Video Stereo Grid Construction Method for Accurate Forest Fire Location

Jichang Cao[1,2(✉)], Yichao Cao[3], Qing Guo[1], Liang Ye[1], and Jialing Zhen[1]

[1] School of Electronic and Information Engineering,
Harbin Institute of Technology, Harbin 150001, Heilongjiang, China
caojichang@126.com
[2] Science and Technology and Industrialization Development Center of Ministry of Housing and Urban-Rural Development, Beijing 100835, China
[3] School of Automation, Southeast University, Nanjing 210096, Jiangsu, China

Abstract. Accurate forest fire location during the initial stages is vital for suppressing forest fires. However, in some mountainous areas, accurate automatic fire location is still a challenging task. There is a lack of a good association mechanism between video space and external geographic information. In this paper, we propose a method to measure the fire location and accurately map it to the stereo space coding, which uses the feedback parameters (yaw angle, pitch angle, geographic location and altitude where the camera installed) and digital elevation model. The position of the fire point from the image coordinate system is obtained by computer vision method, and then transform it to the camera coordinate system and the world coordinate system by coordinate transformation. And we establish the mapping from longitude, latitude and height coordinates into stereo grid code to achieve the construction of video stereo grid space. The proposed method has also been deployed and verified in actual forest video monitoring system.

Keywords: Forest fire location · Video stereo space · Grid space

1 Introduction

1.1 Video Based Forest Fire Detection

Forest fire spreads in a large area in the forest area, and causes a lot of losses to the forest ecosystem. They are one of the most frequent, devastating, and difficult natural disasters in the world. They are the greatest threat to the security of forest resources. China is an ecologically fragile country and a country prone to forest fires. The task of forest fire prevention is the primary task of national forest resource protection, and it is related to the safety of people's lives and property and forest resources.

Forest loss is the most direct loss caused by forest fires. At the same time, forest fires will damage the ecological environment and cause the loss of assets in the disaster area. In order to suppress forest fires, the government needs to invest a lot of manpower, material and financial resources, and the rescue process will also cause serious loss of

S. Shi et al. (Eds.): 6GN 2021, LNICST 439, pp. 182–191, 2022.
https://doi.org/10.1007/978-3-031-04245-4_16

people's lives and property. Therefore, early detection of forest fires and timely and efficient rescue measures are of great significance for protecting forest resources and reducing loss of life and property.

In recent years, with the development of video technology and computer vision technology, video-based forest fire video surveillance systems have provided new approaches and methods for forest fire detection [1]. Forest fire video monitoring technology is a non-contact forest fire monitoring technology using computer vision. The traditional forest fire detection technology uses sensor equipment distributed in the forest area to monitor. But the deployment and maintenance costs are generally high. In recent years, computer hardware and image processing technology have begun to make great progress, and forest fire video surveillance systems have begun to occupy a more important position in forest fire detection with its unique advantages [2].

Front-end monitoring equipment used for forest fire prevention usually includes monitoring cameras, lenses, drive motors, network equipment, etc. The digital camera devices with parameter feedback can obtain the camera's pitch, yaw, spatial position, and lens parameters in real time. By analyzing input images and these important parameters, accurate fire positioning could be achieved.

1.2 Video Geographic Data Association

The research on the integration of video data and geographic information began in 1970. Andrew Lippman [3] proposed the integration of video and geographic information for the first time and completed a set of interactive video geographic information systems. With the continuous development of positioning technology, data storage technology, data computing technology and other technologies, video and geographic data are more and more closely related, and many excellent research results have been obtained.

Geographic Information System (GIS) has been developed since the 1960s and is used to obtain, store, operate, analyze, manage and display all types of geographic data [4]. The video geographic information system is designed to combine geographic information and video to dynamically generate hyper-video that can be navigated through geographic content [5]. A large number of videos are organized and used by integrating with geographic information and using GIS research methods to form a variety of organizations and expressions. The integration of video and geographic data can be divided into real-time association and later through external data association from the different organizational association methods.

Kim et al. [6] proposed to use sensors to collect metadata while collecting videos, and store them in the database in real time to associate with the video frames, and proposed a feasible technical framework for this method. Lewis et al. [7] through in-depth research on the integration of video and geographic information, proposed a flexible and versatile and easy to expand spatial video data model ViewPoint, which can be applied to 2D and 3D GIS analysis and visualization [7]. Ying Lu et al. [8] built a new type of R-tree index to store the field of view (FOV), location and other information of the video, thereby improving the efficiency of video query and retrieval based on geographic information and meeting real application requirements. The current video geographic data organization and expression methods are mostly associated and organized with video as an independent object and geographic information, without in-depth exploration

of the video space; and the organization based on video semantics is also a relatively shallow way of in-video events or The organization of moving targets does not carry out spatial modeling and data organization and management of the entire video space. Therefore, when the video is associated with geographic data, it is necessary to choose an appropriate method to fully organize and express the video space, and the data structure of the multi-level grid has the characteristics of multi-scale, which can adapt to the needs of multiple applications.

In addition to expressing geographic data by latitude and longitude, there are many methods of data modeling expression to supplement geographic information. Among them, the earth subdivision grid has the characteristics of multi-scale and flexible expression, and has been widely used in many geographic data modeling application scenarios. The earth subdivision grid is a way of simulating the earth sphere by continuously subdividing and fitting the earth sphere with a grid. According to the different generation methods, the geospatial grid division model can be divided into: division based on latitude and longitude and division of physical elements [9–11]. Since latitude and longitude is currently the most commonly used method for describing geographic data coordinates, the following will mainly introduce the division method based on latitude and longitude.

The division based on latitude and longitude has two forms of equal division and equiangular division according to the different ways of equal division, both of which are based on latitude and longitude. The equal area method is to consider the area of the grid unit and divide the earth space into equal areas. The division method is more complicated. Literature [12] proposes a projection method to project the earth onto a plane and then back to a spherical surface to obtain a grid of latitude and longitude divided by equal areas. The equiangular method means that the latitude and longitude of the unit grid differ by integer multiples, and the different levels are recursively divided. For three-dimensional video space, the grid structure can provide a feasible and convenient method for modeling and expression. The earth division grid based on latitude and longitude is well integrated with the existing theoretical systems and is easy to expand the model. The stereo data model can express the video space at multiple scales and fully reflect the geographic attributes of the video space. On the basis of three-dimensional grid coding, based on the needs of video applications, the extended grid model expression can fully meet the needs of video space modeling.

2 Proposed Method

2.1 Latitude and Longitude Location Method

In this paper, we propose a forest fire location method based on computer vision method and digital elevation model. The key idea is to calculate the position of the fire point from the image coordinate system by computer vision algorithm, and then transform it to the camera coordinate system and the world coordinate system by coordinate transformation. After the fire point is detected from the frame, the sight from monitoring pan-tilt-zoom camera to the fire point is generated firstly. After the introduction of digital elevation model, we can obtain the surface height information within the monitoring

range. Then, we calculate the corresponding intersection of the line of sight on the elevation model. Finally, we can get the 3D spatial position code from the longitude and latitude information of the intersection. The overall flow chart of proposed method is shown in Fig. 1.

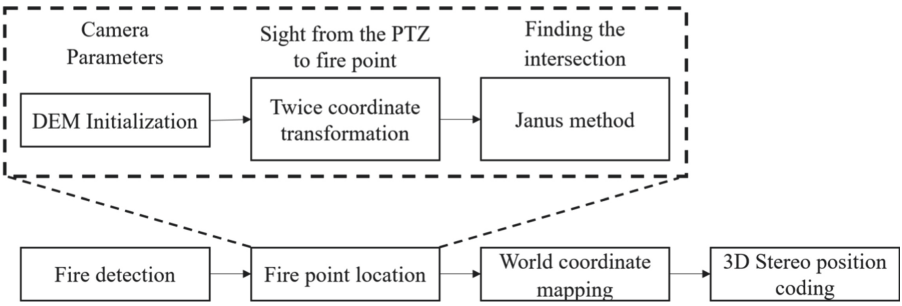

Fig. 1. Schematic flow diagram of the 3D stereo fire position coding from video frames.

As can be seen from Fig. 1, the first step of forest fire location is to detect the early forest fire from continuous frames. In this stage, we use deep learning algorithm to recognize the sequence image and locate the fire source point. Usually, the target detection algorithm will detect a rough target box. We adopt the method in [13] to detect the abnormal smoke target and accurately locate the fire pixel. In Fig. 2, the green box marks the smoke area of the forest fire, and the red dot marks the source area of the fire.

Fig. 2. Forest fire detection and fire smoke source pixel prediction. (Color figure online)

Images are composed of pixels, and pixel coordinates are the position of pixels in the image. To determine the coordinates of the pixels, we must first determine the image coordinate system. Common coordinate systems include image coordinate system, camera coordinate system, world coordinate system, etc.

In order to further locate the fire point in the world coordinates, we need to carry out coordinate transformation for fire source pixel. As shown in the Fig. 3, the image coordinate system takes the image center as the origin, where the pixels $M(x, y)$ represent the coordinates of point M in the image coordinate system, and W and H represent the width and height of the image. Camera model is an important part of visual positioning model, which is based on camera coordinate system, whose origin is located in the center of light of camera imaging lens. Camera coordinate system origin O_c is defined as the center of light of the camera lens, X_c and Y_c axis are parallel to the x- and y-axis of the image coordinate axis respectively, oO_c is the optical axis of the camera. The world coordinate system can be converted to the camera coordinate system by rigid body transformation. As shown in the figure, after the fire pixels are located by the fire source estimation algorithm, the world coordinates can be located in the following way.

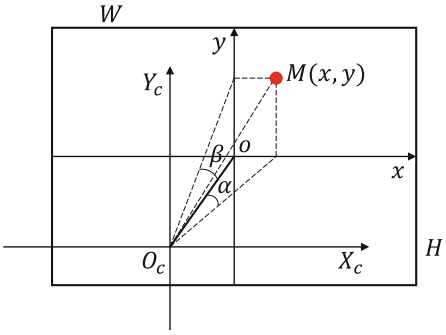

Fig. 3. Schematic diagram for the camera coordinate system.

In this paper, the digital terrain data is based on digital elevation model (DEM), which is a kind of solid terrain model that uses a group of ordered numerical array to represent the ground elevation. It is the digital simulation of the ground terrain (digital expression of the terrain surface shape) through the limited terrain elevation data. Usually in a certain area, dense terrain model points (X, Y, Z) are used to express the ground shape, which is widely used in terrain analysis and becomes an important part of spatial data infrastructure. The digital elevation model reflects the terrain changes in the monitoring area.

This study takes the map data and attribute data of Laoshan area in Nanjing as an example. The map data includes administrative division map and topographic map, and the attribute data includes forest farm name, forest class number and small class number. Figure 4 shows the DEM data visualization of Laoshan area in Nanjing. The red region in the picture indicates the place with higher altitude, and the green region indicates the place with lower altitude.

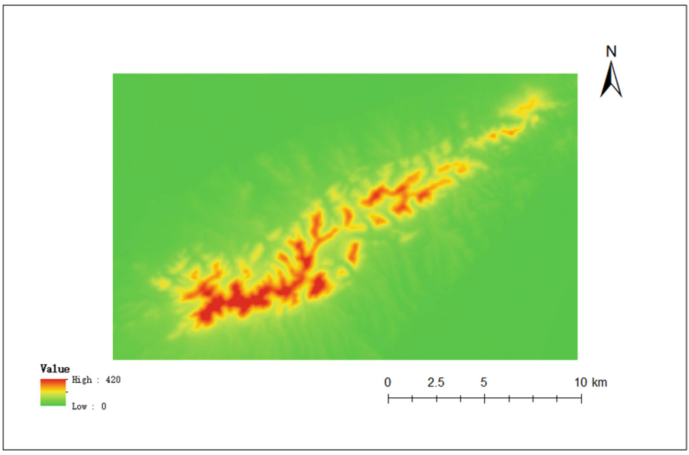

Fig. 4. Visualization of digital elevation model in Laoshan area of Nanjin. (Color figure online)

It is a challenging task to accurately locate the fire point in the world coordinate system from the pixel coordinate system. Not only we need to determine the direction of the fire, but also need to determine the distance from the fire to the camera. Fortunately, we can use the height, direction and angle information of PTZ camera, combined with DEM data to achieve this goal.

Janus method is a commonly used visibility analysis algorithm of grid digital elevation model. Its basic idea is to divide the line of sight into $step = int(max\Delta/m)$ parts by using the maximum movement of the coordinates between the observation point and the target point量 $max\Delta$ and the resolution m of the digital elevation model, and calculate the elevation difference between the terrain elevation value of the dividing point and the corresponding line of sight point ΔL: if $\Delta L > 0$, then two-point intervisibility, and make the next judgment; Otherwise, there is no intervisibility between the two points. In the following content, we describe in detail how to locate the forest fire location.

Firstly, the DEM Data of the monitoring area is initialized, and the longitude and latitude coordinates $P(lng, lat)$ and altitude h_c of the camera are located. The pitch angle Ψ and yaw angle Θ of the camera at the current moment is obtained by the feedback information of the digital PTZ camera; Fig. 5 shows a top view of fire location equipment. In this study, the monitoring PTZ camera will be manually calibrated to ensure the horizontal installation, and the position information (longitude and latitude) of forest fire monitoring point is measured. The front-end image processing equipment has the authority to control the PTZ camera, and can obtain the camera angle at any time.

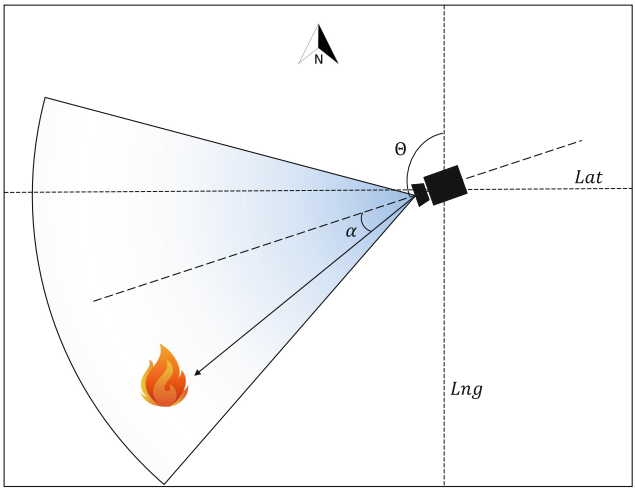

Fig. 5. Top view for the fire location process.

Secondly, computer vision algorithms can process video images in real time and detect forest fire targets in the images. We use the FFENet method [13] to accurately locate the fire pixels in the video. Based on the detection results of the forest fire recognition algorithm and the lens magnification and other parameters, the horizontal angle α and vertical angle β of the fire point relative to the camera's optical axis in the image coordinate system can be calculated. In Fig. 3, O_c is the origin of the camera coordinate system, o is the origin of the image coordinate system, and the image size is $W \times H$. For a surveillance camera with a horizontal field of view of Q_W and a vertical field of view of Q_{H}. If a fire is detected at $M(x, y)$ in the image, the two angles corresponding to the fire can be calculated by the following formula.

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 2Q_W/W & 0 \\ 0 & 2Q_H/H \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{1}$$

Obtain the horizontal angle α and vertical angle β between the line of sight l from the camera to the fire point and the central axis of the camera from the image coordinate system. Because the pitch and yaw angles of the camera can be directly controlled by the digital gimbal, the pitch and yaw angles can be directly mapped to the angle offset of the digital gimbal. In this way, the direction of the fire point relative to the monitoring camera can be obtained.

Finally, we need to calculate the distance between the fire point and the monitoring camera. The line of sight is generated by increasing the offset on the existing yaw and pitch angle according to the geographic location and spatial height of the PTZ. Janus algorithm is used to calculate the intersection coordinates of line of sight l on the digital elevation model, that is, the position of fire point (the position of $\Delta L \rightarrow 0$ in Fig. 6).

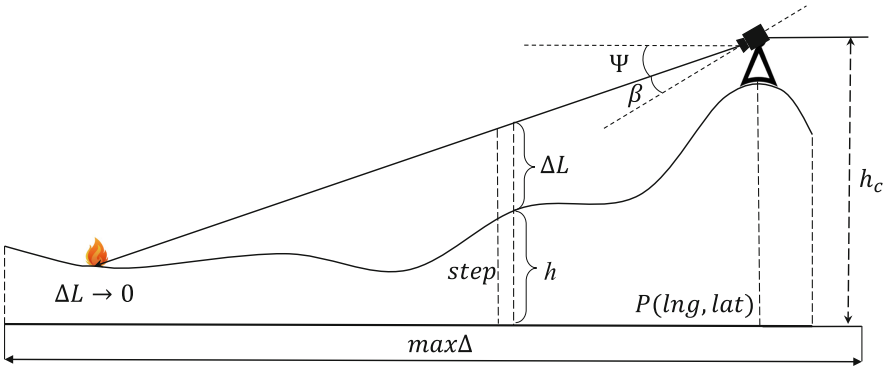

Fig. 6. Cross-sectional profile of the fire point location process.

2.2 Video Stereo Grid Construction Method

The algorithm needs some prior knowledge to complete the construction of the mapping relationship. It mainly needs to determine two parameters, one is the scale of the world coordinate system and the real distance, and the other is the real direction of the world coordinate system, which is defined as the angle between the X axis of the world coordinate system and the north direction. Scale relation needs to know the real distance of a line segment in the video scene, and direction information needs to know the orientation of the video scene in the real 3D world. Therefore, it is necessary to know the true longitude and latitude coordinates of at least two ground pixels in order to solve the two required parameters.

Firstly, the longitude and latitude coordinates of two pixels on the ground in three-dimensional space are assumed to be known. The world coordinate system is a uniform Cartesian three-dimensional coordinate system. The origin of the world coordinate system can be set as one of the two known points. Given the longitude and latitude coordinates of two points, the azimuth of one point relative to another can be calculated by the inverse solution method of geodetic theme And the real distance between two points. Azimuth is the angle from the north of a point to the target line. In this algorithm, it is the angle from the north of the origin of the world coordinate system to the line between two points.

Because the world coordinate system is not necessarily the right-handed system, we can infer whether the Y axis is 90° counterclockwise of the X axis through the camera imaging model. The judgment method is to project the unit world coordinates of the X and Y axes of the two points into the pixel coordinates $P_1 = (u_1, v_1)$, $P_2 = (u_2, v_2)$, and form the vector $\overrightarrow{P_0P_1}$ and $\overrightarrow{P_0P_2}$ with the origin pixel coordinates $P_0 = (u_0, v_0)$. Because the pixel coordinate system is not a right-handed system, the cross product value of the two vectors could be used to judge. If the cross product value is less than zero, the X axis is 90° counterclockwise of the Y axis, otherwise it is not.

At the same time, the angle between the two points and the x-axis in the world coordinate is calculated. Through the relationship between azimuth and XY axis, the angle degree between X axis and due north direction can be calculated. We set a counter

clockwise direction from X to due north. When the Y axis is 90° clockwise of the X axis, $degree = \varphi - \theta$, When the y-axis is 90° counter clockwise to the x-axis, then $degree = \varphi + \theta$. Even if there is only one known longitude and latitude point, the North vector and a known distance can be marked on the video image. It can also calculate the world coordinates of the points on the image, so as to get the required parameters.

When the degree and scale parameters are obtained, the mapping relationship can be established. Because the angle between the x-axis of the world coordinate system and the true north direction and the scale relationship with the real world are known, the projection and the direction angle of the point on the plane of the world coordinate system can be calculated.

$$\begin{cases} \theta = \arccos\left(\frac{x}{\sqrt{x^2+y^2}}\right) \\ \varphi = degree + flag * \theta(flag = -1\ if\ xy\ is\ right - handed\ system\ else\ 1) \\ d = \sqrt{x^2+y^2} * scale \\ H = z * scale \end{cases} \tag{2}$$

After the latitude and longitude coordinates of a point are obtained, the azimuth between the point and the origin of the world coordinate system can be calculated according to the latitude and longitude of this point. According to the two parameters of *degree* and *scale*, the X and Y coordinates of this point in the world coordinate system can be obtained. The Z coordinate can be obtained by dividing the height by *scale* directly.

$$\begin{cases} \theta = flag * (degree - \varphi)(flag = -1\ if\ xy\ is\ right - handed\ system\ else\ 1) \\ x = cos\theta * d/scale \\ y = sin\theta * d/scale \\ z = H/scale \end{cases} \tag{3}$$

The above method establishes the mapping from longitude and latitude to the world coordinate system. The following will introduce how to transform the longitude, latitude and height coordinates into stereo grid code, so as to achieve the construction of three-dimensional grid space. Given that the latitude and longitude coordinates of a certain position are (Lng, Lat, H), the formulas for calculating the N-level stereo grid position code is are follows:

$$CodeLng_n = \begin{cases} \left\lfloor \frac{Lng+256}{2^{9-n}} \right\rfloor_2 & 0 \le n \le 9 \\ CodeLng_9\left(\frac{64}{2^{15-n}}\right) + \left\lfloor (Lng + 256 - CodeLng_9)\frac{60}{2^{15-n}}{}_2 \right\rfloor & 10 \le n \le 15 \\ CodeLng_{15}\left(\frac{64}{2^{21-n}}\right) + \left\lfloor (Lng + 256 - CodeLng_{15})\frac{60}{2^{21-n}} \right\rfloor_2 & 16 \le n \le 32 \end{cases} \tag{4}$$

$$CodeLat_n = \begin{cases} \left\lfloor \frac{Lat+256}{2^{9-n}} \right\rfloor_2 & 0 \le n \le 9 \\ CodeLat_9\left(\frac{64}{2^{15-n}}\right) + \left\lfloor (Lat + 256 - CodeLat_9)\frac{60}{2^{15-n}} \right\rfloor_2 & 10 \le n \le 15 \\ CodeLat_{15}\left(\frac{64}{2^{21-n}}\right) + \left\lfloor (Lat + 256 - CodeLat_{15})\frac{60}{2^{21-n}} \right\rfloor_2 & 16 \le n \le 32 \end{cases} \tag{5}$$

$$CodeH_n = \left\lfloor H \times \frac{2^n}{512 \times 11130} \right\rfloor_2 \quad 0 \le n \le 32 \tag{6}$$

Based on the above formulas, the latitude longitude and height can be converted into stereo space coding, so as to form the mapping path from pixel coordinates to stereo space coding, and construct the spatial geometric model of video stereo grid.

3 Conclusion

In this paper, we propose a video stereo location method for early forest fire. Specifically, the proposed approach can be summarized as two parts: a) computer vision based method is used to locate initial forest fire pixel in image coordinate system, and then transform it to the camera coordinate system and the world coordinate system by coordinate transformation; b) we establish the mapping from longitude, latitude and height coordinates into stereo grid code to achieve the construction of spatial geometric model of video stereo grid. The application in the project confirmed the effectiveness of our proposed method.

References

1. Li, X., Chen, Z., Wu, Q.M.J., Liu, C.: 3D parallel fully convolutional networks for real-time video wildfire smoke detection. IEEE Trans. Circuits Syst. Video Technol. **30**(1), 89–103 (2020). https://doi.org/10.1109/TCSVT.2018.2889193
2. Yuan, F., Zhang, L., Xia, X., Wan, B., Huang, Q., Li, X.: Deep smoke segmentation. Neurocomputing **357**, 248–260 (2019). https://doi.org/10.1016/j.neucom.2019.05.011
3. Lippman, A.: Movie-maps: an application of the optical videodisc to computer graphics. ACM SIGGRAPH Comput. Graph. **14**(3), 32–42 (1980). https://doi.org/10.1145/965105.807465
4. Goodchild, M.F.: Geographical information science. Int. J. Geogr. Inf. Syst. **6**(1), 31–45 (1997)
5. Navarrete, T., Blat, J.: VideoGIS: segmenting and indexing video based on geographic information (2011)
6. Kim, S.H., Arslan Ay, S., Zimmermann, R.: Design and implementation of geo-tagged video search framework. J. Vis. Commun. Image Represent. **21**(8), 773–786 (2010)
7. Lewis, P., Fotheringham, S., Winstanley, A.: Spatial video and GISUJ. Int. J. Geogr. Inf. Sci. **25**(5), 697–716 (2011)
8. Lu, Y., Shahabi, C., Kim, S.H.: An efficient index structure for large-scale geotagged video databases. In: The 22nd ACM SIGSPATIAL International Conference. ACM (2014)
9. Goodchild, M.F., Shiren, Y.: A hierarchical spatial data structure for global geographic information systems. In: CVGIP: Graphical Models and Image Processing, vol. 54, No. 1, pp. 31–44 (1992)
10. Goodchild, M.F.: Discrete global grids: retrospect and prospect. Geogr. Geo-Inf. Sci. **28**(1), 1–6 (2012)
11. Goodchild, M.F.: Geographical grid models for environmental monitoring and analysis across the globe (panel session) (1994)
12. Tobler, W., Chen, Z.: A quadtree for global information storage. Geogr. Anal. **18**(4), 360–371 (1986)
13. Cao, Y., et al.: EFFNet: enhanced feature foreground network for video smoke source prediction and detection. IEEE Trans. Circuits Syst. Video Technol. (2021)