Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Infrared and Visible Airborne Targets Image Fusion with Applications to Sense and Avoid

Zhouyu Zhang * **Youmin Zhang** ** **Yunfeng Cao** * **Meng Ding** ***

* *College of Astronautics, Nanjing University of Aeronautics and Astronautics, 29 Yudao Street, Nanjing 210016 China (e-mail: zhouyuzhang@nuaa.edu.cn, cyfac@nuaa.edu.cn).*
** *Department of Mechanical, Industrial and Aerospace Engineering, Concordia University, Montreal, Quebec H3G 1M8, Canada (e-mail: ymzhang@encs.concordia.ca).*
*** *College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, 29 Yudao Street, Nanjing 210016 China (e-mail: nuaa_dm@nuaa.edu.cn).*

Abstract: Machine vision has revealed great potential in recent years for Sense and Avoid (SAA) ability of Unmanned Aerial Vehicle (UAV). However, the target perception capability of machine vision largely depends on illumination, which restricts UAV to move safely in dark environment. Since images acquired by infrared and visible sensors are complementary in most cases, enhancing image qualities in dark environments by fusion of infrared and visible images is a promising solution. By considering the difficulties of image fusion for airborne targets, a Convolutional Sparse Representation (CSR) based infrared and visible airborne targets image fusion algorithm is proposed in this paper for enhancing SAA capability of UAV in dark environments, which contains three parts: image decomposition, image transformation and image reconstruction. A series of registered infrared and visible images containing airborne targets are selected to evaluate the algorithm proposed in this paper. Simulation results demonstrate the algorithm proposed in this paper effectively increases image qualities in dark environments. In the aspects of fusion metrics, the algorithm proposed in this paper can achieve favorable performance against other image fusion algorithms.

Keywords: Sense and Avoid (SAA), machine vision, target perception, image fusion, Convolutional Sparse Representation (CSR).

1. INTRODUCTION

The Sense and Avoid (SAA) ability of Unmanned Aerial Vehicle (UAV) has been recognized as the most essential factor for the integration of UAV into National Aerospace System (NAS) (Yu and Zhang, 2015). Generally, SAA is composed of two crucial parts: 1) Sensing part, which aims to detect all the airborne targets threatening UAV flight safety with the help of on-board sensing devices; 2) Avoiding part, which aims to eliminate the potential hazard based on the sensing result by trajectory re-planning and corresponding flight control (Fu et al., 2016).

Obviously, sensing part is the foundation of SAA. According to the working pattern of on-board sensing devices, SAA can be divided into two main categories: non-cooperative SAA and cooperative SAA. The non-cooperative SAA is operated with on-board sensing devices free from information exchange. The sensing devices for non-cooperative SAA contain machine vision, Light Detection and Ranging (LiDAR) and acoustic system. In contrast with non-cooperative SAA, the on-board sensing devices for cooperative SAA largely depends on information exchange with airborne targets. The sensing devices for cooperative SAA contain Automatic Dependent Surveillance-Broadcast (ADS-B) and Traffic Alert and Collision Avoidance System (TCAS), which have been widely installed on manned aircraft.

In recent years, the advantages of machine vision for the application of SAA has been widely recognized, and a series of algorithms and systems have been developed for vision based SAA. However, there still exist a few challenges for the sensing part of vision based SAA, and the most crucial one is the high demand for image quality. The factors that may deteriorate airborne image quality can be concluded as follows. 1) Insufficient illumination in dark environments (Wang et al., 2018); 2) Image blur caused by aircraft motion and pose variation (Kupyn et al., 2018); 3) Target occlusion caused by haze and smog (Liu et al., 2019). It is worth noting that previous research work of vision based SAA including airborne target detection, tracking and pose estimation are all carried out with the prerequisite that image quality is good enough. However, low airborne image quality will greatly influence the performance of these algorithms in real application. Among the above mentioned factors, insufficient illumination is the most typical one, and it will directly influence the target perception capability of UAV in dark environment.

As illustrated in Fig. 1, the visible image of the helicopter is not clear in dark environment, especially for the rotors. The infrared image of the helicopter can effectively capture the helicopter structure, however the helicopter texture information of infrared image is insufficient compared with visible image. Since the images obtained by infrared and visible sensors are

(a) Infrared image (b) Visible image (c) Fused image

Fig. 1. Infrared, visible and fused image of an airborne target

complementary, it is desirable to fuse these two kinds of images to enhance image quality, and the fused image as illustrated in Fig. 1(c) is capable of combining the advantages of both infrared and visible images.

Generally, algorithm designed for infrared and visible image fusion can be concluded as three steps: image transformation, image fusion, and image reconstruction. Among the three steps, the method for image transformation is the foundation of the whole algorithm (Liu et al., 2018b). For this reason, the research of image fusion algorithm during the past decade mainly focuses on developing a more concise and effective transformation method. The most widely used transformation methods for image fusion are Sparse Representation (SR), Convolutional Sparse Representation (CSR) and Convolutional Neural Network (CNN).

The application of SR to image fusion has achieved great success in past few years. However, due to the local representation nature of SR, the drawbacks of SR based fusion algorithm are also obvious, which can be concluded as two manifolds (Liu et al., 2016). 1) The context information loss; 2) The high sensitivity to registration errors. To overcome this issue, the fusion framework designed on the basis of global representation algorithms are proposed in recent years, and the most representative algorithms are CNN and CSR (Liu et al., 2018b).

CNN has revealed powerful potential for various computer vision tasks recently. As a supervised learning approach, the framework of CNN can be classified into two main categories, namely the regression CNN and classification CNN. Both the regression CNN and classification CNN have been successfully applied to image fusion (Liu et al., 2018a). However, the restriction of CNN based image fusion may come from the high demand for labeled training samples. CSR is originated from the de-convolutional networks designed for unsupervised image feature analysis (Zeiler et al., 2018). With applications to image fusion, CSR can be treated as an global image transformation approach. The advantages of CSR based image fusion over SR and CNN can be concluded as follows (Liu et al., 2016). 1) The global representation capability; 2) The unsupervised learning nature of CSR makes it free from large amount of labeled ground truth images. Therefore, CSR has revealed great potential for image fusion.

In this paper, a CSR based infrared and visible airborne targets image fusion algorithm is proposed for enhancing SAA capability of UAV in dark environments. Firstly, since infrared and visible images are good at capturing structure and texture information respectively, the source images are decomposed into structure layers and texture layers. Secondly, both the two image layers are transformed into the convolutional sparse domain by CSR. Finally, the transformed Convolutional Sparse Coefficient maps are fused via activity level assessment, and the fused image is obtained by synthesizing the reconstruction results of fused structure and detailed layers.

The structure of this paper is organized as follows. In Section 1, the motivation of this paper is explained. In Section 2, the algorithm for infrared and visible airborne targets image fusion is introduced. In Section 3, a series of experiments are carried out to verify the effectiveness of the algorithm. The conclusion of the paper is given in Section 4.

2. INFRARED AND VISIBLE AIRBORNE TARGETS IMAGE FUSION ALGORITHM

As illustrated in Fig. 2, the general framework of image fusion algorithm in this paper contains three parts: image decomposition, image transformation and image reconstruction. In this section, the algorithms for all the three parts will be introduced in detail.

2.1 Image Decomposition

Typically, as presented in Eq. (1), image I is composed of two layers: the structure layer I_S and the texture layer I_T. The structure layer usually represents the semantic information and captures salient objects inside the image, while the texture layer emphasizes on preserving details of the image. As illustrated in Fig. 2, the semantically meaningful structure layers are usually covered by texture layers. As mentioned above, since infrared image and visible image are good at preserving structure information and texture information respectively, decomposition of the two layers for image fusion is desirable.

$$I = I_S + I_T \tag{1}$$

In this paper, the relative total variation based image decomposition algorithm is adopted for image decomposition (Xu et al., 2012). The objective function for image decomposition is presented as Eq. (2), where $I_S(i)$ and $I(i)$ are pixel values of structure layer and original image at location i, respectively, p is the total pixel number of input image, μ is the parameter controlling smooth degree, and ε is small positive number to avoid denominator being zero. $V_x(i)$ and $V_y(i)$ as presented in Eq. (3) and Eq. (4) are total variations in x and y direction for pixel i, where $R(i)$ is the rectangular region centered at i, $g_{i,j}$ is weighting function designed to avoid spatial affinity, ∂_x and ∂_y are partial derivatives in x and y direction, respectively. The mathematical formulation of $g_{i,j}$ is presented as (5), where σ is the parameter controlling window size. The influence of image decomposition parameter on image fusion will be analyzed in detail in Section 4.

$$\underset{I_S}{\operatorname{argmin}} \sum_{i=1}^{p} (I_S(i) - I(i))^2 + \mu \cdot \left(\frac{V_x(i)}{V_x(i)+\varepsilon} + \frac{V_y(i)}{V_y(i)+\varepsilon} \right) \tag{2}$$

$$V_x(i) = \left| \sum_{j \in R(i)} g_{i,j} \cdot (\partial_x I_S)_j \right| \tag{3}$$

$$V_y(i) = \left| \sum_{j \in R(i)} g_{i,j} \cdot (\partial_y I_S)_j \right| \tag{4}$$

$$g_{i,j} \propto \exp\left(-\frac{(x_i - x_j)^2 + (y_i - y_j)^2}{2\sigma^2} \right) \tag{5}$$

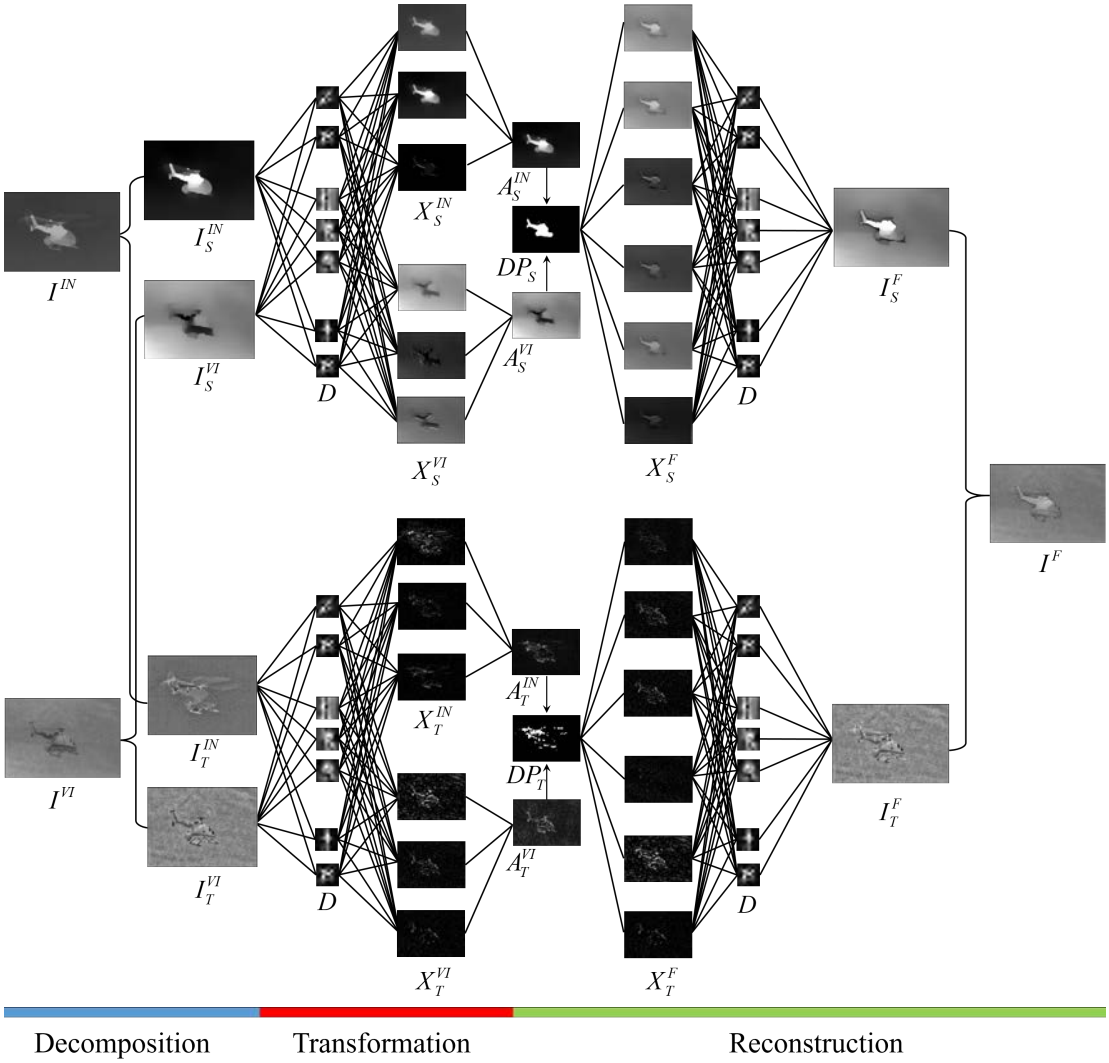


Fig. 2. General framework of infrared and visible airborne targets image fusion

2.2 CSR Based Image Transformation

As mentioned above, effective image transformation is essential for the fusion effect. In this paper, the structure layers I_S^{IN}, I_S^{VI} and texture layers I_T^{IN}, I_T^{VI} of infrared and visible images are transformed into the convolutional sparse domain by elastic net based CSR.

As presented in Eq. (6), the basic idea of CSR is that an input image I can be represented by the sum of the convolutional product of equally sized convolutional dictionary filters $D = \{d_1, d_2, ..., d_m\}$ and convolutional sparse coefficient maps $X = \{x_1, x_2, ..., x_m\}$, where m is the number of convolutional dictionary filters.

$$I = \sum_{i=1}^{m} d_i * x_i \tag{6}$$

For each input image, the convolutional dictionary is pre-learned. Therefore, the computation of convolutional sparse coefficient maps X is essential for image transformation. Conventionally, the computation of X can be operated by l_1 norm regularization, and the objective function can be expressed as Eq. (7), where λ is the regularization parameter. Since l_1 norm regularization could not guarantee group selection when applied to image transformation, the elastic net based regularization is proposed in this paper to combine the advantages of l_1 norm and l_2 norm regularization. The objective function for elastic net based regularization can be expressed as Eq. (8). The solution of Eq. (8) can be acquired by Alternating Direction Method of Multipliers (ADMM).

$$\arg\min_{x_i} \frac{1}{2} \left\| \sum_{i=1}^{m} d_i * x_i - I \right\|_2^2 + \lambda \sum_{i=1}^{m} \|x_i\|_1 \tag{7}$$

$$\arg\min_{x_i} \frac{1}{2} \left\| \sum_{i=1}^{m} d_i * x_i - I \right\|_2^2 + \lambda \sum_{i=1}^{m} \|x_i\|_1 + (1-\lambda) \sum_{i=1}^{m} \|x_i\|_2^2 \tag{8}$$

Therefore, as presented in Eq. (9), given the structure layers of I_S^{IN}, I_S^{VI} and texture layers I_T^{IN}, I_S^{VI} of infrared and visible images, the convolutional sparse coefficient maps X_S^{IN}, X_S^{VI}, X_T^{IN} and X_S^{VI} can be estimated via elastic net regularization based CSR.

$$\arg\min_{x_{S,T}^{IN,VI}{}_i} \frac{1}{2}\left\|\sum_{i=1}^{m} d_i * x_{S,T}^{IN,VI}{}_i - I_{S,T}^{IN,VI}\right\|_2^2 + \lambda\sum_{i=1}^{m}\left\|x_{S,T}^{IN,VI}{}_i\right\|_1 + (1-\lambda)\sum_{i=1}^{m}\left\|x_{S,T}^{IN,VI}{}_i\right\|_2^2 \tag{9}$$

2.3 Fused Image Reconstruction

After the computation of X_S^{IN}, X_S^{VI}, X_T^{IN} and X_S^{VI}, as presented in Eq. (10), the l_1 norm max strategy is adopted to fuse the convolutional sparse coefficient maps of structure layers and texture layers, where $X_S(i,j)$ and $X_T(i,j)$ denotes the content of X_S and X_T at location (i,j), respectively.

$$X_{S,T}^F(i,j) = \begin{cases} X_{S,T}^{IN}(i,j) & \left\|X_{S,T}^{IN}(i,j)\right\|_1 > \left\|X_{S,T}^{VI}(i,j)\right\|_1 \\ X_{S,T}^{VI}(i,j) & \left\|X_{S,T}^{IN}(i,j)\right\|_1 < \left\|X_{S,T}^{VI}(i,j)\right\|_1 \end{cases} \tag{10}$$

Since the fusion of structure layer is operated in the transformation domain, the fusion results of structure and texture layers $X_{S,T}^F = \{x_{S,T}^F{}_1, x_{S,T}^F{}_2, ..., x_{S,T}^F{}_m\}$ need to be transformed back to image domain. As presented in Eq. (11), the reconstruction of the fusion result of structure layer $X_{S,T}^F$ can be acquired by utilizing the convolutional dictionary filter.

$$I_{S,T}^F = \sum_{i=1}^{m_1} d_i * x_{S,T}^F{}_i \tag{11}$$

Finally, as presented in Eq. (12), the fusion image I^F can be obtained by adding the fusion result of structure layers and texture layers.

$$I^F = I_S^F + I_T^F \tag{12}$$

3. EXPERIMENT RESULTS AND ANALYSIS

3.1 Experiment Setup

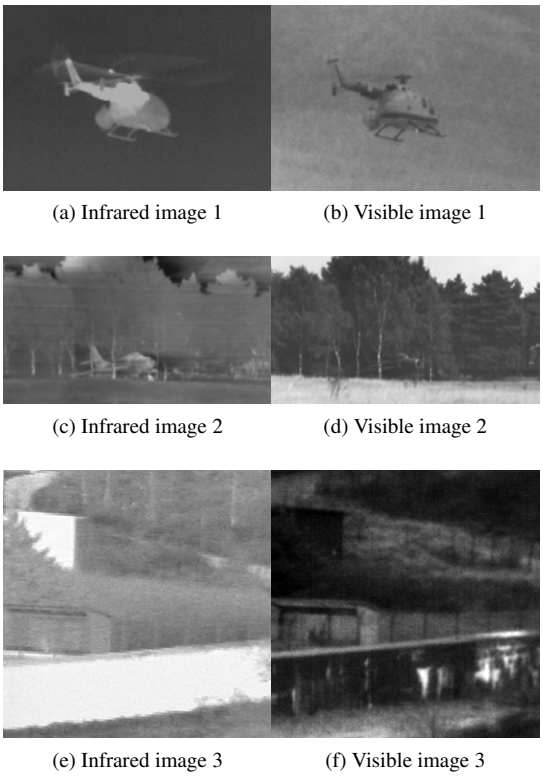
As presented in Fig. 3, three pairs of infrared and visible images containing airborne targets are selected for image fusion experiment in this paper, where the selected images obtained by different sensors are all complementary.

In order to effectively evaluate the algorithm performance, both subjective and objective metrics are adopted to measure the quality of fusion result. Subjective metrics are operated by human eyes observation, while objective metrics are operated by image quality calculation. The name and functionality for each objective metric is presented as follows (Liu et al., 2011).

1) Q_{MI}. Objective metric Q_{MI} is designed on the basis of information theory, and a high value of Q_{MI} represents good fusion result.

2) Q_M. Objective metric Q_M is designed on the basis of multi-scale scheme, and a high value of Q_M represents good fusion result.

3) Q_S. Objective metric Q_S is designed on the basis of image structural similarity, and a high value of Q_S indicates good fusion result.



(a) Infrared image 1 (b) Visible image 1

(c) Infrared image 2 (d) Visible image 2

(e) Infrared image 3 (f) Visible image 3

Fig. 3. Source images for image fusion experiment

Table 1. Parameters for algorithm evaluation

Symbol	Influence	Value range
μ	Controlling texture smooth degree	(0,0.05]
λ	Controlling weight of l_1 norm	[0.0099,0.99]

4) Q_{CB}. Objective metric Q_{CB} is designed on the basis of human perception system, and a high value of Q_{CB} indicates good fusion result.

3.2 Parameter Analysis

Two parameters are recognized to have great influence on fusion result. The names, influences and ranges for these parameters are concluded in Table 1.

Evaluation for parameter μ The range of parameter μ is presented in Table 1, and the value of λ when evaluating μ is 0.001. The fusion results with the variation of μ are presented in Fig. 4. The Q_{MI}, Q_M, Q_S, Q_{CB} for images 1, 2 and 3 are presented in Fig. 5, 6, 7 and 8, respectively. Obviously, the fusion effect decreases with the increase of μ, and the reason is that the higher smooth degree may lose more detail information of the fused image.

Evaluation for parameter λ The range of parameter λ is presented in Table 1, and the value of μ when evaluating λ is 3. The fusion results with the variation of λ are presented in Fig. 9. The Q_{MI}, Q_M, Q_S, Q_{CB} for images 1, 2 and 3 are presented in Fig. 10, 11, 12 and 13, respectively. Generally, the quality of fusion result decreases with the increase of λ. It is worth noting that the increase of λ will cause the coefficient map more sparse, which will finally indicate the decrease of fusion result.

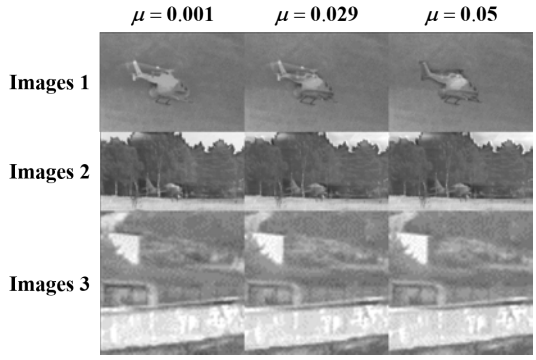
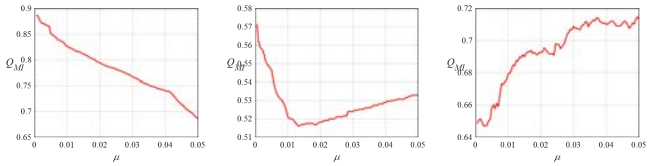
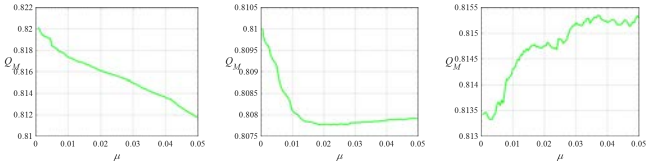


Fig. 4. Fusion results with the variation of μ for images 1, 2 and 3



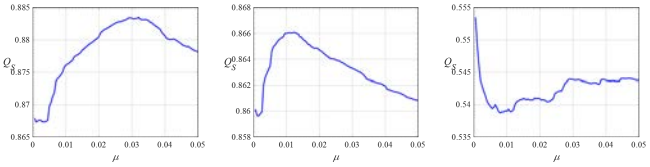
(a) Image 1 (b) Image 2 (c) Image 3

Fig. 5. Q_{MI} with the variation of μ for images 1, 2 and 3



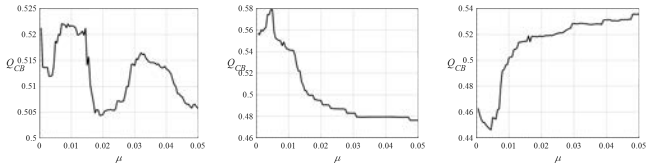
(a) Image 1 (b) Image 2 (c) Image 3

Fig. 6. Q_M with the variation of μ for images 1, 2 and 3



(a) Image 1 (b) Image 2 (c) Image 3

Fig. 7. Q_S with the variation of μ for images 1, 2 and 3



(a) Image 1 (b) Image 2 (c) Image 3

Fig. 8. Q_{CB} with the variation of μ for images 1, 2 and 3

3.3 Comparison Experiments

In order to measure the effect of the algorithm proposed in this paper more effectively, three image fusion algorithms including SR (Yang and Li, 2009), lasso based CSR (Liu et al., 2016), and CNN (Liu et al., 2018a) are selected to compare with the algorithm proposed in this paper. The comparison of fusion results are presented in Fig. 14. From the aspect of subjective

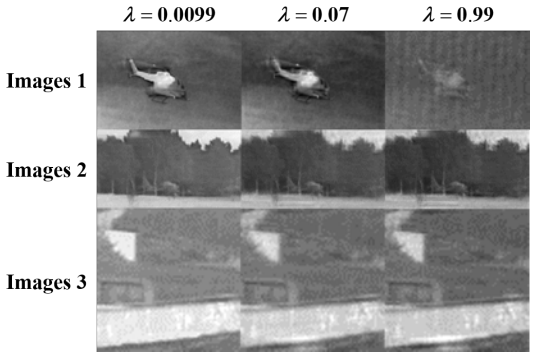
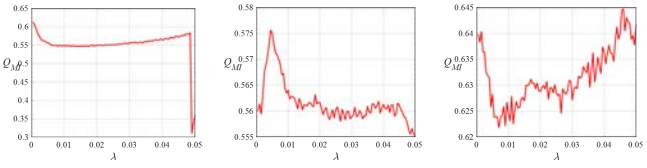
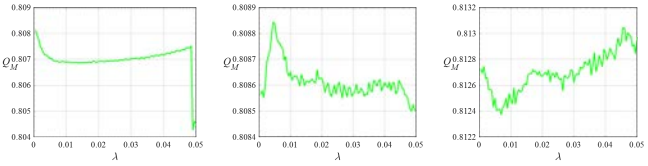


Fig. 9. Fusion results with the variation of λ for images 1, 2 and 3



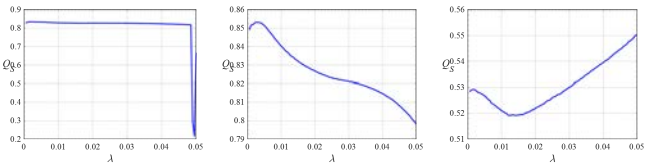
(a) Image 1 (b) Image 2 (c) Image 3

Fig. 10. Q_{MI} with the variation of λ for images 1, 2 and 3



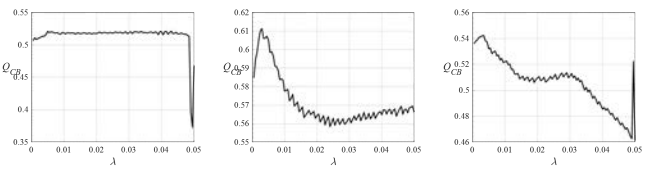
(a) Image 1 (b) Image 2 (c) Image 3

Fig. 11. Q_M with the variation of λ for images 1, 2 and 3



(a) Image 1 (b) Image 2 (c) Image 3

Fig. 12. Q_S with the variation of λ for images 1, 2 and 3



(a) Image 1 (b) Image 2 (c) Image 3

Fig. 13. Q_{CB} with the variation of λ for images 1, 2 and 3

measurement, the algorithm proposed in this paper is capable of preserving image details while strengthening the object. The comparison of objective measurements containing Q_{MI}, Q_M, Q_S, Q_{CB} are presented in Table 2, 3, 4 and 5, respectively. It is obvious that the objective measurements of the algorithm proposed in this paper outperforms other algorithms in most cases.

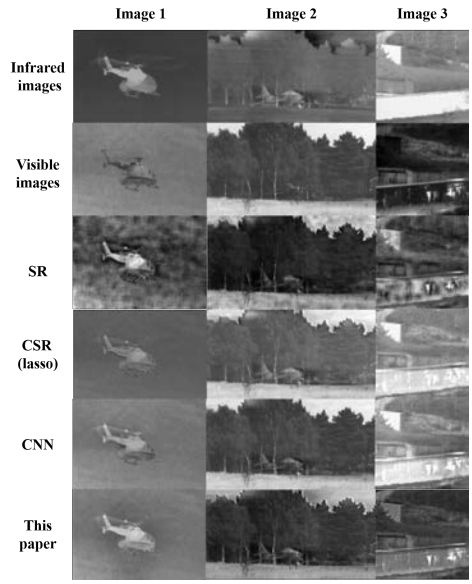


Fig. 14. Comparison of fusion results

Table 2. Comparison of Q_{MI}

Algorithm	Images 1	Images 2	Images 3
SR	7.412	6.037	6.493
CSR (lasso)	7.212	5.897	6.254
CNN	7.015	5.697	6.134
This paper	**6.915**	**5.496**	**6.013**

Table 3. Comparison of Q_M

Algorithm	Images 1	Images 2	Images 3
SR	0.493	0.587	0.629
CSR (lasso)	0.672	0.657	0.728
CNN	0.686	0.682	0.731
This paper	**0.691**	**0.684**	**0.747**

Table 4. Comparison of Q_S

Algorithm	Images 1	Images 2	Images 3
SR	2.233	2.276	2.348
CSR (lasso)	2.637	1.754	2.521
CNN	2.614	1.989	2.409
This paper	**2.713**	**2.012**	**2.706**

Table 5. Comparison of Q_{CB}

Algorithm	Images 1	Images 2	Images 3
SR	0.349	0.367	0.384
CSR (lasso)	0.417	0.208	0.408
CNN	**0.429**	0.473	0.415
This paper	0.413	**0.502**	**0.428**

4. CONCLUSION

In this paper, a CSR based infrared and visible airborne targets image fusion is presented for enhancing SAA ability of UAV. The algorithm contains three parts: image decomposition, image transformation and image reconstruction. Both subjective and objective measurements are selected to evaluate the effectiveness of the fusion algorithm proposed in this paper. Simulation results reveal that the algorithm proposed in this paper is capable of preserving image details while strengthening objects, and outperforms other fusion algorithms in most cases.

5. ACKNOWLEDGEMENT

This study is supported in part by the National Natural Science Foundation of China (No. 61673211, U1633105, 61573282, and 61833013), the Fundamental Research Funds for the Central Universities of China (No. NP2019105), Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX18_0301) and Funding for Outstanding Doctoral Dissertation in NUAA (BCXJ18-11).

REFERENCES

Fu, Y., Zhang, Y.M., and Yu, X. (2016). An advanced sense and collision avoidance strategy for unmanned aerial vehicles in landing phase. *IEEE Aerospace and Electronic Systems Magazine*, 31(9), 40–52.

Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., and Matas, J. (2018). Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8183–8192.

Liu, R., Fan, X., Hou, M., Jiang, Z., Luo, Z., and Zhang, L. (2019). Learning aggregated transmission propagation networks for haze removal and beyond. *IEEE Transactions on Neural Networks and Learning Systems*, 30(10), 2973–2986.

Liu, Y., Chen, X., Cheng, J., Peng, H., and Wang, Z. (2018a). Infrared and visible image fusion with convolutional neural networks. *International Journal of Wavelets, Multiresolution and Information Processing*, 16(03), 1850018.

Liu, Y., Chen, X., Wang, Z., Wang, Z.J., Ward, R.K., and Wang, X. (2018b). Deep learning for pixel-level image fusion: Recent advances and future prospects. *Information Fusion*, 42, 158–173.

Liu, Y., Chen, X., Ward, R.K., and Wang, Z.J. (2016). Image fusion with convolutional sparse representation. *IEEE Signal Processing Letters*, 23(12), 1882–1886.

Liu, Z., Blasch, E., Xue, Z., Zhao, J., Laganiere, R., and Wu, W. (2011). Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: a comparative study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(1), 94–109.

Wang, X., Zhou, Q., Liu, Q., and Qi, S. (2018). A method of airborne infrared and visible image matching based on hog feature. In *MIPPR 2017: Pattern Recognition and Computer Vision*, volume 10609, 106090Y. International Society for Optics and Photonics.

Xu, L., Yan, Q., Xia, Y., and Jia, J. (2012). Structure extraction from texture via relative total variation. *ACM Transactions on Graphics (TOG)*, 31(6), 139.

Yang, B. and Li, S. (2009). Multifocus image fusion and restoration with sparse representation. *IEEE Transactions on Instrumentation and Measurement*, 59(4), 884–892.

Yu, X. and Zhang, Y.M. (2015). Sense and avoid technologies with applications to unmanned aircraft systems: Review and prospects. *Progress in Aerospace Sciences*, 74, 152–166.

Zeiler, M.D., Taylor, G.W., and Fergus, R. (2018). Adaptive deconvolutional networks for mid and high level feature learning in: Computer vision. In *IEEE International Conference On*, volume 2025.