Information Optimization for Image Screening and Transmission in Aerial Detection

Jieqi Li[1(✉)], Xi Liu[2], Lefan Wang[3], Fu Kong[4], Zhengxue Li[1], and Geling Yin[1]

[1] China Academy of Lunch Vehicle Technology, Beijing 10076, China
sxtyljq@163.com

[2] Unit 95899 of PLA, Beijing 100085, China

[3] School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710129, Shaanxi, China

[4] Beijing Mechanical and Electrical Overall Design Department, Beijing 100854, China

Abstract. In aerial detection, photoelectric sensor as the main detection form, can usually obtain a large number of image data in the detection target area, however, the communication bandwidth is often limited. As a result, the contradiction between powerful image acquisition and limited bandwidth causes the massive detected images cannot be completely transmitted. To address this problem, an effective method of acquisition information optimization for image screening and transmission in aerial detection is proposed in this paper. This proposed method is mainly based on the principle of sparse coding, in which the key information is extracted and can reconstruct the other information well via a linear combination. It can autonomously select and transmit the most valuable images without the requirement of prior information. As a result, the transmission of redundant information is greatly reduced, and the requirement for communication bandwidth of detection system is also reduced.

Keywords: Information optimization · Image screening · Aerial detection

1 Introduction

Aerial detection uses aircrafts to conduct scientific detection of areas that are not easy for humans to reach to obtain information about the detection target area. It plays an important role in military, civilian and other fields. In modern war, information dominance is the key factor affecting the overall strategic situation. The requirement of intelligence information in combat is higher and higher, and reliable intelligence information has become the fundamental guarantee of precise strike. In the civil aspect, aerial detection also plays an important role. For example, the aerial detection technology combined with remote sensing has been widely used in basic geological survey, mineral resources exploration, environmental geological exploration, geological disasters and other fields, and has broad development prospects.

With the continuous development of photoelectric sensors, the existing detection technology is developing towards the direction of multi/hyperspectral, high-resolution and all-weather detection. It can obtain a large number of detection images in the target

B. Li et al. (Eds.): IoTaaS 2020, LNICST 346, pp. 745–754, 2021.
https://doi.org/10.1007/978-3-030-67514-1_59

area, however, the bandwidth of the downlink channel is usually limited for such large volume of image data. It has been an increasingly significant problem for the detection system to screen the detected massive images to realize the information optimization and to transmit the most effective information back to the ground information center for response.

At present, there are few researches on the screening and transmission of aerial detection images. The simplest approach is uniformly sampling and transmission, however, the content of images is ignored. Hu *et al.* proposed a SAR image filtering method based on bit plane bit-plane characteristics [4], which is for the typical target of interest and requires the prior knowledge of targets. In addition, image deduplication technology [7, 10], that is to delete duplicate images, can also be thought as an information optimization technology, but it has not been applied to the field of image screening and transmission of detection system.

Based on the keyframe extraction technology in the field of video summarization [1, 3], an acquisition information optimization method is proposed for image screening transmission in aerial detection is proposed. The method does not aim at specific targets and does not need prior knowledge. Through the feature representation of the aerial detection image, the most informative image is selected for transmission according to the characteristics of the image. In each subsequent transmission, the image which is mostly different with the transmitted image is selected. Through the proposed information optimization method, the transmitted data amount and the captured information redundancy is significantly reduced, and the effectiveness of information under the limited bandwidth is improved.

2 Image Screening and Transmission System

The aerial detection image screening and transmission system is shown in Fig. 1. Firstly, the detection sensor detects the target area and collects a large number of aerial detection images. Next, the optimization of image screening and transmission is performed on the detected image sequences, and the most valuable images can be screened from all images. Finally, the screened image is transmitted and returned to the ground information center through the communication system. In this process, the most important part is the optimization of image screening and transmission of massive detection images. Furthermore, the major difficulty is the criterion of image screening, that is, how to choose a small number of key images from the massive detection images without specific targets and prior knowledge.

This proposed method in this paper selects the key images need to be transmitted from the massive detection images based on the principle of sparse representation. The main idea of sparse representation is to select a small number of atoms from an over-complete dictionary to form a sub-dictionary, which is required to reconstruct the original signal with a extremely small error. As well as for the screening and transmission of detection images, a small number of images are selected from the mass detection images to form a key image set. If this key image set can reconstruct the mass detection images without losing the important information carried by the original mass detection images, such a key image set is our need.

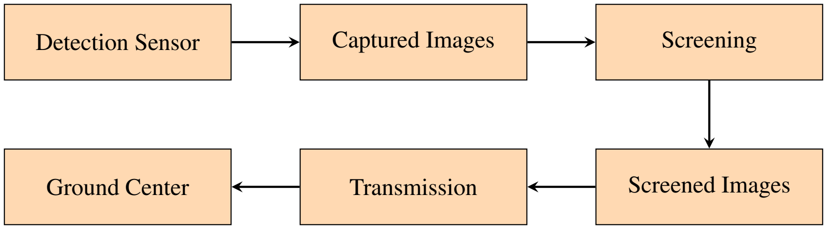

Fig. 1. Aerial image screening and transmission system.

3 Autonomous Screening and Transmission Algorithm

3.1 Screening and Transmission Model

Let $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n] \in \mathbb{R}^{d \times n}$ denote the massive detection image sequence acquired by the aerospace detection system, where $\mathbf{f}_i (i = 1, 2, \cdots, n)$ represents the feature vector of the i-th detection image. Based on the principle of sparse reconstruction, the key image $\mathbf{F}_K = \left[\mathbf{f}_{k_1}, \mathbf{f}_{k_2}, \ldots, \mathbf{f}_{k_p}\right]$ that best covers all of its information is selected from the original detection image sequence for transmission. Once the key image set is selected, so that the original detection image sequence $\mathbf{F}$ not only can be accurately reconstructed but also contains as small as possible columns. Therefore, the aerial detection image screening and transmission model we established is shown in Eq. (1):

$$
\begin{aligned}
&\min_{\mathbf{S}} : \frac{1}{2} \|\mathbf{F} - \mathbf{F}_K \mathbf{A}\|_F + \lambda \cdot \sum(\mathbf{S}), \\
&\text{s.t.} \quad \mathbf{F}_K = \mathbf{FS}, \\
&\mathbf{A} = f(\mathbf{F}, \mathbf{F}_K),
\end{aligned} \tag{1}
$$

in which $\mathbf{S}$ is the key image selection matrix, its diagonal elements are '0' or '1', the other elements are all '0', and '1' means that the corresponding detection image is selected as the key image for transmission, vice versa. In addition, $\mathbf{A}$ represents the reconstruction coefficient when key image set $\mathbf{F}_K$ is used to reconstruct the original image sequence $\mathbf{F}$, and $\sum \mathbf{S}$ represents the number of key frames $\mathbf{F}_K$, namely, the number of transmitted detection images.

In the objective function of Eq. (1), the first term is the error when the original massive detection image sequence is reconstructed by the extracted key image set, and the second term is as small as possible to constrain the extracted key images. In order to acquire the reconstruction coefficients $\mathbf{A}$, we use Orthogonal Subspace Projection (OSP), which projects all the detection images to the subspace determined by the key image set. Therefore, the reconstruction coefficient is obtained as follows:

$$
\mathbf{A} = \left(\mathbf{F}_K^T \mathbf{F}_K\right)^{-1} \mathbf{F}_K^T \mathbf{F}. \tag{2}
$$

3.2 Model Optimization

To solve the model defined in Eq. (1), the key issue is to determine the key image selection matrix $\mathbf{S}$. After obtaining $\mathbf{S}$, the reconstruction coefficient $\mathbf{A}$ can be calculated by Eq. (2), then the reconstruction error can be obtained to determine whether the selected key image set is sufficient. The determination of $\mathbf{S}$ is the determination of the key images to be screened, including the first key image and subsequent key images.

Determine the First Key Image. The first key image is always the one whose feature vector is nearest from the average or the one who has maximum amplitude.

$$\mathbf{f}_{k_1} = \arg\min_{\mathbf{f}_j \in \mathbf{F}} \left\| \mathbf{f}_j - \bar{\mathbf{f}} \right\|_2, \tag{3}$$

where $\bar{\mathbf{f}} = \frac{1}{n}\sum_j \mathbf{f}_j$ means the average of all image feature vectors, or

$$\mathbf{f}_{k_1} = \arg\max_{\mathbf{f}_j \in \mathbf{F}} \left\| \mathbf{f}_j \right\|_2. \tag{4}$$

Equation (3) indicates that the feature of the first key image is closest to the average feature of all detected images, and Eq. (4) means the first key image has maximum amplitude.

Determine the Subsequent Key Images. When the first key image is determined, the image that has the most differences with the transmitted ones is selected as the next key image. In other words, we use the currently transmitted detection image set to reconstruct all non-transmitted detection images, and select the image with the largest reconstruction error to transmit. This is due to the larger the reconstruction error is, the less information of the image is contained by the transmitted image set.

Assuming $\mathbf{F}_K = [\mathbf{f}_{k_1}, \mathbf{f}_{k_2}, \ldots, \mathbf{f}_{k_m}] \in \mathbb{R}^{d \times m}$ denotes detection images that have been transmitted, the next detection image to be transmitted is:

$$\mathbf{f}_{k_{m+1}} = \arg\max_{\mathbf{f}_j \in \mathbf{F}/\mathbf{F}_K} \frac{\left\| \mathbf{f}_j - \mathbf{F}_K \mathbf{a}_j \right\|}{\left\| \mathbf{f}_j \right\|_2}, \tag{5}$$

in which $\mathbf{a}_j$ represents the reconstruction coefficient of the j-th detection image, which can generally be obtained by orthogonal projection algorithm:

$$\mathbf{a}_j = \left(\mathbf{F}_K^T \mathbf{F}_K \right)^{-1} \mathbf{F}_K^T \mathbf{f}_j. \tag{6}$$

After obtain $\mathbf{f}_{k_{m+1}}$, the key image set can be updated as

$$\mathbf{F}_K = \mathbf{F}_K \cup \mathbf{f}_{k_{m+1}}. \tag{7}$$

According to the above analysis, the designed image screening and transmission process is shown in Fig. 2.

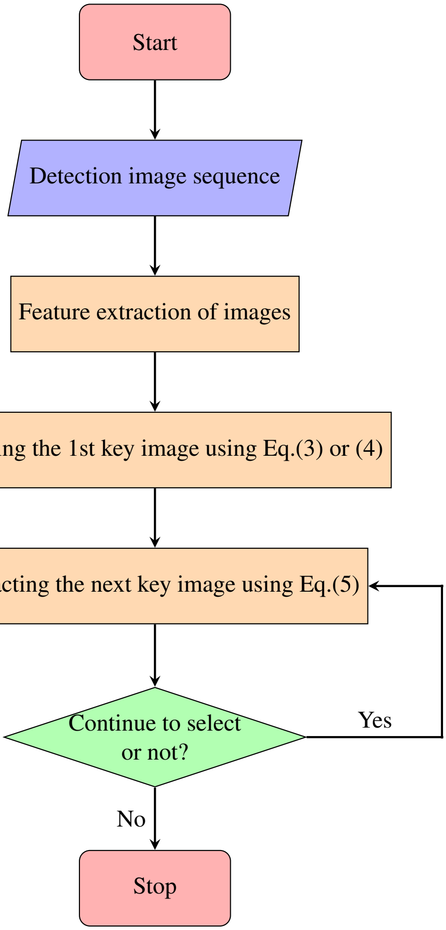

Fig. 2. The proposed aerial detection image screening and transmission process.

4 Feature Representation of Detection Images

As mentioned above, during the screening of massive detection images, how to characterize each image in the detection image sequence directly affects the final screening and transmission effect. If the processing is directly based on the pixel data of the original image, the similarity of the image is easily affected by factors such as illumination, geometric deformation, etc., and the computation is usually time-consuming. Features representation can accurately describe the information that users care about in the image, thereby, the impact of factors such as illumination on image similarity is reduced. Commonly used image features include color, texture, shape and spatial relationship.

Color is a global feature that describes the surface properties of the scene corresponding to the image or image area. The most commonly used method to express color is color histogram, that is not affected by image rotation and translation changes. However, the disadvantage is that it does not express the distribution information in color space. Commonly used color histogram feature representation methods include: RGB (red, green, blue) space histogram, HIS (hue, intensity, saturation) space histogram, HSV (chroma, purity, lightness) space histogram, etc. The color set [8] is an approximation to the color histogram. Firstly, the RGB color space is transformed into a visually balanced color space (HSV), and the color space is quantized into several bins. Then, the image is divided into several regions by automatic color segmentation technology, and each regions is indexed by a color component of the quantified color space, so the image is expressed as a binary color index table. Color moment uses the mathematical moments of color distribution to represent the image [11]. Since color distribution is mainly concentrated in low-order moments, the first, second, third-order moment (mean, variance, and skewness) are enough to express the color distribution of the image. The color coherence vector divides the pixels belonging to each bin of the histogram into two parts. If the area of the continuous region occupied by some pixels in the bin is larger than a given threshold, the pixels in the region are regarded as coherence pixels.

As a kind of global feature, texture feature describes the surface properties of the image area. Different from the color feature, the texture feature is not based on the feature of pixels, and needs to perform statistical calculation in the area containing multiple pixels. As a statistical feature, texture features usually have rotational invariance and strong resistance to noise. Common texture feature extraction methods include structural analysis methods, statistical methods, model based methods and signal processing based methods [12]. Only applicable to regular textures, the structural analysis method assumes that the texture is formed by a certain regular arrangement of texture primitives. The statistical method analyzes the texture of images from the perspective of regional statistics, which can be carried out in space and frequency domains. The main methods include edge histogram, autocorrelation function, edge frequency, gray-level co-occurrence matrix, etc. Based on the structural model of the image, the model based method take advantage of the parameters of the model as the texture feature. Typical methods are random field model methods, such as Markov Random Field (MRF) model and Gibbs Random Field model. Generally, signal analysis based methods utilize filter, such as spatial filter, frequency domain filter, and Gabor filter, to extract texture from the filtered image.

There are two types of representation methods for shape features, one is contour features, and the other is regional features. The contour feature of the image principally focused on the outer boundary of the object, while the regional feature of the image is primarily related to the entire shape area. The typical methods for shape description include boundary feature method, Fourier shape descriptor method, and moment invariant feature method. The boundary feature method obtains the shape parameters of the image by describing the boundary feature, and the representative algorithm includes the Hough transform method for detecting parallel lines and the boundary direction histogram method. The Fourier shape descriptor method uses the Fourier transform of the object boundary as the shape description, and utilizes the closure and periodicity of the region boundary to transform a two-dimension problem into a one-dimension. Invariant

(a) Uniform sampling **(b)** Ours1 **(c)** Ours2

Fig. 3. Comparisons of the selected key images by different methods.

moments exploit the moments in statistics to characterize the geometric characteristics of the image area, which have the invariance to rotation, translation, and expansion.

In order to obtain better image screening performance, the features used should own invariance to illumination, translation and rotation, so as to avoid screening out some detection images that only have simple rotation, translation or light intensity change. Therefore, combined features can be used to represent images, such as a 360 dimensional descriptor consisted of census transform histogram [9] and color moment, which has achieved satisfactory performance in the research of video summarization [2,6]. In recent years, with the development of deep learning, deep features based on convolution neural networks [5] can also be used for image feature description.

5 Experiment and Discussion

In the experiment, the processed video frames are used to simulate the detection image sequence. The video used in our experiment is from episode 10 of "Aerial China3", which is aerial documentary with 10 episodes. The video is down-sampled every five frames to reduce the computation, and the resolution is 1920 × 1080. For the feature representation of images, the experiment uses 360-dimensional features, consisting of 252-dimensional CENTRIST features and 108-dimensional color features. In the experiment, 15 key images are selected from 600 images for transmission, and the results of the algorithm proposed in this paper are compared with uniform sampling.

We first give a qualitative representation of the key images selected by our method and uniform sampling, and the comparison results are shown in Fig. 3, in which '*Ous1*' and '*Ous2*' indicates our method selects the 1st key image using Eq. (3) and Eq. (4), respectively. According to the results of Fig. 3, the similarity between the transmitted images filtered by the proposed method in this paper is obviously small, which indicates that the information content of the transmitted images filtered by our method is more sufficient and abundant. However, it can also be observed that the images selected by uniform sampling have large similarities, for example, the 2nd, 3rd, 4th, 5th and 6th images have certain similarity. This shows that uniform sampling can not consider the content changes of image sequences.

For further observation, the experiment is quantitatively evaluated, and the average correlation defined in Eq. (8) between each currently screened transmission image (image to be transmitted) and the previously transmitted image is calculated. The experimental results are shown in Fig. 4. According to the quantitative evaluation in Fig. 4, the average correlation between the newly selected image and the images that have been selected and transmitted through the proposed algorithm is significantly lower than the correlation through uniform sampling and filtering. In summary, the detection image screening and transmission algorithm proposed in this paper can not only reduce the transmission data volume, but also transmit the most informative images.

$$Cor_{avg} = \frac{1}{|kf^{(pre)}|} \sum_i cor(kf^{(now)}, kf_i^{(pre)}), \tag{8}$$

where $kf^{(now)}$ denotes the newly selected keyframe at current time, $kf_i^{(pre)}$ denotes the i-th key image that is previously selected and transmitted, $|kf^{(pre)}|$ is the length, i.e., the number of previous key images.

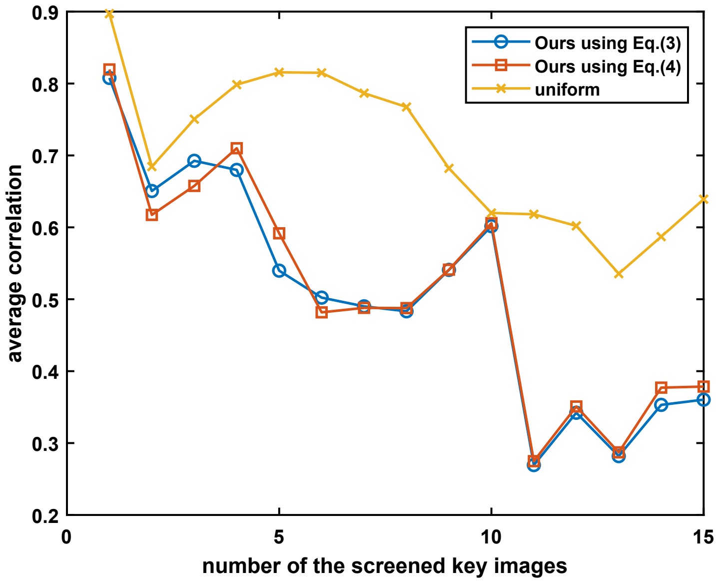

Fig. 4. Comparisons of the selected key images by different methods.

6 Conclusion

In order to effectively solve the contradiction between the massive images of the target captured by the sensor and the limited transmission bandwidth of the communication system, information optimization method for screening and transmission of detection images is proposed in this paper. The proposed method is based on sparse representation principle, which assumes than the screened key image set can reconstruct all the important information in the original image sequence. On the premise that there is no specific target and prior information, the most valuable images are selected and transmitted in the aerial detection system, so as to realize the optimization of information acquisition. Experimental results prove that the proposed method can select more informative images and reduce the transmission of redundant information. Accordingly, the requirement for communication bandwidth of the detection system is reduced under the same detection image acquisition capacity.

References

1. Ajmal, M., Ashraf, M.H., Shakir, M., Abbas, Y., Shah, F.A.: Video summarization: techniques and classification. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2012. LNCS, vol. 7594, pp. 1–13. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33564-8_1
2. Cong, Y., Liu, J., Sun, G., You, Q., Li, Y., Luo, J.: Adaptive greedy dictionary selection for web media summarization. IEEE Trans. Image Process. **26**(1), 185–195 (2017)

3. De Avila, S.E.F., Lopes, A.P.B., da Luz, A., de Albuquerque Araújo, A.: VSUMM: a mechanism designed to produce static video summaries and a novel evaluation method. Pattern Recognit. Lett. **32**(1), pp. 56–68 (2011)
4. Hu, C., Liu, F., Zhou, J.: SAR images screening based on bit-plane characteristics. J. Comput. Appl. **29**(11), 3021–3026 (2009)
5. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
6. Ma, M., Mei, S., Wan, S., Hou, J., Wang, Z., Feng, D.D.: Video summarization via block sparse dictionary selection. Neurocomputing **378**, 197–209 (2020)
7. Rashid, F., Miri, A., Woungang, I.: Secure image deduplication through image compression. J. Inf. Secur. Appl. **27**, 54–64 (2016)
8. Smith, J.R., Chang, S.F.: VisualSEEk: a fully automated content-based image query system. In: ACM International Conference on Multimedia, pp. 87–98 (1997)
9. Wu, J., Rehg, J.M.: CENTRIST: a visual descriptor for scene categorization. IEEE Trans. Pattern Anal. Mach. Intell. **33**(8), 1489–1501 (2010)
10. Xu, J., Zhang, W., Zhang, Z., Wang, T., Huang, T.: Clustering-based acceleration for virtual machine image deduplication in the cloud environment. J. Syst. Softw. **121**, 144–156 (2016)
11. Yu, H., Li, M., Zhang, H.J., Feng, J.: Color texture moments for content-based image retrieval. In: International Conference on Image Processing, pp. 929–932 (2002)
12. Zhang, J., Tan, T.: Brief review of invariant texture analysis methods. Pattern Recognit. **35**(3), 735–747 (2002)