J Inf Process Syst, Vol.14, No.6, pp.1405~1419, December 2018
https://doi.org/10.3745/JIPS.04.0096

ISSN 1976-913X (Print)
ISSN 2092-805X (Electronic)

JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Infrared and Visible Image Fusion Based on NSCT and Deep Learning

Xin Feng*

Abstract

An image fusion method is proposed on the basis of depth model segmentation to overcome the shortcomings of noise interference and artifacts caused by infrared and visible image fusion. Firstly, the deep Boltzmann machine is used to perform the priori learning of infrared and visible target and background contour, and the depth segmentation model of the contour is constructed. The Split Bregman iterative algorithm is employed to gain the optimal energy segmentation of infrared and visible image contours. Then, the nonsubsampled contourlet transform (NSCT) transform is taken to decompose the source image, and the corresponding rules are used to integrate the coefficients in the light of the segmented background contour. Finally, the NSCT inverse transform is used to reconstruct the fused image. The simulation results of MATLAB indicates that the proposed algorithm can obtain the fusion result of both target and background contours effectively, with a high contrast and noise suppression in subjective evaluation as well as great merits in objective quantitative indicators.

Keywords

Boltzmann Machine, Depth Model, Image Fusion, Split Bregman Iterative Algorithm

1. Introduction

Image fusion is to use different sensors to provide complementary information to increase the information of the image and to obtain more reliable and accurate image information. It is widely applied in the fields such as geographic information system, machine vision and biomedical engineering [1]. The mechanism of visible light is different from infrared imaging. The former mainly relies on the spectral reflection of the object to image, while the latter is formed through thermal radiation of the object [2]. Therefore, the visible light usually has a rich background information and it can better describe the environmental information in the scene, while infrared light can give better target characteristics. The fusion of infrared and visible light is to use the complementary information of these two images to combine the background information of visible light with the target features of infrared light, so as to improve the target recognition ability and environmental interpretation ability of human or machine [3,4].

However, infrared and visible images have different edge characteristics. To begin with, the visible

Manuscript received July 12, 2017; first revision August 29, 2017; accepted October 23, 2017.
Corresponding Author: Xin Feng (149495263@qq.com)
* College of Mechanical Engineering and Key Laboratory of Manufacturing Equipment Mechanism Design and Control, Chongqing Technology and Business University, Chongqing, China (149495263@qq.com)

light edges are much steeper than that of the infrared light, and their edges may miss and offset corresponding to the same scene in a capacity. Secondly, there are many low-frequency components in the infrared image, and the correlation between adjacent pixels is relatively low, which factors affect the accuracy of infrared and visible fusion significantly. At present, the fusion methods of basis-based wavelet have been used in the study on fusion of infrared and visible images. Zheng et al. [5] proposed a fusion method between infrared and visible image that based on shearlet transformation of infrared light and visible light images, which has a certain improvement in subjective visual effects and objective quantitative indicators. Wang et al. [6] proposed an infrared and visible image fusion method based on non-down sampled contourlet and sparse representations, which mainly formulated different fusion rules for different sparsity coefficients to achieve better results. Gan et al. [7] proposed an image fusion method based on discrete cosine transformation coded wavelet. Kong et al. [8] proposed an improved infrared and visible light fusion method based on nonsubsampled contourlet transform (NSCT) transform and intensity-saturation-hue-saturation transform. Shen et al. [9] put forwards to fuse infrared and visible images based on a Tetrolet transform, which method recovers the sparse coefficients of various rules by the optimization algorithm to obtain the fused image and reduces the amount of fused data effectively. However, these methods do not work well on the edge of the target and background, and artifacts are likely to occur.

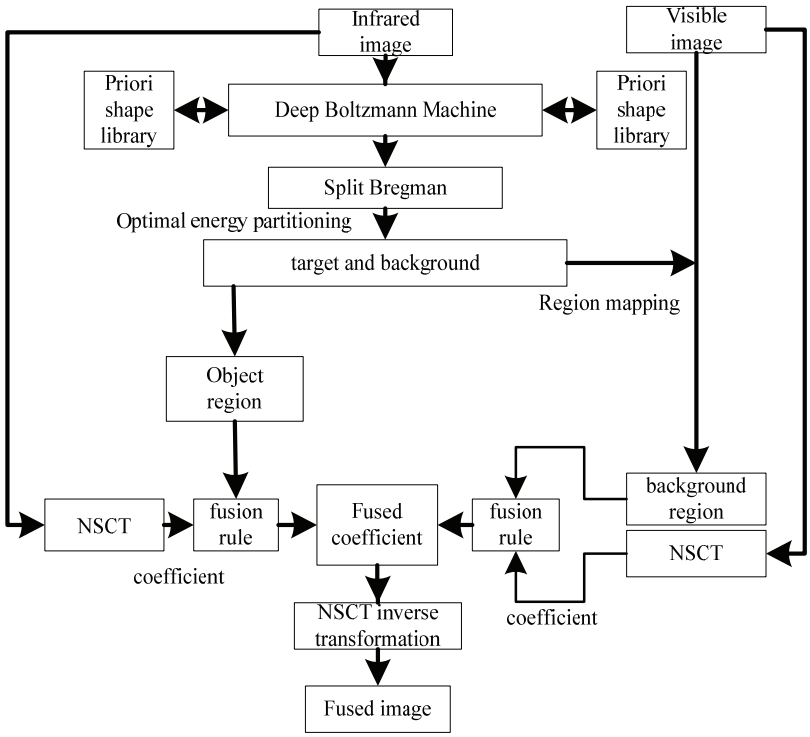


Fig. 1. Flow diagram of this paper.

In this paper, an infrared and visible light fusion algorithm is proposed based on depth model segmentation, and the flow is shown in Fig. 1. The second section establish a contour priori learning of deep Boltzmann machine (DBM) model. The third part uses a priori adaptive learning to get a clear

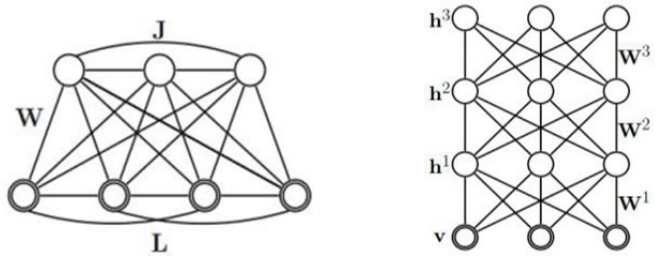
contour of the object and the background, which can effectively preserve edge sharpness of target and background while handling the problem of noise jamming. The fourth section makes a comparison between the proposed method with the NSCT method, shearlet method and DBM-DWT method, and it proves that the method has certain advantages.

2. Establishment of a Depth Model

The deep learning model has attracted wider attention from scholars at home and abroad because of its excellent ability to learn complex data [10]. The deep learning model has a multi-sensing unit, which can form abstract data of higher level by combination of the low-level features to represent the distribution structure of data (images, sound and text) in a better way. Among them, the DBM [11] is developed rapidly as one of the structures. In [12], the authors proposed a target shape template based on DBM, which method is not only capable of producing a desired target shape consistent with the actual situation but also can be used to generate shapes different from the priori library. The main idea of this section is to estimate the high-level shape or contour through the deep learning of prior shape, and to integrate it into the underlying variational segmentation model by energy effectively, thus establishing the energy equation. Finally, the energy is optimized to obtain the required contour.

2.1 Contour Priori Learning of Deep Boltzmann Machine

The DBM is an extension of restricted Boltzmann neural network, and it includes more than one hidden layers of abstraction, with full layers between layers and no connections within layers. Generally speaking, contour can also be described as a two-layer representation, namely, high-level characteristics (local targets or overall targets) and underlying characteristics (edges or corners, etc.). The underlying characteristics have invariance property, while the high-level characteristics are globally descriptive and can be well understood in the case of external disturbances such as noise or deformation. In this paper, three-layer DBM is used for a priori learning. The model and the ordinary BM are shown in Fig. 2.



(a) (b)

Fig. 2. Ordinary BM and deep BM models: (a) ordinary BM model and (b) three-layer of DBM model is usedin this paper.

If a two-dimensional binary visible unit vector v is used to represent an arbitrary shape, h^1 and h^2 are the underlying and high-level binary implicit unit vectors, then the energy of the depth Boltzmann machine with respect to the state $\{v, h^1, h^2\}$ can be defined as:

$$E^{DBN}\left(v,h^1,h^2;\theta\right)=-v^TW^1h^1-(h^1)^TW^2h^1-(a^1)^Th^1-(a^2)^Th^2-b^Tv \tag{1}$$

where $\theta=\left\{W^1,W^2,a^1,a^2,b\right\}$ is the model parameter, W^1 and W^2 are weight matrix (visible connection term) from visible unit to hidden unit and implicit unit to hidden unit, respectively. a^1 and a^2 represent the offset of the hidden unit (self-connection item), and b represents the self-connection items of the visible cell. The probability of the unit vector v is:

$$P(v,\theta)=\frac{1}{Z(\theta)}\sum_{h^1,h^2}\exp\left(-E^{DBN}\left(v,h^1,h^2;\theta\right)\right) \tag{2}$$

where the constant term is defined as [12]:

$$Z(\theta)=\sum_{v}\sum_{h^1,h^2}\exp\left\{-E^{DBM}\left(v,h^1,h^2;\theta\right)\right\} \tag{3}$$

For the given alignment of the training shape $\left\{v_1,v_2,\cdots,v_n\right\}$, DBM learning is used to determine the weight matrix and self-connection in formula (1).

2.2 The Establishment of Partition Model

In the previous section, the weight and offset terms obtained by the DBM are implicitly determined by the probability distribution in the training library, which can be expressed as:

$$P\left(v,h^1,h^2;\theta\right)\propto\exp\left(-E^{DBM}\right) \tag{4}$$

The three-layer learning can effectively capture the hierarchical structure of the prior shape and the simple local features of the underlying detection shape, and then it will feed back to the upper layer, so that the upper layer can capture global features which are more complex [13].

Once the hidden layer state is determined, a contour can be generated by conditional probability inference.

Since the contour generation state is expressed in a probabilistic form, the probability is expressed here to represent the contour, and the vector v of the two-dimensional visible unit is defined by the shape q. q is the contour probability representation $q:\Omega\to[0,1]$.

Define upper layer contour constraints:

$$E\left(q,h^1,h^2\right)=E_{DBN}\left(q,h^1,h^2,\theta\right) \tag{5}$$

where $\theta=\left\{W^1,W^2,a^1,a^2,b\right\}$ is the parameter model obtained from the previous section, and the states of the hidden layer states h^1 and h^2 can be estimated by the logic function as :

$$P\left(v_i=1\middle|h^1\right)=\sigma\left\{\sum_j W_{i,j}^1h_j^1+b_i\right\} \tag{6}$$

where $\sigma(x)$ is defined as a logical function:

$$\sigma(x)=1/(1+\exp(-x)) \tag{7}$$

Define the underlying model energy formula as:

$$E_i(q) = \int_\Omega r_o(x) q(x) dx + \int_\Omega r_b(x)(1 - q(x)) dx + \int_\Omega r_e(x) |\nabla q(x)| dx \tag{8}$$

Here, r_o and r_b are used to describe the area of the visible and infrared light image target and background, respectively, and r_e is used for edge detection. Given an initial target region, the (8) energy function is minimized to obtain the segmentation result of the target based on the underlying gray region and boundary contour information. Eqs. (5) and (8) are combined to establish the total energy equation [14]:

$$E(q, h^1, h^2; \theta) = \underbrace{\|\nabla q\|_e + \alpha_1 q^T r}_{data\ \ term} - \alpha_2 \underbrace{\left(q^T W^1 h^1 + h^{1T} W^2 h^2 + a^{1T} h^1 + a^{2T} h^2 + q^T b\right)}_{shape\ \ term} \tag{9}$$

$\|\nabla q\|_e = \int_\Omega r_e(x) |\nabla q(x)| dx$ is a weighted total variational model, $r = r_o - r_b$. The contour prior of this three-layer structure is used as an upper layer information to give an guidance to the underlying data drive, and the segmentation result is obtained by minimizing the energy function.

The energy functional (9) with respect to shape is convex functional over a convex set, and it can be efficiently solved by split Bregman method to obtain a global minimizer. Each layer of hidden units can be computed by mean-field approximate inference, just as what has been done for DBM. It will be detailed in Algorithm 1.

Given the learned model parameters $\{W^1, W^2, a^1, a^2, b\}$ and a new image u with a test shape, q is initialized as mean shape of the data-set, and h^2 as zero vector. Repeat the following steps 1 to 3 until convergence.

1. $h^1 \leftarrow \sigma\left(q^T W^1 + W^2 h^2 + a^1\right)$
2. $q \leftarrow \arg\min |\nabla q|_e + \alpha_1 q^T r - \alpha_2 \left(q^T W^1 h^1 + q^T b\right)$.
3. $h^2 \leftarrow \sigma\left(\left(h^1\right)^T W^2 + a^2\right)$

Since the energy equation is a convex function for q, the parameters q, h^1 and h^2 can be optimized using the split Bregman algorithm.

Algorithm 1. Split Bregman algorithm

1. First define $z^k = \left(c_1^k - u\right)^2 - \left(c_2^k - u\right)^2 - \alpha\left(W^1 h^1 + b\right)$,
2. Calculate $\left(q^{k+1}, \vec{d}^{k+1}\right) = \arg\min \left\|\vec{d}\right\|_e + \alpha_1 q^T z^k + \frac{\lambda}{2}\left\|\vec{d} - \nabla q - \vec{e}^k\right\|^2$
3. $\vec{d}^{k+1} = shrink_g\left(\vec{e}^k + \nabla q^{k+1}, \lambda\right)$
4. $\vec{e}^{k+1} = \vec{e}^k + \nabla q^{k+1} - \vec{d}^{k+1}$
5. Calculate $\Omega_\tau^k = \left\{x : q^{k+1}(x) > \tau\right\}$
6. Upgrade $c_1^{k+1} = \int_{\Omega_\tau^k} u dx$ and $c_2^{k+1} = \int_{\Omega/\Omega_\tau^k} u dx$

Iterate from steps 1 to 6 until the condition $\left\|q^{k+1}-q^{k}\right\|^{2}<\varepsilon$ is satisfied.

Where u is the image to be segmented, c_1 and c_2 represent the average gray value of the interior and exterior of the region to be segmented in the image u to be segmented, respectively. $\vec{d}$ is the substitute variable introduced.

3. Fusion

3.1 The Establishment of Partition Model

Define the trust level of the target area for the partitioned area:

$$C_i=\left[1+e^{-\lambda_1(\mu_f-\mu_1)}\right]^{-1}\left[1+e^{-\lambda_2(\mu_f-\mu_b-\mu_2)}\right]^{-1} \tag{10}$$

where μ_f and μ_b represent the gray level mean of the foreground area of the i^{th} segmentation target and the gray level mean of the background area respectively. λ_1 and λ_2 control the shape of the Sigmoid (*S* function) function. μ_1 and μ_2 are the displacement (offset) of the *S* function. If the gray level mean of a certain area is higher than the gray level average of the background area, the trust degree of the area belonging to the target area will be 1, otherwise 0. Then, the target area and the background area of the infrared light image will be mapped into the visible light image to realize the area segmentation of the original image.

3.2 NSCT Decomposition

The NSCT transform [15] cancels the up sampling and down sampling of the image during image decomposition and reconstruction, so that it is not only multi-scale, local and directional, but also translation-invariant and sub-band images between the images of the same size and other characteristics.

The NSCT transform is mainly composed of a nonsubsampled pyramid (NSP) and a nonsubsampled directional filter banks (NS-DFB). The NSP filter satisfies the Bezout identity $H_0(z)G_0(z)+H_1(z)G_1(z)=1$, where $H_0(z)$ and $H_1(z)$ are low-pass and high-pass decomposition filters, respectively, and $G_0(z)$ and $G_1(z)$ are corresponding synthesis filters. Similar to the NSP structure, the decomposition filter $U_0(z)$ and $U_1(z)$ in the NS-DFB and the synthesis filter $U_1(z)$ and $U_0(z)$ also satisfy the Bezout identity $U_0(z)V_0(z)+U_1(z)V_1(z)=1$.

In this paper, NSCT transform is used to decompose the original infrared image, that is, to decompose through NSP and NSDFB, so as to acquire the sub-band coefficient $C_{J,r}$ of original infrared image in high-frequency and the sub-band coefficient C_J in low-frequency.

3.3 Fusion Rule Making

For the target area, in order to maintain the target characteristics of the infrared image as much as possible, the following rules are formulated:

$$C^F(m,n)=C^I(m,n)\ ,\ (m,n)\in R_T \tag{11}$$

where R_T represents the target area, and F and I represent the fused infrared image and the source infrared image, respectively. $C(m,n)$ is the high frequency sub-band and low frequency sub-band coefficient after NSCT decomposition.

For the background area, the fusion is to extract more background details. Since the infrared and visible images have great differences in the gray values of the corresponding background parts, the structural similarity [16] is defined in the light of their corresponding background regions:

$$S_{SSIM}(I_v,I_i)=\frac{\left(2\overline{I}_v\cdot\overline{I}_i+C_1\right)}{\left(\overline{I}_v^{\,2}+\overline{I}_i^{\,2}+C_1\right)}\frac{\left(2\sigma_{v,i}+C_2\right)}{\left(\sigma_v^{\,2}+\sigma_i^{\,2}+C_2\right)} \tag{12}$$

where $\overline{I}_v$ and $\overline{I}_i$ represent the mean values of the regions corresponding to the infrared light image and the visible light image, respectively, σ_i^2 and σ_v^2 are the variances of the corresponding regions, respectively, and $\sigma_{v,i}$ is the covariance of the corresponding region. The parameters C_1 and C_2 in the structural similarity are calculated as $(k_1 l)^2$ and $(k_2 l)^2$, respectively, where l is 255 (pixel values range of the eight-bit grayscale image), and k_1 and k_2 are minimum constants ($k_1 << 1, k_2 << 1$), respectively. For the convenience of calculation, the C_1 and C_2 values take special cases, that is $C_1 = C_2 = 0$.

In this paper, the coefficient is set by experience, and the experience threshold is 0.65. The following fusion rules are defined:

$$\begin{cases} \underbrace{C_J^F(m,n)=\left(C_J^v(m,n)+C_J^i(m,n)\right)/2}_{low\quad coefficient} \\ \underbrace{\begin{aligned} &C_{J,r}^F(m,n)=\begin{cases} C_J^v(m,n) & if \quad E_{J,r}^v(m,n)\ge E_{J,r}^i(m,n) \\ C_J^i(m,n) & if \quad E_{J,r}^v(m,n)< E_{J,r}^i(m,n) \end{cases} \\ &when \quad S_{SSIM}(v,i)\ge\varepsilon \end{aligned}}_{high\quad coefficient} \\ \\ \underbrace{\begin{aligned} &C_{J,r}^F(m,n)=\begin{cases} C_J^v(m,n) & if \quad X^v(R)> X^i(R) \\ C_J^i(m,n) & if \quad X^v(R)\le X^i(R) \end{cases} \\ &when \quad S_{SSIM}(v,i)<\varepsilon \end{aligned}}_{high\quad coefficient} \end{cases} \tag{13}$$

where $C_J(m,n)$ and $C_{j,r}(m,n)$ are the low-frequency sub-band coefficients and the high-frequency sub-band coefficients in the j-scale r direction. The regional energy $E_{J,r}(m,n)$ and region Significance $X(R)$ is defined as:

$$E_{J,r}(m,n)=\sum_{x=-(M-1)/2}^{(M-1)/2}\sum_{y=-(N-1)/2}^{(N-1)/2}\left|C_{J,r}(m+x,n+y)\right|^2 \tag{14}$$

$$X(R)=\frac{1}{|R|}\sum_{\forall(m,n)\in R}\left|C_{J,r}(m,n)\right|^2\log\left|C_{J,r}(m,n)\right|^2 \tag{15}$$

where $|R|$ is the number of pixels in the region R, and $M\times N$ is the size of the local region, which is generally 3×3 or 5×5. For the timeliness of the algorithm, the value in the algorithm is 5×5.

After processing through the coefficient combination rule, the obtained coefficients are inversely transformed by NSCT to gain the final fused image.

4. Experimental Results and Analysis

4.1 Shape Iteration Results

MPEG7-CE-Shape database and 300 pieces of 300 class objectives are chosen for training. The hidden layer unit number in Boltzmann machine is set as 1600 and 800, and the number of iterations is set as 500 and 200 times. The iteration times for overall training is 1000, and simulation experience is carried out in MATLAB environment.

Result is fitted on a piece of image contour with interference, as shown in Fig. 3. It showcases several samples of the image during iteration process, and the curve in the first line is precisely in conformity with the target boundary. The second line indicates the corresponding contour probability representation that the segmentation shape that is almost the same as the contour can be obtained after 50 times of iterations.

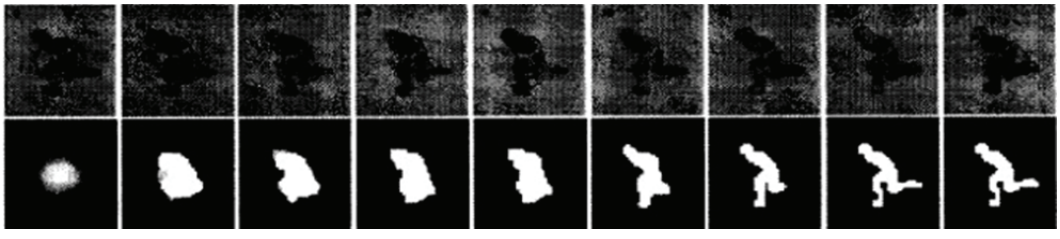
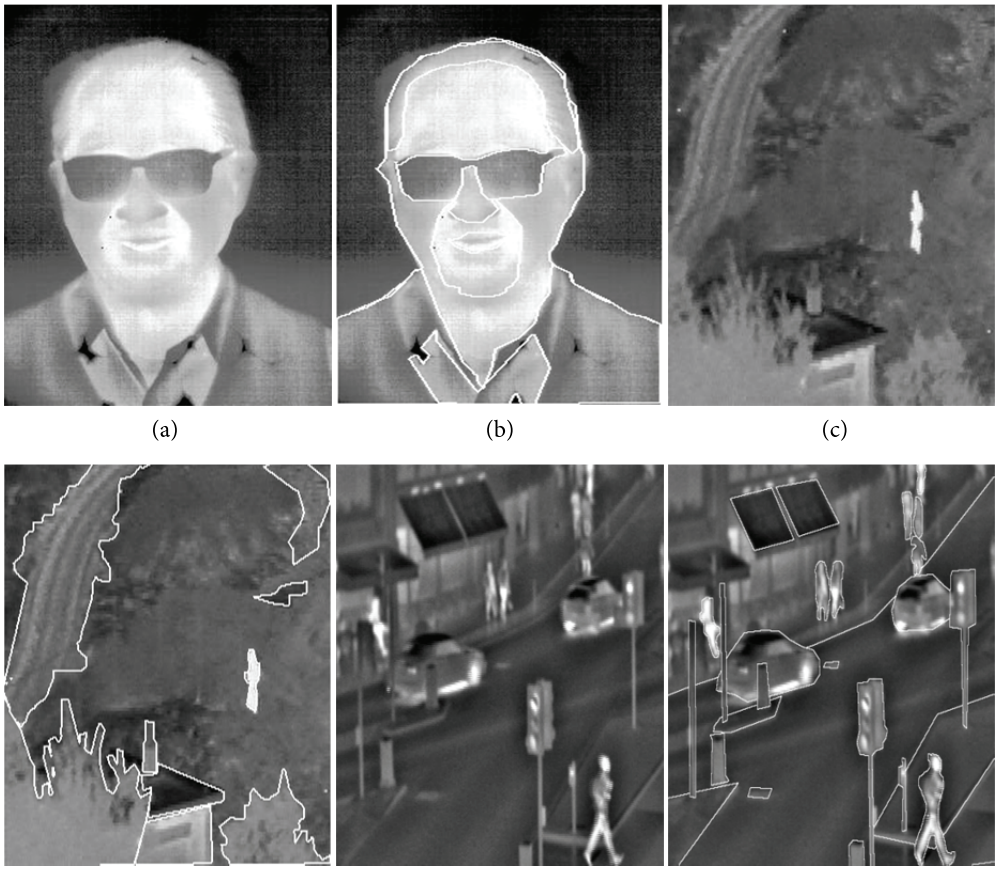


Fig. 3. Iterative process of segmentation algorithm.

4.2 Regional Segmentation Results

Fig. 4 shows the results of region segmentation based on the depth Boltzmann model. Fig. 4(a), (c) and (e) indicate the infrared images of the three infrared and visible images. Fig. 4(b), (d) and (f) are the results of the segmentation by the method. It can be seen that the algorithm of learning segmentation using the depth Boltzmann model of this paper can effectively obtain accurate target and background contours, which is beneficial to the determination and formulation of fusion rules in the target background region after NSCT transformation.



(a) (b) (c)

(d) (e) (f)

Fig. 4. Results of image region segmentation. (a) Infrared image 1, (b) regional segmentation result, (c) infrared image 2, (d) regional segmentation result, (e) infrared image 3, and (f) regional segmentation result.

4.3 Fusion Results

The performance of the fusion method has been verified by three sets of visible light and infrared light images, and the effectiveness of the proposed method has been testified by the NSCT fusion method [17], the shearlet fusion method [5], the region segmentation in this paper, and the DBM-DWT [7]. The NSCT fusion method uses the parameter setting criteria [17], the number of decomposition layers is set as 3, the scale decomposition filter uses a "maxflat" filter, and the direction decomposition filter uses a "dmaxflat" filter. The decomposition grade is set to be 2,3,3 from coarse to fine. In the shearlet fusion method, three-layer decomposition is carried out. In this paper, decomposition scales of the NSCT transform filter is 1, and other settings are the same as those in [17]. The parameters of the DWT decomposition are set in accordance with the literature [7]. The computer configurations in the test were: Windows XP operating system, Pentium Dual-Core E5400@2.70 GHz CPU, 2 GB memory. The platform used for algorithm programming is MATLAB 7.0.1.

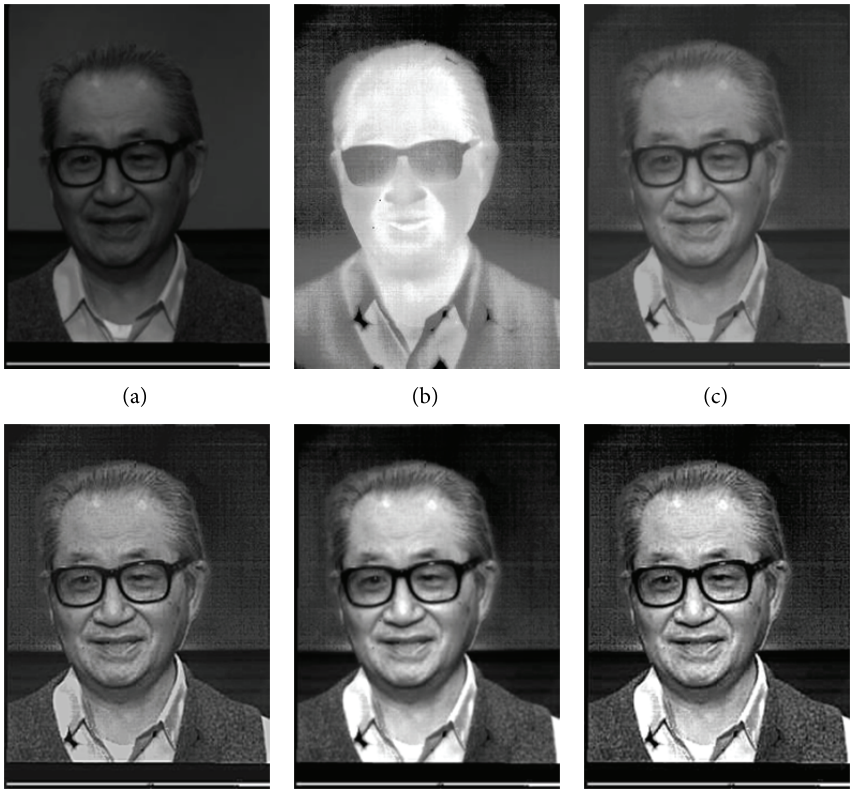


Fig. 5. Comparison of several fusion methods (Equinox faces image). (a) Visible image, (b) infrared image, (c) NSCT method, (d) shearlet method, (e) DBM-DWT method, and (f) proposed method in this paper.

The source image of the first set of visible and infrared images is the registered Equinox faces image. Among them, Fig. 5(a) is the visible light image; Fig. 5(b) is the infrared light image; Fig. 5(c) shows the results of the NSCT fusion method, which basically has no ghosting and block, with a high definition, which reflects the ability of the NSCT transform to capture edge information. Fig. 5(d) is the result of the shearlet fusion method, in which the visual effect is slightly better than the former, with clearer details. Fig. 5(e) suggests the result of the fusion of the DBM-DWT method, that is, the internal contrast of the segmentation region is relatively high, yet there are some ghosting and blocks in the segmentation edge. Fig. 5(f) is the fusion method of this paper whose contrast is the highest in comparison, and even a small amount of texture information can be captured, with quite complete figure, which fully demonstrates the merits of NSCT transform in the image fusion cannot be replaced by DWT transform due to its translation invariance and multi-directional representation.

The second group is the registered TNO UN Camp image [18], as shown in Fig. 6. Fig. 6(a) is the original infrared light image, Fig. 6(b) is the original visible light image. By comparing the results of the four fusion methods in Fig. 6(c)–(f), we can find that the contrast obtained by the fusion method is relatively high, and maintains the complete image edge contour and texture detail information.

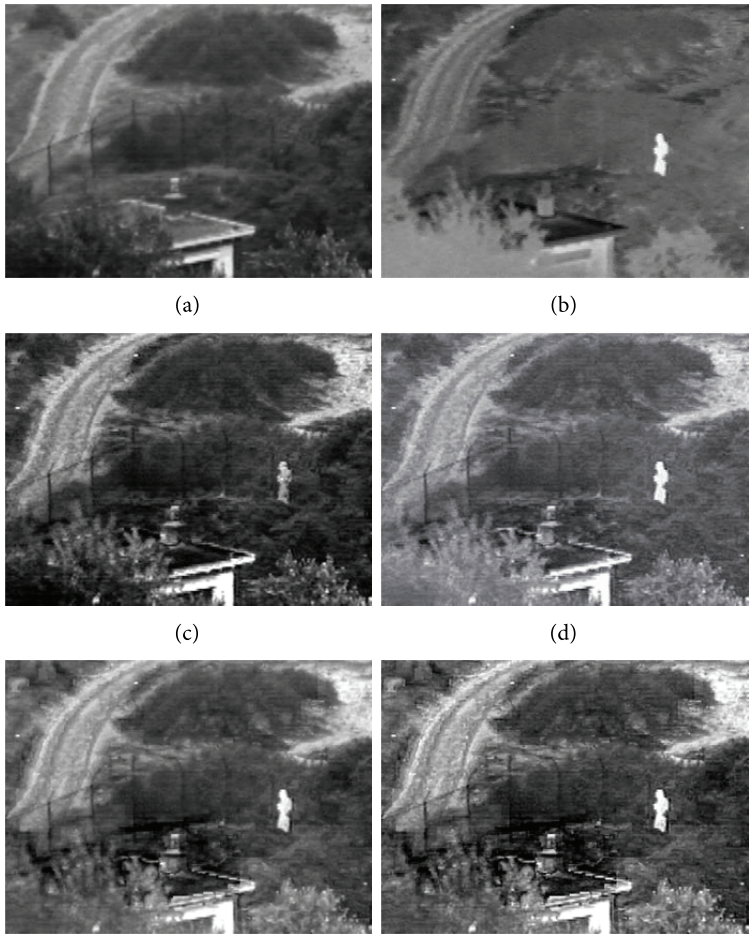


Fig. 6. Comparison of various fusion methods' effect (TNO UN Camp image). (a) Visible image, (b) infrared image, (c) NSCT method, (d) shearlet method, (e) DBM-DWT method, and (f) proposed method in this paper.

The third group is the registered Bristol Queen's Road image, as shown in Fig. 7. Fig. 7(a) showcases the original visible light image and Fig. 7(b) is the original infrared light image. Gaussian white noise with a mean value of 0 variance of 0.01 is added to the original image, and the image after noise is added as shown in Fig. 7(c) and (d), so as to further verify the superiority of the proposed algorithm in noise suppression. From Fig. 7(e)–(h), the noise fusion result of Fig. 7 and the enlarged result of Fig. 8, it can be clearly seen that the fusion method has the best effect on noise suppression after learning by the DBM method. The depth of Boltzmann-based segmentation method has the ability to learn and segment accurate target and background contours, and it can suppress noise better than pure NSCT and shearlet methods. Due to the fusion cannot be distinguished and the noise singularity of the high-frequency sub-band portion will also be affected by noise through NSCT transform and shearlet transform, the method in the paper has higher contrast and higher definition compared with the DBM-DWT method.

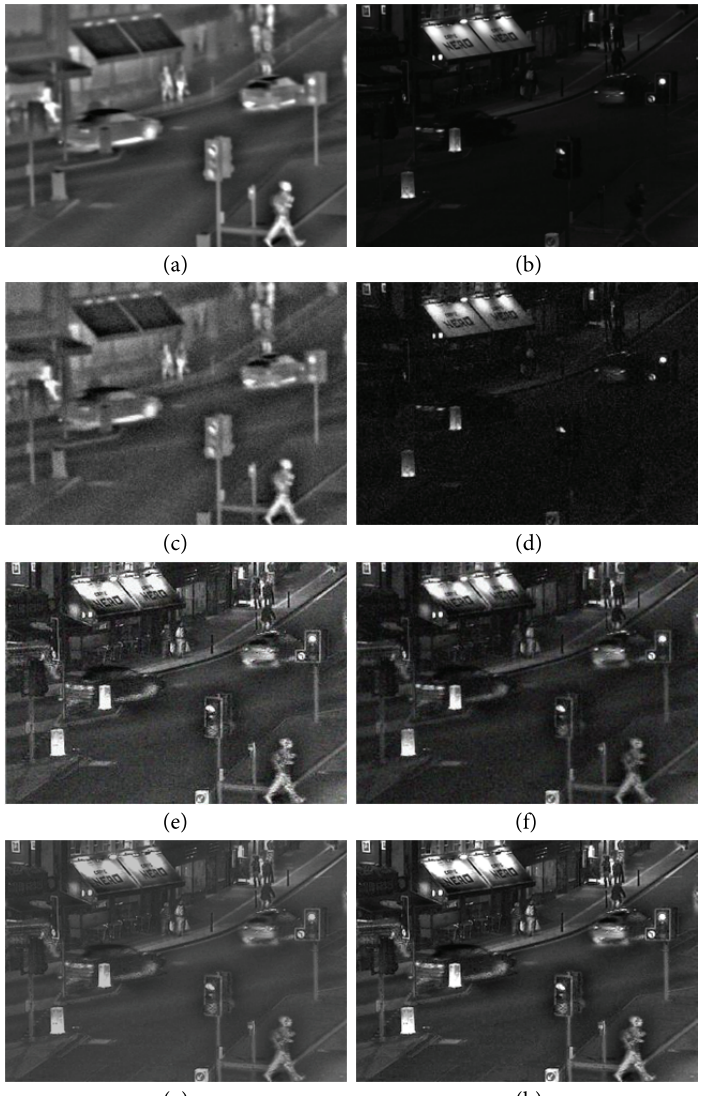
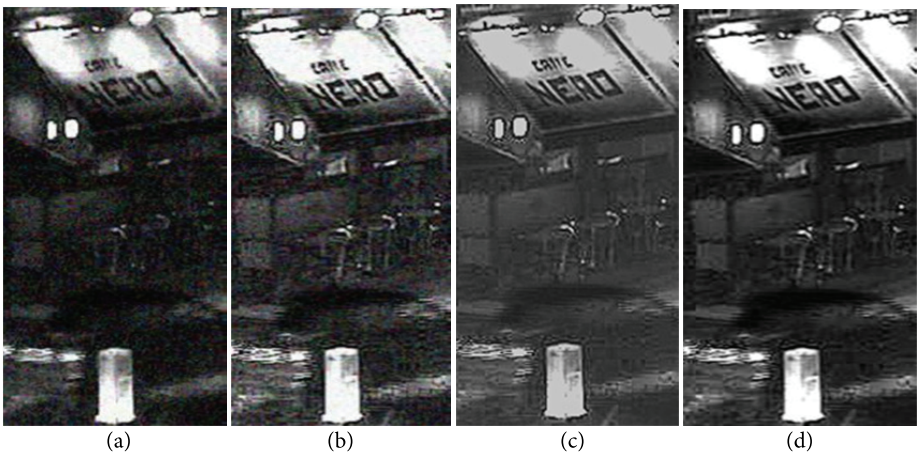


Fig. 7. Comparison of various fusion methods' effect (Bristol Queen's Road image). (a) Infrared image, (b) visible image, (c) adding noise infrared image, (d) adding noise visible image, (e) NSCT method, (f) shearlet method, (g) DBM-DWT method, and (h) proposed method in this paper.

Fig. 8 is a comparison of the image of all the fusion methods of Fig. 7 after enlargement of local area, wherein Fig. 8(a) is the fusion result of the NSCT method in [17], with clearer contour and better preserved edge characteristic. In addition, there is no Ghosting substantially, but with a lot of noise, which means that it has no suppression on noise. Fig. 8(b) shows the result of the improved shearlet transform fusion method proposed in [5]. As the result indicates, its edge is smoother, the distortion is smaller, and the contrast is higher, but there is no noise suppression. Fig. 8(c) shows the fusion result of the DBM-DWT method. It can be seen that the sharpness is higher than the first two methods due to its suppression of noise, but the contrast is relatively low, and the edge of the region is slightly distorted. Fig. 8(d) showcases the fusion result of the method. It can be clearly seen that the image overcomes the

edge oscillation while retaining the edge characteristics, and the visual effect is better than the former three methods, with the noise interference effectively suppressed.



(a) (b) (c) (d)

Fig. 8. Fusion effect comparison of local area. (a) NSCT method, (b) shearlet method, (c) DBM-DWT method, and (d) proposed method in this paper.

Table 1. Fusion evaluation index

Fusion method	IE	AG	PSNR	$Q_{AB/F}$	T /s
The first set of visible and infrared images					
NSCT	7.66	7.45	39.58	0.79	101
Shearlet	7.86	7.61	38.26	0.76	289
DBM-DWT	7.74	7.85	38.89	0.72	262
Proposed method	7.96	8.05	39.83	0.86	351
The second set of visible and infrared images					
NSCT	6.74	6.43	30.56	0.68	488
Shearlet	6.53	6.15	32.47	0.71	541
DBM-DWT	6.59	6.41	32.12	0.64	447
Proposed method	6.83	6.93	34.15	0.79	474
The third set of visible and infrared images					
NSCT	5.23	4.22	19.23	0.51	453
Shearlet	5.46	4.86	21.23	0.55	563
DBM-DWT	6.18	4.18	30.12	0.61	621
Proposed method	6.36	5.91	32.56	0.70	787

In order to evaluate the fusion performance objectively, five image evaluation indicators in [19] were used to quantify and compare, namely IE (information entropy), AG (average gradient), PSNR (peak signal to noise ratio), Q (edge retention), and T (image recovery time). The quantitative evaluation index of the fusion methods used in the experiments on infrared and visible images are shown in Table 1.

As shown in Table 1, the DBM-DWT method is slightly different from the NSCT method and the shearlet method in terms of edge retention, and other indicators are similar to the first two methods, with some indicators even better. In addition to the runtime cost of this method slightly more than that of the DBM-DWT method, the shearlet method and the NSCT method, other indicators are better. It

can be seen from the third set of data that the proposed algorithm and the DMB-DWT method have obvious suppression effects compared with the other three methods against the influence of noise, so the algorithm has certain advantages as far as comprehensive performance concerned.

5. Conclusion

In this paper, an infrared and visible light fusion algorithm is proposed based on depth model segmentation. The learning method is used to obtain the clear contour of the target and background adaptively, so as to effectively preserve the edge definition of the target and the background and effectively suppress the noise. Then, the target contour and the background contour are merged with the corresponding fusion rules separately. The method can effectively overcome the shortcomings of infrared and visible image fusion, that is, they are easy to be interfered by noise and produce artifacts, which will lead the target contour to be unclear and have low contrast. Finally, the effectiveness of the proposed algorithm is proved through experiments.

Acknowledgement

The paper is supported by National Natural Science Foundation of China (No. 31501229, 61861025), Chongqing Nature Science Foundation for Fundamental science and frontier technologies (csct2015 jcyjA40014, cstc2015jcyja50027, cstc2018jcyjAX0483), Chongqing Municipal Education Commission Foundation and Frontier Research Project (No. KJ1500635). Initial Scientific Research Fund of Young Teachers in Chongqing Technology and Business University (No. 2014-56-07).

References

[1] Z. Wang, Y. Ma, and J. Gu, "Multi-focus image fusion using PCNN," *Pattern Recognition*, vol. 43, no. 6, pp. 2003-2016, 2010.

[2] D. P. Bavirisetti and R. Dhuli, "Fusion of infrared and visible sensor images based on anisotropic diffusion and Karhunen-Loeve transform," *IEEE Sensors Journal*, vol. 16, no. 1, pp. 203-209, 2016.

[3] V. Bhateja, H. Patel, A. Krishn, A. Sahu, and A. Lay-Ekuakille, "Multimodal medical image sensor fusion framework using cascade of wavelet and contourlet transform domains," *IEEE Sensors Journal*, vol. 15, no. 12, pp. 6783-6790, 2015.

[4] Y. Yang, S. Tong, S. Huang, and P. Lin, "Multifocus image fusion based on NSCT and focused area detection," *IEEE Sensors Journal*, vol. 15, no. 5, pp. 2824-2838, 2015.

[5] H. Zheng, C. Zheng, X. Yan, and H. Chen, "Visible and infrared image fusion algorithm based on shearlet transform," *Chinese Journal of Scientific Instrument*, vol. 33, no. 7, pp. 1613-1619, 2012.

[6] J. Wang, J. Peng, and G. He, X. Feng, and K. Yan, "Fusion method for visible and infrared images based on non-subsampled contourlet transform and sparse representation," *Acta Armamentarii*, vol. 34, no. 7, pp. 815-820, 2013.

[7] T. Gan, S. T. Feng, S. P. Nie, and Z. Q. Zhu, "Image fusion algorithm based on block-DCT in wavelet domain," *Acta Physica Sinica*, vol. 60, no. 11, article no. 114205, 2011.

[8] W. Kong, L. Zhang, and Y. Lei, "Novel fusion method for visible light and infrared images based on NSST–SF–PCNN," *Infrared Physics & Technology*, vol. 65, pp. 103-112, 2014.
[9] Y. Shen, X. Feng, and Y. Hou, "Infrared and visible images fusion based on Tetrolet transform," *Spectroscopy and Spectral Analysis*, vol. 33, no. 6, pp. 1506-1511, 2013.
[10] G. Dahl, A. R. Mohamed, and G. E. Hinton, "Phone recognition with the mean-covariance restricted Boltzmann machine," *Advances in Neural Information Processing Systems*, vol. 23, pp. 469-477, 2010.
[11] R. Salakhutdinov and H. Larochelle, "An efficient learning procedure for deep Boltzmann machines," *Neural Computation*, vol. 24, pp. 1967-2006, 2012.
[12] S. M. Ali Eslami, N. Heess, and J. Winn, "The shape Boltzmann machine: a strong model of object shape," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, 2012, pp. 406-413.
[13] L. J. Latecki, R. Lakamper, and T. Eckhardt, "Shape descriptors for non-rigid shapes with a single closed contour," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Hilton Head Island, SC, 2000, pp. 424-429.
[14] W. Li, Q. Li, W. Gong, and S. Tang, "Total variation blind deconvolution employing split Bregman iteration," *Journal of Visual Communication and Image Representation*, vol. 23, no. 3, pp. 409-417, 2012.
[15] A. L. Da Cunha, J. Zhou, and M. N. Do, "The nonsubsampled contourlet transform: theory, design, and applications," *IEEE Transactions on Image Processing*, vol. 15, no. 10, pp. 3089-3101, 2006.
[16] D. Brunet, E. R. Vrscay, and Z. Wang, "On the mathematical properties of the structural similarity index," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1488-1499, 2012
[17] Q. Zhang and B. L. Guo, "Fusion of infrared and visible light images based on nonsubsampled contourlet transform," *Journal of Infrared and Millimeter Waves (Chinese Edition)*, vol. 26, no. 6, pp. 476-480. 2007.
[18] TNO UN Camp image [Online]. Available: http://www.imagefusion.org/.ftpquota.
[19] X. Li and S. Y. Qin, "Efficient fusion for infrared and visible images based on compressive sensing principle," *IET Image Processing*, vol. 5, no. 2, pp. 141-147, 2011.

Xin Feng https://orcid.org/0000-0001-8793-3775

He received Ph.D. degrees in School of Computer Science and Engineering from Lanzhou University of Technology in 2012. Since March 2013, he has been teaching in the School of Chongqing Technology and Business University.