Key Frames Extraction Based on Optical-flow and Mutual Information Entropy

Guangbin Bao, Dongfang Li* and Yuli Mei

School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China
*Corresponding author: 308165685@qq.com

Abstract. Key frames provide a suitable video summary and framework for video indexing, browsing and retrieval, and how to extract key frames is the core in the field of video retrieval. In order to solve the problem of redundancy and missing selection in key frame extraction, this paper proposed a key frame extraction algorithm based on optical flow and mutual information entropy. This algorithm integrates mutual information entropy and optical flow characteristics to extract key frames. First, the optical flow method calculated the total optical flow of each frame of the image, and selected the video frame with extreme optical flow difference in the neighbourhood as the key frame, and the other video frames were put into the candidate key frame set; by calculating the mutual information entropy of the key frame set, the minimum mutual information entropy is taken as the threshold; Then, The mutual information entropy of the candidate key frame set was calculated, and the image frames larger than the threshold value were put into the key frame set; Finally, redundant key frames are deleted through the measurement of inter-frame similarity, and the remaining key frames are the key frames to be extracted. The experimental results show that compared with other methods, the accuracy of extracted key frames by this approach is significantly improved on precision, recall and F-score, and the extracted key frames can reflect video content more effectively.

1. Introduction

With the rapid development of multimedia technology and video monitoring technology, as well as the popularity of mobile phones, tablet PC and various digital devices, the number of videos is growing rapidly, and various good and bad videos are flooding in the network. How to analyse and filter the video data quickly and effectively is one of the hot issues in the current research, especially the key frame extraction. Effective extraction of key frames and generation of video summaries are conducive to improving the efficiency of detecting bad video information, filtering bad video, purifying network space, and providing users, especially teenagers, with a green and safe online environment.

In order to extract key frames effectively, researchers proposed a lot of methods to extract key frames. For example, SUN[1] proposed an approach by combining mutual information (MI) and image entropy to extract keyframes, but the effect of keyframe processing between shots is not ideal. Brindha [2] proposed a triangle model method to extract keyframes, but the selected key frame cannot represent the information of the original input video. W. Wolf [3] proposed an optical flow analysis approach to calculate the amount of use in the lens, and selected the local minimum of the amount of motion as the key frame, but the calculation amount of this method is large. Li D[4] proposed a new method to extract key frames, but there was a key frame redundancy problem in extracting key frames. Jiahao[5] proposed a method, which can also extract key frames effectively, but the process is a bit cumbersome.
In view of the above problems, this paper proposes a key frame extraction algorithm based on optical flow and mutual information entropy (OMIE).

2. Proposed Method

2.1. Outline of the Proposed Algorithm

Our method consists of two steps. In the first step, according to the difference of optical flow, the video frame is divided into video key frames set (VKS) and candidate key frames set (CKS). The mutual information entropy $I_{vks}(f_k, f_{k+1})$ of two adjacent frames in VKS is calculated. The similarity of two frames is measured by $I_{vks}(f_k, f_{k+1})$, and the similarity threshold $T$ is set to make $T = \min I_{vks}$. In the second step, calculate the mutual information entropy in the candidate key frame and compare it with the threshold value. If it is greater than the threshold value $T$, put it into the key frames set (KS); otherwise, measure the similarity between two frames. If it is less than the threshold value, it is determined as the key frame.

2.2. Key Frame Information Extraction Based on Optical Flow

2.2.1 Optical flow method

Optical flow was first proposed by Gibson in 1950. In optical flow, the structure and motion relationship of the object in the video can be obtained by observing the instantaneous velocity of the pixel motion on the image. Assuming that the brightness is constant, the time is continuous or the range of motion is not violent, the following formula 1 can be obtained after $dt$ time.

$$I(x, y, t) = I(x + dx, y + dy, t + dt)$$  \hspace{1cm} (1)

By expanding the right Taylor of the formula 1, we can get formula 2.

$$I(x, y, t) = I(x, y, t) + \frac{\partial I}{\partial x} dx + \frac{\partial I}{\partial y} dy + \frac{\partial I}{\partial t} dt + \varepsilon$$ \hspace{1cm} (2)

Then the equations (1) and (2) above can be used to obtain the optical flow constraint equation.

$$I_x \cdot u_x + I_y \cdot u_y + I_t = 0$$ \hspace{1cm} (3)

Where, $I_x$ and $I_y$ represent the change of light intensity of adjacent pixels in the same image in the X and Y directions respectively. $I_t$ represents the intensity change of the same pixel at adjacent moments. $\overline{u} = (u_x, u_y)^T$ is the optical flow at that moment.

2.2.2 Key frame information extraction

Because optical flow can effectively reflect the motion information of moving objects in video, this paper proposes a method to obtain video key frames by calculating the optical flow difference of picture frames. The steps are as follows.

Step 1: The total optical flow information $M(t)$ of each frame in the video is calculated.

$$M(t) = \sum_x \sum_y \left( (u(x, y, t))^2 + (v(x, y, t))^2 \right)^{\frac{1}{2}}, \quad t = 1, 2, \cdots, n$$ \hspace{1cm} (4)

Where $u(x, y, t)$ is the motion rate of gray information in the X direction at time $t$; $v(x, y, t)$ is the motion rate of gray information in the Y direction at time $t$. Figure 1 (a) shows the optical flow information of each frame in the video.

Step 2: Calculate the optical flow difference $D(t)$ between two adjacent frames.

$$D(t) = M(t) - M(t - 1)$$ \hspace{1cm} (5)

In order to reflect the change of D (T) more intuitively, the D(t) is magnified 100 times in practice and connected into a curve. The position of the minimum value in Figure 1 (b) is the place where the moving optical flow has a sudden change. Then the video frame with the minimum value can be regarded as the key frame.
Figure 1. (a) shows the optical flow information of the video; (b) shows the optical flow difference of the video.

Step 3: According to the results of optical flow difference extraction, calculate argmin D(t). Select the first n argmin D(t) according to the number of keyframes given in the video, find the keyframes corresponding to the minimum D(t) in the image, put them into the VKS, and put the others into the CKS.

2.3. Key Frame Extraction Based on Mutual Information Entropy

2.3.1. Entropy and mutual information entropy

In video, because the information contained in a single video frame is uncertain, entropy and information entropy[6-7] are introduced to describe the uncertainty of information. Assuming that the distribution of image's pixels is independent and uncorrelated, the probability distribution of image pixels is expressed as \( p = \{p_1, p_2, \cdots, p_n\} \). For two images X and Y, the information entropy and joint information entropy can be defined as shown in formula.

\[
H(X) = -\sum_i p_X(i) \log p_X(i) \quad (6)
\]

\[
H(Y) = -\sum_j p_Y(j) \log p_Y(j) \quad (7)
\]

\[
H(X,Y) = -\sum_{i,j} p_{X,Y}(i,j) \log p_{X,Y}(i,j) \quad (8)
\]

Where \( p_X(i) \) and \( p_Y(j) \) represent respectively the probability density function of image X and image Y. \( p_{X,Y}(i,j) \) is the joint probability density function of image X and Y.

The mutual information between X and Y can be expressed as \( I(X,Y) \).

\[
I(X,Y) = H(X) + H(Y) - H(X,Y) \quad (9)
\]

2.3.2. Key frame extraction

Because mutual information entropy can effectively measure the similarity of pictures, this paper calculates the mutual information entropy to determine whether the candidate key frame is a key frame. Its main steps are as follows.

Step 1. Calculate the mutual information entropy feature \( I_{vks}(f_k, f_{k+1}) \) in the video key frame set and set the threshold T.

\[
I_{vks}(f_k, f_{k+1}) = H(f_k) + H(f_{k+1}) - H(f_k, f_{k+1}) \quad 1 \leq k < n \quad (10)
\]

\[
T = \min I_{vks}(f_k, f_{k+1}) \quad (11)
\]

Step 2. Calculating the mutual information entropy feature \( I_{cks}(f_j, f_{j+1}) \) in the CKS. If \( I_{cks}(f_j, f_{j+1}) \) is greater than threshold T, the frame is placed in the VKS. If the \( I_{cks} \) is less than the threshold T, then the feature vector similarity measurement is carried out for this frame and the key
frames in VKS. Here, the feature vector similarity measurement adopts the OOS\[8\] algorithm and set the value of similarity \(S\) to 10\%. If the \(\text{OOS}(f_{j+1}, f_k)\) is less than \(S\), the two frames are considered to be similar and put into KS; if the OOS is greater than \(s\), the two frames are considered to be different and put into CKS.

Step 3. According to the calculation results, the video frames that meet the requirements are saved as key frames, and the results are output. Algorithm 1 shows the key frame extraction process.

3. Experimental Result Analysis

Key frame extraction evaluation criteria are generally divided into three categories\[9\]: result description, objective evaluation and subjective evaluation. In this paper, objective evaluation and subjective evaluation are used to evaluate the algorithm model.

3.1 Objective Evaluation Results and Analysis

In order to verify the performance of video key frame extraction algorithm in this paper, objective evaluation criteria are used. The formula for precision, recall and F-score is as follows.

\[
\text{precision} = \frac{N_{\text{matched}}}{N_{SF}}
\]

(12)

\[
\text{recall} = \frac{N_{\text{matched}}}{N_{UF}}
\]

(13)

\[
\text{F-score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{(\text{precision} + \text{recall})}
\]

(14)

Where \(N_{\text{matched}}\) represents the correct number of key frames extracted by the algorithm; \(N_{SF}\) represents the number of key frames provided by the system; \(N_{UF}\) represents all key frames extracted by the algorithm; when the precision and recall are relatively high, the F-score will be higher.

It can be seen from Table 1 that the F-score of the proposed OMIE algorithm is the highest among all algorithms, respectively 12\%, 23\%, 16\%, 14\%, higher than OV, DT, STIMO, KMC algorithm, indicating that the overall performance of the OMIE algorithm in this paper is higher than other algorithms. we also found that the recall of the proposed OMIE algorithm are the highest among all algorithms. Respectively 7\%, 28\%, 9\%, 17\%, higher than other methods, and the precision of proposed OMIE method is higher than other methods, respectively 16\%, 17\%, 22\%, 10\%.

Table 1. The precision, recall and F-score of different algorithms on Open Video Project

| Algorithm | OV[10] | DT[11] | STIMO[12] | KMC[9] | OMIE  |
|-----------|--------|--------|-----------|--------|-------|
| Precision | 0.62   | 0.61   | 0.56      | 0.68   | 0.78  |
| Recall    | 0.69   | 0.48   | 0.67      | 0.59   | 0.76  |
| F-score   | 0.65   | 0.54   | 0.61      | 0.63   | 0.77  |
| Average length | 9.8 | 6.3 | 10.1 | 6.93 | 9.32 |

As can be seen from Figure 2, the number of correct key frames extracted by our algorithm is the most in all algorithms, which shows that the performance of the OMIE algorithm in this paper is higher than other algorithms.
3.2 Subjective Evaluation Results and Analysis

The subjective evaluation adopts the key frame extraction rating method, and uses three levels of Good, Acceptable and Bad\cite{9} to evaluate the key frame extraction results. Three males and two females unrelated to this research were invited as testers to subjectively evaluate the results of the key frames extracted by the above 5 algorithms.

|          | User1 | User2 | User3 | User4 | User5 | Average Value |
|----------|-------|-------|-------|-------|-------|---------------|
| **Good** | 43 (86%) | 44 (88%) | 42 (84%) | 44 (88%) | 45 (90%) | 43.6 (87.2%) |
| **Acceptable** | 6 (12%) | 6 (12%) | 7 (14%) | 6 (12%) | 5 (10%) | 6 (12%) |
| **Bad** | 1 (2%) | 0 (0%) | 1 (2%) | 0 (0%) | 0 (0%) | 0.4 (0.8%) |
| **Total** | 50 | 50 | 50 | 50 | 50 | 50 |

From table 2, it can be seen that the percentage of grade good accounts for 87.2%, indicating that the key frames extracted by OMIE algorithm can well present video content. The percentage of the grade Acceptable accounts for 12%, indicating that the OMIE algorithm can better present the video content for the key frames extracted by some videos. The percentage of grade bad is only 0.8%, which shows that the key frames extracted by OMIE algorithm can't present the video content of a few videos.

In order to further verify the effectiveness of the model in this paper, 10 test videos of different types are randomly selected from the OVP \cite{13} and the average playing time of each video is about 2-5 minutes. The same testers are invited to evaluate the key frame extraction results of the above five algorithms.

From Figure 3, it can be seen that the evaluation of the OMIE model in this paper is higher than other algorithms. Specifically, our model has the highest subjective evaluation score for the key frame extraction results of 8 videos in 10 videos, and the average subjective evaluation score is higher than other comparison algorithms. This shows that our algorithm has achieved a better user experience.
In addition, it analyzes the reliability of subjective result statistics, judges whether the score includes the serious subjective deviation of users, and analyzes the user preference[9]. As can be seen from Figure 4 that various users have similar preferences in these methods, indicating that the subjective evaluation results are reliable.

**Figure 3.** Subjective evaluation results of test videos

**Figure 4.** User preference statistics

### 4. Conclusion

This paper proposes a key frame extraction algorithm based on optical flow and mutual information entropy (OMIE). By comparing the subjective and objective experiments on OVP data sets and the mainstream key frame extraction algorithms, the advancedness of the algorithm in this paper is verified. The experimental results show that the method has the following advantages: (1) effectively improve the accuracy of shot detection; (2) effectively solve the problem of redundant frame and missing frame; (3) the key frame can better reflect the shot content, and can extract more key frames under the same frame number.

### 5. Acknowledgements

This work was supported by Gansu Province Natural Science Foundation of China Nos.18JR3RA156 and Lanzhou Municipal Science and Technology Program Nos.2017-4-105.

### 6. References

[1] SUN L N and ZHOU Y H. A key frame extraction method based on mutual information and
image entropy. Hangzhou: International Conference on Multimedia Technology, 2011, 35-38.

[2] Brindha N and Visalakshi P. Content based video feature extraction and classification using perceived motion energy spectrum-svm classifier. International Journal of Computing & Information Technology, 2018, 10(1), 1-10.

[3] Wlof, W. Key frame selection by motion analysis. 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings. IEEE, 2002, 1228-1231.

[4] Li D, Jizhong S, and Penghui G. Detection and judgment of crowd emergencies based on optical flow and information entropy. Journal of Zhejiang University (Science Edition), 2019, 46(1):39-47.

[5] Jiahao Zhao and Bo Q. Key frame extraction algorithm based on mi-surf features. Wireless Internet Technology, 2019, 16(03):117-119+122.

[6] Ye G, Pan C., Huang X., Zhao Z, and He J. A chaotic image encryption algorithm based on information entropy. International Journal of Bifurcation and Chaos, 2018, 28(01), 1850010.

[7] Gang SY. Performance comparison of statistics based similarity measures for image registration. Chinese Journal of Computers, 2004, 27(9), 1278-1283.

[8] ZHAO W L, NGO C W, TAN H K, et al. Near-duplicate Keyframes identification with interest point matching and pattern learning. IEEE Transactions on Multimedia, 2007(5):1037-1048.

[9] Truong B T and Venkatesh S. Video abstraction: a systematic review and classification. Transactions on Multimedia Computing, Communications and Applications, ACM, 2007, 3(1):1-37.

[10] Deng C, Ji R, Liu W, et al. Visual reranking through weakly supervised multi-graph learning. Proceedings of the IEEE International Conference on Computer Vision, 2013:2600-2607.

[11] Mundur P, Rao Y, and Yesha Y. Keyframe-based video summarization using delaunay clustering. International Journal on Digital Libraries, 2006, 6(2), 219-232.

[12] Furin M, Geraci F, Montangero M, and Pellegrini, M. Stimo: Still and moving video storyboard for the web scenario. Multimedia Tools and Applications, 2010, 46(1), 47-69.

[13] OVP. Video Open Project Storyboard [EB/OL]. 2020, https://open-video.org.