Infrared Thermography and Big Data for Detection of People with Fever and Determination of High-Risk Areas in Epidemic Situations

*ROBALINO ESPINOZA Viviana Lorena**, *TAMAYO FREIRE Alexis Shipson*

College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, P.R. China

(Received 20 May 2021; revised 20 June 2021; accepted 30 June 2021)

Abstract: Technological advances in computer science and their application in our daily life allow us to improve our understanding of problems and solve them effectively. A system design to detect people with fever and determine high-risk areas using infrared thermography and big data is presented. In order to detect people with fever, face detection algorithms of Viola-Jones and Kanade-Lucas are investigated, and comparison between them is presented using a training set of 406 thermal images and a test set of 2 072 thermal images. Thermography analysis is performed on detected faces to obtain the temperature level on Celsius scale. With this information a sample database is created. To perform big data experimental analysis, Power Bi tool is used to determine the high-risk area. The experimental results show that Viola-Jones algorithm has a higher performance recognizing faces of thermal images than Kanade-Lucas, having a high detection rate, less false-positives rate and false-negatives rate.

Key words: face detection; thermography; image analysis; big data

CLC number: TP391 **Document code:** A **Article ID:** 1005-1120(2021)S-0122-07

0 Introduction

Coronaviruses (CoV) in the virus family cause various conditions, from a simple cold to a severe illness. Middle Eastern Respiratory Syndrome coronavirus (MERS-CoV) and severe acute respiratory syndrome coronavirus (SARS-CoV), Covid-19 announced on 11 March 2020 by the World Health Organization (WHO), presents symptoms such as fever, tiredness, dry nasal discharge, sore throat, or dysentery[1]. These symptoms begin gradually and approximately have an infection fatality ratio of 0.5%—1%, which is the ratio of the number of deaths from the disease by the number of infected individuals[2-3]. The human body emits infrared radiation when the temperature is above absolute zero (−273 ℃). The radiation amount emitted depends directly on body temperature. Most of the electromagnetic spectrum, such as microwave radiation and infrared wavelength between 0.7 to 1 000 microns, are invisible to the human eye; however, they can be captured with infrared cameras. A person is considered to have a fever when their temperature exceeds 37 ℃[4], customarily measured with thermometers, but this process can be automated using computer vision techniques. Humans perform face recognition without any apparent effort. The automatic identification of faces by a computer is an important part using facial recognition algorithms. The problem is that some algorithms like Hough transform, Canny, principal component analysis (PCA), among others, are geometrics which can perform facial recognition but have a different purpose, like detecting edges[5]. Instead, Viola-Jones and Kanade-Lucas are focused on face recognition.

1 Related Work

Viola-Jones algorithm applies a cascade classifi-

*Corresponding author, E-mail address: Vivianalorena123@gmail.com.

How to cite this article: ROBALINO ESPINOZA Viviana Lorena, TAMAYO FREIRE Alexis Shipson. Infrared thermography and big data for detection of people with fever and determination of high-risk areas in epidemic situations[J]. Transactions of Nanjing University of Aeronautics and Astronautics, 2021, 38(S): 122-128.

http://dx.doi.org/10.16356/j.1005-1120.2021.S.015

er using the machine learning method. The descriptors use Haar-like features, which are digital image features used in object recognition for facial detection. If it finds a face, this fragment is passed to the next classifier.

Techniques such as integral image and cascading classifiers make the Viola-Jones algorithm highly efficient[6]. Fig.1 shows how these cascade classifiers work.

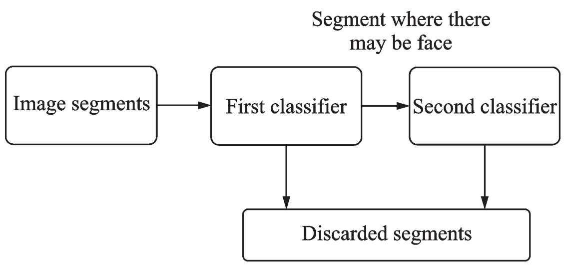


Fig.1 Viola-Jones cascade classifier

The training process uses AdaBoost (Adaptive Boosting) to select features and construct the classifier. A extensive set of images with a size corresponding to the size of the detection window is prepared. Each image corresponds to a value $y_1=1$ for face detection and $y_1=0$ for non-face detection.

For $y_1=0$ and $y_1=1$, where p_- and p_+ are the number of non-faces and faces in the image, respectively, initialized weights

$$w_{i,l} \frac{1}{2p_-}, \frac{1}{2p_+} \tag{1}$$

Normalized the weights as follow so that $w_{i,l}$ is a probability distribution

$$\frac{w_{i,l}}{\sum_{l=1}^{n} w_{i,l}} \tag{2}$$

Train a classifier h_j which is restricted to using a single feature, for each feature j, the classifier error rate is evaluated to $w_{i,1}$

$$\varepsilon_j = \sum_{l=0}^{L-1} w_{i,l} \left| h_j(x_l) - y_t \right| \tag{3}$$

Choose the classifier h_j, with the lowest error ε_i.Update weights

$$\beta_i = \frac{\varepsilon_i}{1-\varepsilon_i} \tag{4}$$

The final classifier is

$$\alpha_i = \lg \frac{1}{\beta_i} \tag{5}$$

The algorithm uses the value of two features. Adding the pixels within two rectangular regions, the regions have the same size and shape, vertically and horizontally adjacent, then a rectangle calculates the difference between diagonal pairs, finally having a detector resolution of 24×24 pixels.

Kanade Lucas algorithm is focused on dense alignment problems in images and objects. In the algorithm, there is a linear relationship between the geometric displacement and the appearance of the pixels; this relationship is rarely linear, so a linearization in the process is repeated until convergence. The intensity of the pixels is not so deterministic with the geometric displacement. In the linear relationship it must be established stochastically by a learning process. This algorithm has an efficiency value with which it can estimate the linear relationship, and this efficiency value allows the dependence between the coordinates of the pixels. The parameters of the linear relationship are called the image gradients; finite differentiation operations estimate the gradients. It detects a series of points in a sequence of images; These points are tracked based on linear translation, but the points offset deformation can appear in different correlative macros. This algorithm fails on the boundaries, which means that if the image has excessive motion. The scattering points go outside the local window, so Kanade-Lucas algorithm cannot process these points[7].

The tracking of the facial region of Kanade Lucas depends on the movement of the feature centers of successive sides, the process of facial tracking is shown in Eqs.(6—8)

$$R_t = R_{t-1} + (C_t - C_{t-1}) \tag{6}$$

$$C_t = \frac{1}{|f_t|} \sum_i f_t(i) \tag{7}$$

$$C_{t-1} = \frac{1}{|f_{t-1}|} \sum_i f_{t-1}(i) \tag{8}$$

where R_t and R_{t-1} represent the areas of faces in two adjacent video frames, f_t and f_{t-1} the landmarks of the current and previous frames, C_t and C_{t-1} the position centers of entities in two consecutive frames.

The Kanade-Lucas algorithm detects the person's face in video frames through the cascading object detector used in the Viola-Jones algorithm; In addition, the algorithm uses a classification model trained for face detection. In the face tracking the

Kanade-Lucas-Tomasi (KLT) algorithm uses the cascade object detector in each frame. Other researches show that this algorithm has a high computational cost, fail to detect the face when the person turns or tilting the head. This limitation is because of the type of classification model used for face detection[8].

In order to evaluate these algorithms, three metrics are used. Detection rate is defined as the fraction of detected faces by the number of images in the dataset. The false-positive rate is the proportion of all negatives that still yield positive test outcomes, and the false-negative rate is the proportion of positives that yield negative test outcomes with the test[9].

The histogram analysis indicates the temperature through a temperature distribution, transforming the intensity of the image to length intervals to obtain values from the thermal images, allowing to analyze the presence of fever[10].

Big data is a large amount of data that exceeds the processing capacity of databases. To get value out of this, this data must be processed. The characteristics commonly used are the three Vs: Volume, velocity, and variety, which are very helpful in understanding the nature of the data and the platform to use it. When placing the data in a test area, it is necessary to combine extraction, transformation, loading (ETL) and transform it into a format that facilitates its analysis, allowing access to clean data for analysis after loading it into the database. It is essential that data transformations involve quality assessment prior to data analysis, as its structure must be correct for analysis. For extraction, loading, and transformation (ELT), an inventory must be carried out, and the current data available in the dataset must be compared. In the sub-phase, the available data is extracted, and the connections to the data that are in process are determined, such as online transactions processing (OLTP), online analytical processing (OLAP), among other sources.

The data analysis lifecycle showed in Fig.2 defines recommended practices in the analysis process from project selection to completion. A lifecycle is a set of methods established for data analysis and decision; the data analysis lifecycle includes six phases.

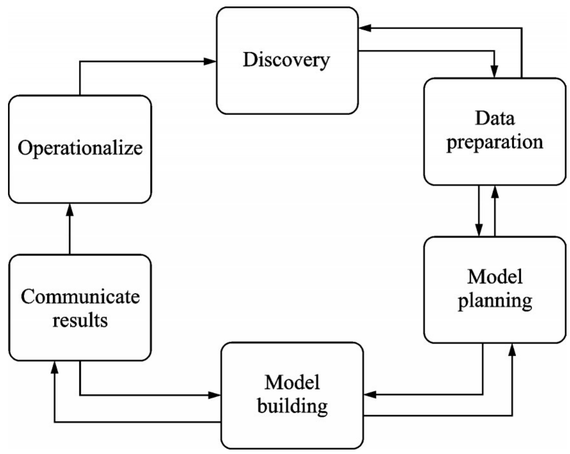


Fig.2 Data analytics lifecycle

(1) Discovery: Important activities in this phase include framing the problem as an analytics challenge begin learning the data, and formulating initial hypotheses to test.

(2) Data preparation: Requires of an analytics sandbox to perform analytics which allows the team to familiarize with the data and conditioning it.

(3) Model planning: In this phase, the team explores the data and learns the relationship between variables to selects the most appropriate models and variables, determining techniques, methods, and workflow to follow in the following phase.

Model building: The team develops datasets for testing, builds, and executes models based on the previous phase, also considers whether its existing tools will suffice for running the models.

(4) Communicate results: The team determines the project's success according to the criteria of the first phase, identifies key findings, and develops a narrative to summarize and convey findings.

Operationalize: The team gets final reports and may implement the models in a production environment.

The tool used for analysis is Power Bi which allows the visualization of large amounts of data using a tabular data model obtaining reports or dashboards quickly and intuitively[11].

Previous related research shows that using the Otsu method in Viola-Jones at the preprocessing stage reduces false-positive rate[12]. The Viola-Jones algorithm shows a good performance on face detection using the thermal images in LWIR spectrum and visible spectrum compared with Gabor feature extraction and classification using support vector machines[13]. The recognition of head movements based on the Lucas-Kanade method identifies the movements of the head with precision by detecting the nostrils with the Lucas-Kanade optical flow algorithm in real time[14]. The different comparative studies of local histogram equalization, global histogram equalization and fast quadratic dynamic equalization help to the graphical representation of the probability of occurrences by equating the probability distribution of the occurrence of values with the intensity values in the image[15]. Another research shows that the detection of faces with eyeglasses in thermal images, Viola-Jones performed well in contrast with the predefined group truths[16]. These investigations do not compare Viola-Jones with Kanade-Lucas algorithm and do not use big data, so the research objective determines the need of design system using infrared thermography and big data to detect people with fever and determinate high-risk areas in epidemic situations.

2 Methodology

The experiment is carried out in four stages. The first stage select an efficient algorithm to detect faces. We use the dataset IRIS Thermal/Visible Face Database available in Ref.[17], which uses a thermal sensor Raytheon Palm-IR-Pro and has 4 228 thermal images of faces. We selected 406 thermal images with lighting variations and poses with a resolution of 320×240 as the training dataset. 2 000 thermal images with variations in lighting and poses with a resolution of 320×240, and 72 images of everyday objects such as Kettle and Torch with a resolution of 256×256 obtained from a random search on the World Wide Web, as a test set to compare main algorithms using Matlab 2015a and get the detection rate, false-positive and false-negative rates.

The second stage consists of thermographic analysis, which is performed using histogram analysis on the test set of 2 072 thermal images and objects; this analysis indicates the temperature on a fixed scale. The temperature distribution is taken from the thermal image. If the temperature exceeds 37 Celsius degrees it will be considered as a fever case. To perform the following stages we will assign a date randomly (dd/mm/yyyy), time (hh:mm:ss), and a district name Jianye, Xuanwu, Qixia and Pukou which will represent an area in Nanjing to illustrate the experiment on a map.

The third stage consists of data collection. MySql database tool is used to store the date, time, thermographic result, latitude, longitude, and district name. With the previous stages a database is created. The gathered data is prepared and checked for missing data and outliners using statistical software R included in Power Bi tool.

The fourth stage consists of big data analysis. The database created in the previous stage is imported to the Power Bi tool where it is analyzed calculating the growth of fever cases percentage per month and location through line chart, and showing circles of different sizes on a map depending on the number of cases found, determining the high-risk areas.

3 Experiment

3.1 First stage: Algorithm selection

Viola-Jones and Kanade Lucas algorithms are compared to determine which is more efficient in detecting faces using thermal images.

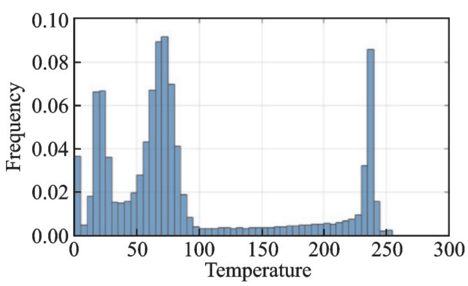
A training set of 406 thermal images with lighting variations and poses with a resolution of 24×24 is used to train the algorithm to select features and construct the classifier. A test set of 2 000 thermal images with variations in lighting and poses with a resolution of 320×240, and a set of 72 images of everyday objects with a resolution of 256×256. Fig.3 shows a sample of the thermal image dataset.

Fig.3 Thermal image dataset sample

3. 2 Second stage: Thermographic analysis

This analysis indicates the temperature on a fixed scale; histogram consists of transforming the intensity of the input image obtaining length intervals, the histogram vector contain integer counts to obtain appropriate range intensity values of [0, 1] for double class images, [0, 255] for unit8 and [0, 65 535] class images for unit16 class images. In Matlab 2015a, the histogram command uses the binning algorithm to automatically obtain results in bins with a uniform width with a range of elements in x, and underlying shows the distribution. The bins are shown in rectangles each rectangle indicates the number of elements in the bin.

(a) Floating point thermal (temperature) image



(b) Histogram of thermal temperature image

Fig.4 Thermographic analysis

3. 3 Third stage: Data collection

This database has seven entities ID: dat, tim, temperature, locationLa, locationLo, ubicat. Temperature entity took the data from the thermographic analysis. Date (dd/mm/yyyy), time(hh:mm:ss), and district name Jianye, Xuanwu, Qixia and Pukou with its respective coordinates randomly assign, which represent an area in Nanjing to illustrate the experiment on a map. The gathered data are prepared and checked for missing data and outliners getting 1 922 samples in a MySql database.

3. 4 Fourth stage: Big data analysis

Power Bi tool attracts all the information to analyze in five main steps:

(1) Gather: The data is collected from the third stage and placed in a single location; this data is transformed, resulting in clean data without errors and inconsistencies.

(2) Store: The data is stored and transformed; this is known as a data warehouse stored in a specialized format to allow recovery during the report and analysis. It is also structured so that thousands or millions of transactions can be added efficiently.

(3) Model: The data model shows in Fig.5 presents the data to the user without table name, foreign fields, or complex language.

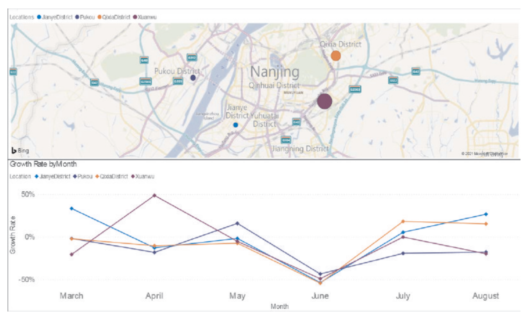


Fig.5 Data model

(4) Visualize: After data is available in the model, graphs, tables, and dashboards shown in Fig.6 and Fig.7 are used to visualize the report.

(5) Share: In order to be shared for visualizations, a common location is created where users can go to find the reports they need. The location is securely secured to ensure that authorized users can access the data.

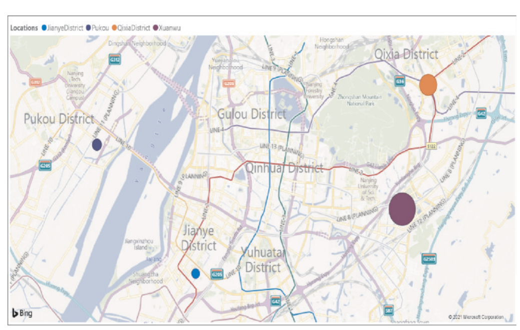


Fig.6 High-risk areas illustrated on a map

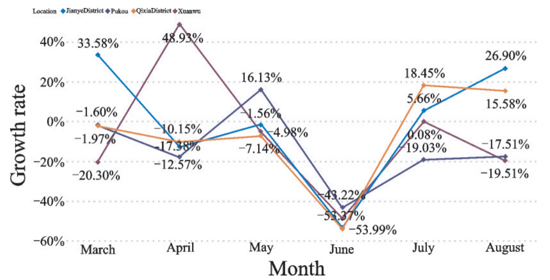


Fig.7 Growth rate of fever cases per month and location

4 Analysis

As shown in Table 1, the experimentations results show that Viola-Jones has a correct detection rate of 92.8%, a false-positive rate of 6.7%, and a false-negative rate of 0.5%. Kanade-Lucas has a correct detection rate of 86.2%, a false-positive rate of 10.5%, and a false-negative rate of 3.3%. Thermography analysis results show the transformation of the intensity of thermal image into image intensity ranges between 0 and 255 determining temperature, the gathered data is checked for missing data and outliners getting 1 922 samples in a MySql database, 138 false-positives and 12 false-negatives are discarded from the database to perform data big analysis. Showing the growth rate of fever cases in Xuanwu district increases by 48.93% in April compared to March, which makes it a high-risk area. Then in May, the number of cases decreases 7.14% compared to April. On the other hand, in May, Pokou has an increase of 16.13% compared to April. So it is considered a high-risk area.

Table 1 Face detection, false-positive and false negative on the test set containing 2 072 images %

Algorithm	Detection rate	False positive	False negative
Viola-Jones	92.8	6.7	0.5
Kanade-Lucas	86.2	10.5	3.3

For June, a considerable reduction is seen in all areas reaching the lowest levels during this period analyzed. However, in July, two of the four areas are considered high-risk, showing the Qixia district an increase of 18.45%, and Jianye district an increase of 5.66% compared to June. Finally, in August, a high-risk level remains in areas of Jianye district and Qixia district, which shows an increase of 26.90% and 15.58% respectively and the Xuanwu district are considered low risk showing a decrease of 19.51% compared to July. Furthermore, the map shows the high-risk areas with circles illustrating the higher-risk area with a larger circle size than the lower-risk areas.

5 Conclusions

We have presented a system design to detect people with fever. This system uses the Viola-Jones algorithm to detect faces on thermal images. The results show low false-positives rates and low false-negatives rates. Once the face is detected, thermal analysis of each face determines its temperature; these data are collected to create a database for subsequent big data planning and analysis, as well as interpretation. The appropriate model and variables to be used are determined. Constructing the big data analytics model is used to determine the high-risk areas according to fever cases growth rate per month and area.

References

[1] CUCINOTTA D, VANNELLI M. WHO declares COVID-19 a pandemic[J]. Acta Bio-medica: Atenei Parmensis, 2020, 91(1): 157-160.

[2] PEREZ F J, LAUER S A, KAISER L, et al. Serology-informed estimates of SARS-COV-2 infection fatality risk in Geneva[M]. Switzerland: MedRxiv, 2020.

[3] STRINGHINI S, WISNIAK A, PIUMATTI G, et al. Seroprevalence of anti-SARS-COV-2 IgG antibodies in Geneva[M]. Switzerland: The Lancet, 2020.

[4] GAUSSORGUES G, CHOMET S. Infrared thermography[M]. United Kingdom: Spring Science & Business Media, 1993.

[5] LI Y, CHEN L, HUANG H, et al. Nighttime lane makings recognition based on Canny detection and Hugh transform[C]//Proceedings of IEEE International Conference on Real-Time Computing and Robotics. Cambodia: IEEE, 2016: 415-441.

[6] VIOLA P, JONES M. Rapid detection using a boosted cascade of simple features[C]//Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Kauai: IEEE, 2001.

[7] SAHA A, DAS A. A method for tracking of facial action points using pyramidal Lucas Kanade algorithm

for expression recognition[J]. Multim Process Technol, 2017, 8: 133-141.

[8] MSTAFA R J, ELLEITHY K M. A video steganography algorithm based on Kanade-Lucas-Tomasi tracking algorithm and error correcting codes[J]. Multimedia Tools and Applications, 2016, 75(17): 1-23.

[9] RUSSELL J. Machine learning fairness in justice systems: Base rates, false positives, and false negatives [C]//Proceedings of 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA). Miami: IEEE, 2020: 817-820.

[10] KUMAR C S, ROY R, KUMAR A P, et al. Analysis of histogram processing for brain MRI using Matlab [J]. International Journal of Innovative Research in Engineering & Science, 2015, 3: 20-30.

[11] WILLEY J. Data science & big data analytics: Discovering, analyzing, visualizing and presenting data [M]. Canada: John Wiley & Sons Inc, 2015.

[12] MA C, TRUNG N T, UCHIYAMA H, et al. Adapting local features for face detection in thermal image [J]. Sensors, 2017, 17(12): 2741.

[13] TROFIMOV V S, SHESTAKOV I L, BLEDNOV R G. About possibility of temperature trace observing on the human skin using commercially available IR camera[J]. Proceedings of the SPIE, 2016, 9971(1): 1-5.

[14] ZHAO Z, WANG Y, FU S, et al. Head movement recognition based on Lucas- Kanade algorithm [C]// Proceedings of 2012 International Conference on Computer Science and Service System. Nanjing: Computer Science & Service System, 2012: 2303-2306.

[15] SURABHI P, MAUSUMI G. Comparative analysis of histogram equalization techniques[C]//Proceedings of 2014 International Conference on Contemporaty Computing and Informatics (IC3I). India: Computer Science, 2014: 167-168.

[16] YASHWANTH G, MHETRE B M, AGARWAL A. Thermal to visual face recognition using transfer learning[C]//Proceedings of Identity Security and Behavior Analysis (ISBA) 2019 IEEE 5th International Conference. India: IEEE, 2019: 1-8.

[17] RIAD H, JAMES W. IEEE OTCBVS WS series bench [EB/OL]. (2006-03-04)[2020-08-12]. http://vcipl-okstate.org/pbvs/bench/index.html.

Author Ms. **ROBALINO ESPINOZA Viviana Lorena** is a postgradute student at College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China. She received her B.S. degree in Electronics, Telecommunications and Networks from the Polytechnic School of Chimborazo, Riobamba, Republic of Ecuador, in 2018. Her research interests include digital image processing, databig, learningdeep, visionmachine, applications in telecommunications and networks.

Author contributions Ms. **ROBALINO ESPINOZA Viviana Lorena** designed the study and wrote the manuscript. Mr. **TAMAYO FREIRE Alexis Shipson** conducted the analysis and wrote the manuscript. All authors commented on the manuscript draft and approved the submission.

Competing interests The authors declare no competing interests.

(Production Editor:LIU Yandong)

红外热像和大数据用于检测发热人群和高危地区

ROBALINO ESPINOZA Viviana Lorena,
TAMAYO FREIRE Alexis Shipson
(南京航空航天大学电子信息工程学院,南京 211106, 中国)

摘要:提出了一种利用红外热像仪和大数据对发热人群进行检测并确定高危区域的系统设计方法。为了检测发热人群,研究了Viola-Jones和Kanade-Lucas两种人脸检测算法,并对其使用406张热图像训练集和2 072张热图像测试集进行了比较。对检测到的面部进行热成像分析,以获得摄氏度尺度上的温度水平。通过此信息创建了一个示例数据库。在进行大数据实验分析时,使用Power Bi工具确定高风险区域。实验结果表明,Viola-Jones算法对热图像人脸的识别性能优于Kanade-Lucas算法,这种算法检出率高,假阳性和假阴性率低。

关键词:人脸检测;热成像;图像分析;大数据