, ..., n$$

where $CCLB_i$ represents the i-th scale cascade convolutional rain removal operation, and Z_i represents the output of the i-th scale rain removal operation. Finally, through the upsampling operation and multi-scale fusion operation, the complementary features above each scale are obtained, so

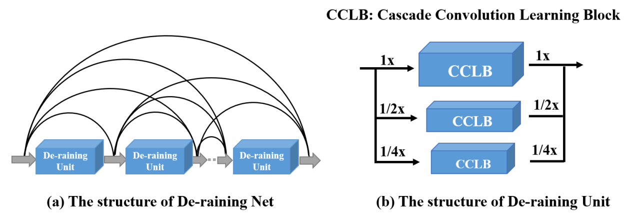


Figure 5: Detailed structure of Semi-supervised matching network

that the network can obtain the most effective features as much as possible and improve the rain removal ability of the network:

$$\hat{F} = \sum_{i=1}^{n} Z_i$$

4 Experiments

4.1 Experimental configuration and dataset

The complete network structure proposed in this paper runs under the Python framework of GeForce RTX 3090 graphics card and Ubuntu 18.04.5 LTS system. In the training process, 100 64×64 image pairs are randomly cut from the training data set as input and optimized by ADAMwe set initial learning rate 5×10^{-4} and we set matching loss function weight ω 1, batch size 16, scale number n 3, number of eigenvector channels 32.

In this work, we uses two classic composite datasets Rain100L Rain100H et al.(Yang et al. 2017b),and a real rain map dataset collected for model training.After all,in the above two classic synthetic datasets and generalized test datasets Rain12 ,Yang et al.(Yang et al. 2017b),Zhang et al.(Zhang, Sindagi, and Patel 2019),and Wang et al.(Yang et al. 2017b) proposed three real rain map datasets for model testing.

4.2 Experiment and analysis of de-raining on synthetic rainy images

In order to objectively evaluate the de-raining algorithm based on semi-supervised learning proposed in this paper, Structural SIMilarity (SSIM)(Wang et al. 2004) and Peak Signal-to-Noise Ratio (PSNR)(Zhang, Sindagi, and Patel 2019) are selected as evaluation indicators. It is tested on two classical synthetic datasets Rain100H, Rain100L and generalization test dataset Rain12 with state-of-the-art algorithms RESCAN (Recurrent Squeeze-and-Excitation Context Aggregation Net)(Li et al. 2018b), PreNet (Progressive Image Deraining Networks)(Ren et al. 2019), SSIR (Semi-Supervised Transfer Learning)(Wei et al. 2019) and JDNet (Joint Network)(Wang et al. 2020). Test results are shown in Table 1, the results of this algorithm has been rough display.

By analyzing table 1, it can be concluded that the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) of the algorithm in this paper are higher than other algorithms in the data set Rain100H and Rain100L. It can be seen that the algorithm in this paper has greatly improved in

Algorithms	Rain100H		Rain100L		Rain12	
	PSNR/dB	SSIM	PSNR/dB	SSIM	PSNR/dB	SSIM
RESCAN	25.92	0.84	36.12	0.97	32.35	0.89
PreNet	27.89	0.89	36.69	0.98	34.77	0.96
SSIR	22.47	0.71	32.37	0.92	24.12	0.78
JDNet	30.02	0.92	38.65	0.99	37.02	0.97
ours	30.68	0.93	39.53	0.99	36.61	0.97

Table 1: Quantitative de-raining results of average PSNR and SSIM on the synthesized datasets

(a)Image (b)RESCAN (c)JDNet (d)SIRR (e)Syn2Real (f)Ours

Figure 6: Visual comparison of different algorithms on real rainy images

removing rain streaks and restoring detail features. Although the PSNR evaluation index of the proposed algorithm in the synthetic dataset Rain12 is slightly lower than that of JDNet, the subsequent experimental results show that the semi-supervised rain removal algorithm proposed in this paper is significantly better than JDNet in real rain removal.

4.3 Experiment and analysis of de-raining on real-word rainy images

In order to verify the generalization ability of the semi-supervised rain removal algorithm proposed in this paper in the real rain map, the real rain map data sets proposed by Yang et al.(Yang et al. 2017b), Zhang et al.(Zhang, Sindagi, and Patel 2019) and Wang et al.(Wang et al. 2019) were tested and the rain removal results of RESCAN and JDNet based on supervised learning rain removal algorithms and SIRR and Syn2Real (Synthetic-to-Real Transfer Learning) based on semi-supervised learning rain removal algorithms were compared. Since the real rain image has no label image, the performance can only be evaluated by visual observation of the image results. This paper selects some test results as shown in Fig.6.

Conclusion

In this paper, an end-to-end neural network based on semi-supervised learning is established for a single image de-raining task. Before the realization of de-raining, the first-order mean vector information and second-order covariance statistical information of the synthetic rainy images and the real rainy images are used by the semi-supervised matching network to minimize the difference in the feature distribution between the synthetic rainy images and the real rainy images, and improve the generalization of the de-raining model in the real rainy images. At the same time, a multi-scale de-raining network is proposed, which connects multiple de-raining units through residual dense connection, and fully extracts the effective complementary features in multiple scales to achieve efficient de-rainingl effect. Through the training and test result analysis on the dataset, it can be obtained that the proposed algorithm not only obtains higher scores in the PSNR and SSIM evaluation indicators of the synthetic dataset, but also greatly improves the visual effect of de-raining in the real rainy images compared with the existing algorithm.

References

Fu, X.; Huang, J.; Ding, X.; Liao, Y.; and Paisley, J. W. 2017. Clearing the Skies: A Deep Network Architecture for Single-Image Rain Removal. *IEEE Transactions on Image Processing*, 26: 2944–2956.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Huang, G.; Liu, Z.; Van Der Maaten, L.; and Weinberger, K. Q. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708.

Li, X.; Wu, J.; Lin, Z.; Liu, H.; and Zha, H. 2018a. Recurrent Squeeze-and-Excitation Context Aggregation Net for Single Image Deraining. In *ECCV*.

Li, X.; Wu, J.; Lin, Z.; Liu, H.; and Zha, H. 2018b. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *Proceedings of the European conference on computer vision (ECCV)*, 254–269.

Ren, D.; Zuo, W.; Hu, Q.; Zhu, P.; and Meng, D. 2019. Progressive image deraining networks: A better and simpler baseline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3937–3946.

Wang, C.; Wu, Y.; Su, Z.; and Chen, J. 2020. Joint self-attention and scale-aggregation for self-calibrated deraining network. In *Proceedings of the 28th ACM International Conference on Multimedia*, 2517–2525.

Wang, T.; Yang, X.; Xu, K.; Chen, S.; Zhang, Q.; and Lau, R. W. 2019. Spatial attentive single-image deraining with a high quality real rain dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12270–12279.

Wang, Y.; Gao, B.; lok Woo, W.; Tian, G.; Maldague, X.; Zheng, L.; Guo, Z.; and Zhu, Y. 2018. Thermal pattern contrast diagnostic of microcracks with induction thermography for aircraft braking components. *IEEE Transactions on Industrial Informatics*, 14(12): 5563–5574.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4): 600–612.

Wei, W.; Meng, D.; Zhao, Q.; and Xu, Z. 2018. Semi-supervised CNN for Single Image Rain Removal. *ArXiv*, abs/1807.11078.

Wei, W.; Meng, D.; Zhao, Q.; Xu, Z.; and Wu, Y. 2019. Semi-supervised transfer learning for image rain removal. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 3877–3886.

Wenhan Yang, J. F. J. L. Z. G. S. Y., Robby T. Tan. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 580–587.

Yang, W.; Tan, R. T.; Feng, J.; Liu, J.; Guo, Z.; and Yan, S. 2017a. Deep Joint Rain Detection and Removal from a Single Image. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1685–1694.

Yang, W.; Tan, R. T.; Feng, J.; Liu, J.; Guo, Z.; and Yan, S. 2017b. Deep joint rain detection and removal from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1357–1366.

Yasarla, R.; Sindagi, V. A.; and Patel, V. M. 2020. Syn2real transfer learning for image deraining using gaussian processes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2726–2736.

Ye, Y.; Chang, Y.; Zhou, H.; and Yan, L. 2021. Closing the Loop: Joint Rain Generation and Removal via Disentangled Image Translation. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2053–2062.

Yue, Z.; Xie, J.; Zhao, Q.; and Meng, D. 2021. Semi-Supervised Video Deraining with Dynamical Rain Generator. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 642–652.

Zhang, H.; and Patel, V. M. 2018. Density-Aware Single Image De-raining Using a Multi-stream Dense Network. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 695–704.

Zhang, H.; Sindagi, V.; and Patel, V. M. 2019. Image deraining using a conditional generative adversarial network. *IEEE transactions on circuits and systems for video technology*, 30(11): 3943–3956.