Expression Recognition Algorithm Based on Infrared Image

Ying Cui[1] and Shi Qiu[2(✉)]

[1] College of Equipment Management and Support, Engineering University of PAP, Xi'an 710086, China
[2] Key Laboratory of Spectral Imaging Technology CAS, Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119, China

Abstract. It's important to recognize facial expressions in social communication. To solve the problem that facial expression recognition by visible light is vulnerable to interference, we built a model from the perspective of thermal infrared. Based on the distribution characteristics of thermal infrared, the face region is firstly located by building a multi-projection model toward color. Then, the level set function of the local Gaussian fitting model was optimized, the regular term was removed, and the larger iteration step size was selected to achieve accurate face segmentation on the premise of segmentation accuracy. Finally, based on the structure of traditional deep learning network, the characteristics of DPN and CBAM network are given full play to realize expression recognition by thermal infrared images.

Keywords: Infrared image · Face · Multi-projection · Level set · Identify

1 Introduction

Facial expression recognition is the main means of analyzing inner activity, which is of great significance in social and medical fields. At present, facial expression recognition is mainly based on visible images. Sarode [1] uses computers to recognize facial expressions. Berretti [2] extracted SIFT features to achieve expression matching. Jain [3] recognizes facial expressions based on the shape change model. Moore [4] used Local Binary Patterns to recognize expressions from multiple perspectives. Guo [5] classifies facial expressions through videos. Lajevardi [6] selected representative features to carry out the study. Liu [7] constructed a deep network to classify facial expressions. Luo [8] extracted PCA and LBP features and used SVM to classify facial expressions. Owusu [9] established a neural-AdaBoost to extract facial features hierarchically. Saeed [10] uses geometrical features to recognize the expression of a single frame image. Yu [11] constructed multiple deep networks to recognize static image expressions. Lopes [12] constructed convolutional networks to recognize expression. Chen [13] researched on multi-feature fusion based on video sequences to realize expression recognition. Elaiwat [14] established a spatio-temporal RBM to realize expression recognition. Xie

W. Fu et al. (Eds.): ICMTEL 2021, LNICST 388, pp. 321–330, 2021.
https://doi.org/10.1007/978-3-030-82565-2_26

[15] established the FRR-CNN network for expression recognition. Meng [16] proposed Identity-aware convolutional neural network for facial expression recognition. Zhang [17] analyzed the relationship between characters from facial expressions. Li [18] fused CNN with the attention model to realize expression recognition of a covered face. Georgescu [19] integrated depth features with handcrafted features to realize expression recognition. Shao [20] established 3D-CNN to realize expression classification. Wang [21] realized expression recognition based on figure posture and facial image. Li [22] introduced the attention mechanism to realize expression recognition. Visible light imaging is consistent with human perception and can objectively reflect the characteristics of objects. Although many studies have been carried out based on visible-light images, the following problems still exist: 1) It is easily disturbed by the external environment and focus on the face is hard. 2) The established model has limited utilization characteristics.

Given the above problems, based on the characteristics of thermal infrared imaging, we use the infrared images to carry out research. 1) The face extraction model was established, and the area of the face was focused. 2) Build an expression recognition model to analyze facial expressions.

2 Algorithm

To meet the needs of getting facial expression accurately, the algorithm in this paper builds a face extraction model based on the acquisition of thermal infrared images, focuses the face area, and then proposes an expression recognition network to realize expression recognition based on the existing deep learning framework. The specific flow chart is shown in Fig. 1.

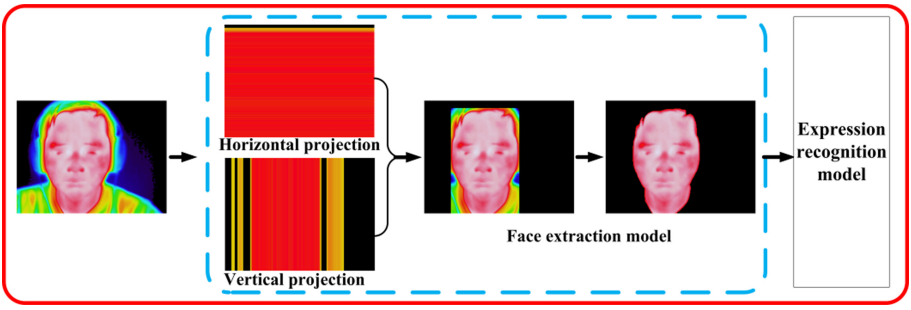

Fig. 1. Algorithm flow chart

2.1 Face Extraction Model

The infrared images can directly display the heat distribution [23]. The human body produces heat, which makes people with different skin tones tend to be red in the infrared

image. Based on this feature, facial detection of different skin tones can be realized. We can build a model to measure the similarity of colors:

$$D_{(SR,SG,SB)}(x, y) = |C_R(x, y) - SR| + |C_G(x, y) - SG| + |C_B(x, y) - SB| \quad (1)$$

where $C_R(x,y)$, $C_G(x,y)$ and $C_B(x,y)$ are pixel values of R, G, and B of the infrared image respectively, and (SR, SG, SB) is the target pixel value.

The facial region is continuous and tends to be red, so we construct the horizontal projection $\boldsymbol{M}_X$ and the vertical projection $\boldsymbol{M}_Y$ to determine the face region indirectly.

$$\begin{cases} \boldsymbol{M}_X = \min\left(D_{(SR,SG,SB)}(x, 1), \cdots, D_{(SR,SG,SB)}(x, n), \cdots, D_{(SR,SG,SB)}(x, H)\right) \\ \boldsymbol{M}_Y = \min\left(D_{(SR,SG,SB)}(1, y), \cdots, D_{(SR,SG,SB)}(n, y), \cdots, D_{(SR,SG,SB)}(W, y)\right) \end{cases} \quad (2)$$

where (SR, SG, SB) = (255, 0, 0). Then measure the difference between horizontal and vertical projections:

$$E(x, y) = \begin{cases} 0\left(\begin{array}{l} abs(M_X(x, y) - M_Y(x, y)) > T_1 || M_X(x, y) \to (0, 0, 0) \\ ||M_Y(x, y) \to (0, 0, 0)||I(x, y) \to (0, 0, 0) \end{array}\right) \\ I(x, y)\ others \end{cases} \quad (3)$$

Then determine the area where the face is, and then carry out the operation in this area.

According to the infrared imaging principle, the parts with poor heat dissipation are hair, behind ears, clothing, etc. [24]. The transition zone between face, clothing, and hair presents an obvious continuous red transition zone due to friction heating. Based on this feature, the facial contour can be extracted effectively.

Signed distance function is often used in traditional level sets [25]:

$$u(x, y, t) = \begin{cases} d\left[(x, y), C\right] \\ -d\left[(x, y), C\right] \end{cases} \quad (4)$$

where C is the evolution curve, $d[(x,y),C]$ is the distance function from the point(x,y) to curve C. Since the gradient of the level set is collinear with the normal, but the direction is opposite, the value of the point in a certain region is defined as negative inside and positive outside. The gradient modulus of the signed distance function is identical to 1, which ensures that the change of $u(x,y,t)$ is uniform everywhere and the numerical calculation is stable.

According to the curve evolution theory, the evolution equation of the level set function is:

$$\frac{\partial u}{\partial t} = \beta |\nabla u| \quad (5)$$

where β is the normal rate and t is the evolution time. After many iterations, the level set function deviates from the inner negative and outer positive characteristic of the signed distance function, thus destroying the stability of the iteration. When $|\bigtriangledown u| >> 1$, the level set function will be rush or ravines, cause the energy function into local minimum; When $|\bigtriangledown u| << 1$, the level set function is too flat, boundary fitting is very difficult.

Local Gaussian Distribution Fitting (LGDF) model [26] is a classic segmentation algorithm based on the level set active contour model, and its energy generic function is:

$$E(\varphi, u_1, u_2, \sigma_1, \sigma_2) = A(\varphi) + B(\varphi, u_1, u_2, \sigma_1, \sigma_2) \tag{6}$$

$$\begin{cases} A(\varphi) = \alpha \int_\Omega \frac{1}{2}(|\nabla\varphi| - 1)^2 dx + \beta \int_\Omega \delta(\varphi)|\nabla\varphi| dx \\ B(\varphi, u_1, u_2, \sigma_1, \sigma_2) = \lambda \iint C(\varphi, u_1, \sigma_1) H(\varphi(y)) dydx \\ \quad + (1-\lambda) \iint C(\varphi, u_2, \sigma_2)(1 - H(\varphi(y))) dydx \\ C(\varphi, u, \sigma) = K(x-y)\left(\log\sigma(x) + \frac{(I(y)-u(x))^2}{2\sigma^2(x)}\right) \end{cases} \tag{7}$$

where $A(\varphi)$ is the regular term, $K(x\text{-}y)$ is the Gaussian window, u_1 and u_2 are the local mean values inside and outside the contour, σ_1 and σ_2 are the local variances inside and outside the contour. The level set function is guaranteed to remain a signed distance function and to be smooth during evolution. $B(\varphi,u_1,u_2,\sigma_1,\sigma_2)$ is controlled by the local binomial fitting term $C(\varphi,u,\sigma)$.

LGDF model uses the mean value and variance to describe the local pixel distribution, which can effectively solve the segmentation problems of uneven grayscale and low contrast images, but at the same time, it also increases the calculation cost.

Aiming at the deficiency of the signed distance function, we built parameterized level set according to mathematical theory:

$$\varphi(x, \mathbf{W}) = 1 - \prod_{i=1}^{N}\left(1 - r_{ij}(x)\right) \tag{8}$$

$$r_{ij}(x) = \frac{1}{1 + \exp\left(\sum_{k=0}^{n} w_{ijk} x_k\right)} \tag{9}$$

where $\boldsymbol{W} = [w_{ijk}]$ is the target contour, and the updating process of parameter w_{ijk} is the contour evolution process. $r_{ij}(x)$ determined by w_{ijk} is used to represent the half-space, and the level set function is composed of polyhedra so that $\varphi(x,\boldsymbol{W}) \in [0,1]$ is guaranteed. The level set of $\varphi(x,\boldsymbol{W}) = 0.5$ is used as the boundary between foreground and background. When $\varphi(x,\boldsymbol{W}) > 0.5$, the level set function belongs to the foreground region, when $\varphi(x,\boldsymbol{W}) < 0.5$, the level set function belongs to the background region, so $\varphi(x,\boldsymbol{W})$ replaces the regular term. $\boldsymbol{W}$ adopts the interactive method, The user initializes φ by defining N faces in the region of interest in the image, and the polyhedron is approximately a half-sphere.

When $\varphi(x,\boldsymbol{W})$ is introduced into LGDF, the level set energy functional is:

$$\begin{aligned} P(W) = {} & \lambda \iint C(\varphi, u_1, \sigma_1)\varphi(x, W) dydx \\ & + (1-\lambda) \iint C(\varphi, u_2, \sigma_2)(1 - \varphi(x, W)) dydx \end{aligned} \tag{10}$$

$$\begin{cases} u(x) = \frac{\int K(y-x)I(y)H(\varphi-0.5)dy}{\int K(y-x)H(\varphi-0.5)dy} \\ \sigma^2(x) = \frac{\int K(x-y)\left(I(y-u(x))^2\right)H(\varphi-0.5)dy}{\int K(x-y)H(\varphi-0.5)dy} \end{cases} \tag{11}$$

The evolution process of φ requires neither the regular term nor initialization, which simplifies the energy functional and reduces the calculation cost.

Image segmentation using gradient descent flow:

$$\frac{\partial \varepsilon}{\partial w_{ijk}} = (\lambda_1 e_1 - \lambda_2 e_2)\frac{\partial \varphi}{\partial w_{ijk}} \tag{12}$$

$$\begin{cases} e_n(x) = \int K(x-y)C(\varphi, u, \sigma)dy \\ \frac{\partial \varphi}{\partial w_{ijk}} = x_k\left[r_{ij}(x)\right]\prod_{l=1}^{M}\left(1-r_{ij}(x)\right)\prod_{r\neq i, r=1}^{N}\left(1-\prod_{j=1}^{M}\left(1-r_{ij}(x)\right)\right) \end{cases} \tag{13}$$

In the iteration process, the discriminant parameter vector $\boldsymbol{W}$ is constantly updated:

$$w_{ijk} \rightarrow \tau\frac{\partial P}{\partial w_{ijk}} - w_{ijk} \tag{14}$$

where τ is the step, when $P(\boldsymbol{W})$ is the smallest, the parameter is optimal. Therefore, this algorithm is not restricted by the CFL standard and can choose the larger τ to accelerate convergence.

2.2 Expression Recognition Model

While the ResNet network does a good job of refining features, Densenet supports deeper detail exploration. DPN network combines the advantages of RESNET and DENSENET models. DPN network has a highly coupled two-channel link structure, which can effectively solve the problem of gradient disappearance in deep network training. By combining the channel selection mechanism with the spatial orientation selection mechanism, CPAM can achieve multi-directional convolution to obtain better results.

For this reason, we combined the two to construct a C-DPN model and normalized the size of the face image was sent into the Block. Both blocks have transitions. The Transition layer structure reduces the amount of data. There are 4 blocks in the C-DPN network, and the number of M-blocks in each Block determines the network depth, which can realize data partition. Rich feature information is extracted from the matrix through the Block, and the data dimension reduction is realized through the adaptive average pooling layer to reduce the computation. Adding the Dropout layer in front of the full connection layer can further reduce the feature redundancy, speed up the computing speed, and solve the overfitting problem to some extent. Finally, the feature matrix is expanded through the full connection layer, and the weight is obtained to realize expression state classification. Its network structure is shown in Fig. 2.

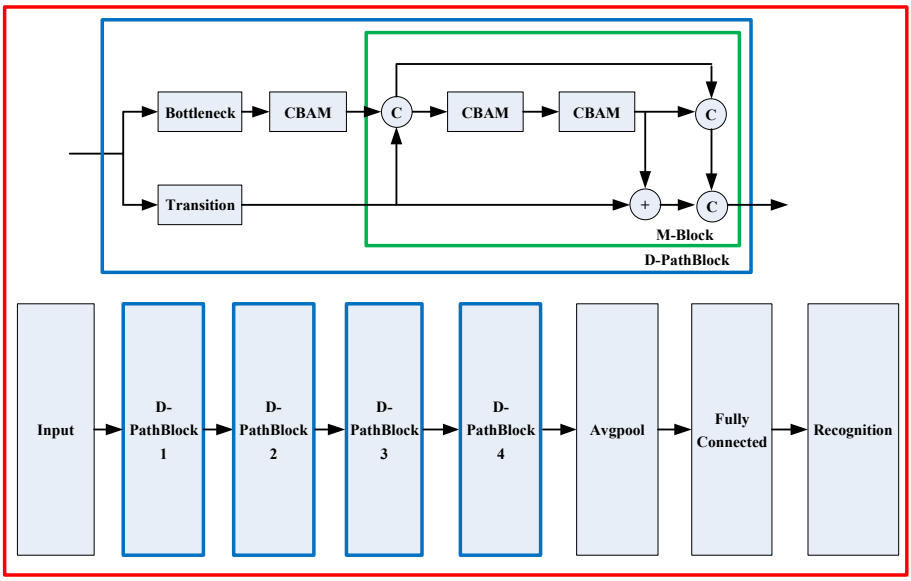

Fig. 2. Algorithm network structure

3 Experiment and Result Analysis

The database was composed of 150 frames of an image taken by a thermal infrared camera with 7 groups of different expressions, and the facial boundaries in the database were manually marked as the gold standard. The experimental equipment is a PC, equipped with Win7 operating system and VC++ software writing platform.

3.1 Algorithm Renderings

To measure the contour extraction effect of the algorithm, we introduce DC (Dice coefficient) to measure the detection accuracy [27]:

$$DC(\Omega_s, \Omega_r) = \frac{2Area(\Omega_s \cap \Omega_r)}{Area\Omega_s + Area\Omega_r} \tag{15}$$

where Ω_s is the manually marked result, as the gold standard. $\Omega_r$is the algorithm extracting effect. The closer the DC value is to 1, the better.

Compared with the traditional LGDF, MSLCV, and our algorithm, the results are shown in Table 1. The segmentation performance, number of iterations, and calculation time of the proposed algorithm are all optimal. This is because the parameterized level set function constructed in this paper replaces the regular term in LGDF and MSLCV algorithm to improve the segmentation performance. The algorithm needs no initialization and has strong robustness. The iteration step τ is not limited by CFL and can be increased to reduce the number of iterations without reducing the segmentation accuracy, which accelerates the curve evolution to the real contour.

Table 1. Comparison of algorithm effects

Image sequence	*DC*		
	LGDF	MSLCV	Ours
Angry	0.86	0.87	0.89
Disgust	0.84	0.88	0.91
Fear	0.80	0.82	0.86
Happy	0.87	0.89	0.92
Normal	0.86	0.88	0.90
Sadness	0.84	0.86	0.89
Surprise	0.86	0.87	0.90
Image sequence	Number of iterations		
	LGDF	MSLCV	Ours
Angry	564	124	61
Disgust	682	171	76
Fear	721	214	92
Happy	506	101	50
Normal	410	80	45
Sadness	680	160	72
Surprise	701	201	86
Image sequence	Computing time/s		
	LGDF	MSLCV	Ours
Angry	45.8	24.1	12.5
Disgust	48.7	26.5	14.6
Fear	52.4	27.8	18.4
Happy	43.2	31.6	20.6
Normal	40.3	35.5	24.5
Sadness	47.6	37.6	26.6
Surprise	50.1	39.5	30.4

In order to display the algorithm performance, we demonstrate the segmentation effect of 7 groups of data. It can be seen from Table 2 that the temperature of the skin is higher than that of the hair and clothing. The algorithm proposed in this paper takes full advantage of the characteristics of thermal infrared imaging to build a multi-projection model with color orientation, which can meet the requirements of facial localization at different scales. The improved LGDF model can accurately segment the facial region.

3.2 Facial Expression Recognition Algorithm

UF1 (Unweighted F1-score), UAR (Unweighted Average Recall) and ROC (Receiver Operating characteristic Curve) were used to measure the evaluation effect.

Table 2. Algorithm renderings

	Angry	Disgust	Fear	Happy	Normal	Sadness	Surprise
Orininal							
Facial positioning							
Facial segmentation							

$$UF1 = \frac{1}{C}\sum_{C}\frac{2TP}{TP + FP + FN} \tag{16}$$

$$UAR = \frac{1}{C}\sum_{c}\frac{TP}{N} \tag{17}$$

where C represents the total number of types, and TP, FP, and FN represent the proportion of calculated results.

By comparing our algorithm with the current mainstream algorithm, the results are shown in Table 3, and the recognition effect of counting angry expressions is shown in Fig. 3. Adaboost [9] algorithm uses multi-scale thought focusing feature to achieve feature extraction. Multiple deep network [11] algorithm builds the depth model from multiple angles, which enhances the robustness of the model. The FRR-CNN [15] integrates the two network structures to improve the accuracy of detection. D-H-F [19] integrates depth features with traditional features to realize expression recognition. The algorithm proposed in this paper gives full play to the advantages of DPN and CBAM, extracts image features and realizes expression recognition, and achieves good results.

Table 3. Comparison of algorithm effects

	AdaBoost	Multiple deep network	FRR-CNN	D-H-F	Ours
UF1	0.63	0.66	0.68	0.70	0.73
UAR	*0.65*	0.69	0.70	0.74	0.76

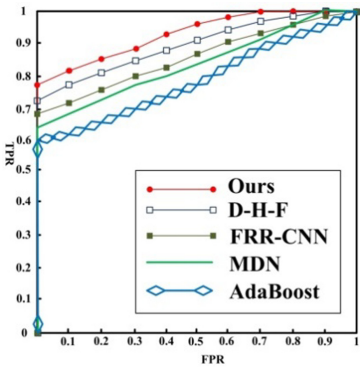

Fig. 3. Recognition effect of angry expressions

4 Conclusion

Because of the need to analyze facial expressions in social communication and the inaccuracy of facial expression analysis caused by the interference of visible light imaging, a new facial expression analysis algorithm based on the thermal infrared image was proposed. A color-oriented multi-projection model was established to locate the face region, the LGDF level set algorithm was optimized, and the parameterized level set framework was constructed to achieve fast and accurate face image segmentation. Taking full advantage of the network, the expression recognition model is built to analyze the expression. Follow-up studies on facial expression analysis and social relationship prediction will be carried out.

Acknowledgement. This work is supported by Postdoctoral Science Foundation of China under Grant No. 2020M682144. The Open Project Program of the State Key Lab of CAD&CG (Grant No. A2026), Zhejiang University.

References

1. Sarode, N., Bhatia, S.: Facial expression recognition. Int. J. comput. Sci. Eng. **2**(5), 1552–1557 (2010)
2. Berretti, S., Del Bimbo, A., Pala, P., Amor, B.B., Daoudi, M.: A set of selected SIFT features for 3D facial expression recognition. In: 2010 20th International Conference on Pattern Recognition pp. 4125–4128. IEEE (2010)
3. Jain, S., Hu, C., Aggarwal, J.K.: Facial expression recognition with temporal modeling of shapes. In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops) pp. 1642–164. IEEE (2011)
4. Moore, S., Bowden, R.: Local binary patterns for multi-view facial expression recognition. Comput. Vis. Image Underst. **115**(4), 541–558 (2011)
5. Guo, Y., Zhao, G., Pietikäinen, M.: Dynamic facial expression recognition using longitudinal facial expression atlases. In: European Conference on Computer Vision pp. 631–644. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33709-3_45

6. Lajevardi, S.M., Hussain, Z.M.: Automatic facial expression recognition: feature extraction and selection. SIViP **6**(1), 159–169 (2012)
7. Liu, M., Li, S., Shan, S., Chen, X.: Au-aware deep networks for facial expression recognition. In: 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG) pp. 1–6. IEEE (2013)
8. Luo, Y., Wu, C.M., Zhang, Y.: Facial expression recognition based on fusion feature of PCA and LBP with SVM. Optik-Int. J. Light and Electron Opt. **124**(17), 2767–2770 (2013)
9. Owusu, E., Zhan, Y., Mao, Q.R.: A neural-AdaBoost based facial expression recognition system. Expert Syst. Appl. **41**(7), 3383–3390 (2014)
10. Saeed, A., Al-Hamadi, A., Niese, R., Elzobi, M.: Frame-based facial expression recognition using geometrical features. Adv. Hum. Comput .Interact. **2014**, 1–13 (2014)
11. Yu, Z., Zhang, C.: Image based static facial expression recognition with multiple deep network learning. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction pp. 435–442 (2015)
12. Lopes, A.T., De Aguiar, E., Oliveira-Santos, T.: A facial expression recognition system using convolutional networks. In: 2015 28th SIBGRAPI Conference on Graphics, Patterns and Images pp. 273–280. IEEE (2015)
13. Chen, J., Chen, Z., Chi, Z., Fu, H.: Facial expression recognition in video with multiple feature fusion. IEEE Trans. Affect. Comput. **9**(1), 38–50 (2016)
14. Elaiwat, S., Bennamoun, M., Boussaïd, F.: A spatio-temporal RBM-based model for facial expression recognition. Pattern Recogn. **49**, 152–161 (2016)
15. Xie, S., Hu, H.: Facial expression recognition with FRR-CNN. Electron. Lett. **53**(4), 235–237 (2017)
16. Meng, Z., Liu, P., Cai, J., Han, S., Tong, Y.: Identity-aware convolutional neural network for facial expression recognition. In: 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017) pp. 558–565. IEEE (2017)
17. Zhang, Z., Luo, P., Loy, C.C., Tang, X.: From facial expression recognition to interpersonal relation prediction. Int. J. Comput. Vis. **126**(5), 550–569 (2018)
18. Li, Y., Zeng, J., Shan, S., Chen, X.: Occlusion aware facial expression recognition using CNN with attention mechanism. IEEE Trans. Image Process. **28**(5), 2439–2450 (2018)
19. Georgescu, M.I., Ionescu, R.T., Popescu, M.: Local learning with deep and handcrafted features for facial expression recognition. IEEE Access **7**, 64827–64836 (2019)
20. Shao, J., Qian, Y.: Three convolutional neural network models for facial expression recognition in the wild. Neurocomputing **355**, 82–92 (2019)
21. Wang, K., Peng, X., Yang, J., Meng, D., Qiao, Y.: Region attention networks for pose and occlusion robust facial expression recognition. IEEE Trans. Image Process. **29**, 4057–4069 (2020)
22. Li, J., Jin, K., Zhou, D., Kubota, N., Ju, Z.: Attention mechanism-based CNN for facial expression recognition. Neurocomputing **411**, 340–350 (2020)
23. Baker, E.A., Lautz, L.K., McKenzie, J.M., Aubry-Wake, C.: Improving the accuracy of time-lapse thermal infrared imaging for hydrologic applications. J. Hydrol. **571**, 60–70 (2019)
24. Raccuglia, M., Heyde, C., Lloyd, A., Hodder, S., Havenith, G.: The use of infrared thermal imaging to measure spatial and temporal sweat retention in clothing. Int. J. Biometeorol. **63**(7), 885–894 (2019). https://doi.org/10.1007/s00484-019-01701-5
25. Vercruysse, D., Sapra, N.V., Su, L., Trivedi, R.: Analytical level set fabrication constraints for inverse design. Sci. Rep. **9**(1), 1–7 (2019)
26. Li, Y., Cao, G., Yu, Q., Li, X.: Active contours driven by non-local Gaussian distribution fitting energy for image segmentation. Appl. Intell. **48**(12), 4855–4870 (2018). https://doi.org/10.1007/s10489-018-1243-x
27. Qiu, S., Luo, J., Yang, S., Zhang, M., Zhang, W.: A moving target extraction algorithm based on the fusion of infrared and visible images. Infrared Phys. Technol. **98**, 285–291 (2019)