A Systematic Review of the Literature in Dynamic Video Summarization

Mahsa Rahimi Resketi, Homayun Motameni, Ebrahim Akbari, Hossein Nematzadeh
Department of Computer Engineering
Islamic Azad University, Sari, Mazandaran, Iran

Abstract

Due to the increase of video data generated in today's life, finding ways to mine these data or summarize them has become a great challenge. Thus, the main objective of this work is to investigate why video summarization is needed and how it could be done in the best possible way. To meet this goal, a systematic literature review (SLR) was conducted on articles published from 2020 to 2021 to identify the primary studies on the use of video skimming in data mining, following a predefined review protocol. The contributions and the types of evidence reported regarding the benefits of this method were also examined. In summary, the main findings of this work are: (1) the use of video summarization is widely variant, and more new domains are being introduced each year, (3) industrial researchers have not participated in academic studies, and (4) different evaluation metrics and datasets have been introduced, and studies are attempting to introduce new methods. Finally, this work showed several promising research opportunities that are pretty important and interesting but underexplored in current research and practice.

Keywords: Data mining; Multimedia; Video summarization; Video skimming; Abstraction

1. Introduction

Today, with the prevalence of cameras everywhere, easy access to the internet, and remarkable advancement in video processing, giant amounts of video data are available. While they are of great importance and can help in many ways, accessing them is challenging and time-consuming. These amounts of data would decrease the performance of various video processing-based applications like video searching, indexing, video recommendation, etc. Video summarization is an effective way to reduce these problems and decrease difficulties.

Video summarization is mainly classified into keyframe video summarization (or static video summarization) and video skimming (or dynamic video summarization). The former produces video summaries by selecting a group of keyframes representing the whole video, while the latter summarizes a video into a shorter version.

This paper attempts to focus on video skimming as well as its techniques and taxonomies. The rest of the paper is organized as follows. Section 2 reviews the backgrounds and related work. Section 3 describes the proposed framework, and Section 4 explores the research questions. Finally, the conclusion and directions for further work are presented in Section 5.

2. Proposed Methods

Systematic Literature Review (SLR) is a way to identify and evaluate the available state-of-the-art studies carried out on a given subject, based on a series of research questions using SLR, researchers attempt to find evidence for each research question and then use the collected data to do required analyses. To perform SLR in this paper, the guidelines and the systematic review protocol template proposed by Kitchenham and Charters [1] were used.

In this paper, StArt (State of the Art through Systematic Reviews) [2] is used to provide support to researchers conducting SLRs.

2.1. Research questions

Every SLR starts with a group of questions that aim to declare the primary goal of the research and to identify the most essential factors addressed in state-of-the-art research. This paper intends to answer one main question: why is video summarization essential in the industry?

Table 1. Research questions and motivations

Questions	*Description and motivation*
RQ1. What is the domain of a Video? (News, sports, rushes, movies, surveillance, user/consumer, etc.)	This question identifies the main domain that the output of the proposed method is used
RQ2. Which features are used to summarize a video?	As a video has different features, this question aims to define the features used to summarize video data.
RQ3. What are the odds, limitations, and challenges of the method?	This question intends to define the odds or challenges in summarizing videos.
RQ4. Which methods are used for video summarization?	This question defines and explores different ways of summarizing videos.
RQ5. Which datasets are used?	This question intends to explore different datasets used in the summarization methods.
RQ6. What are the evaluation methods?	The most important factor in proving the correctness of an article is evaluation.

This question would raise a group of detailed questions that may help other researchers who want to start working on this subject. All these questions are explored and declared in the next sections. Research questions, their descriptions, and motivations are presented in Table 1.

 | DOI: 10.20533/ijmip.2042.4647.2022.0064

2.2. Inclusion and exclusion criteria

A specific criterion is needed to select articles in the review because many of them are not of sufficient scientific quality and meanwhile, reading all of them is time-consuming and cannot provide accurate results. Therefore, selecting articles according to a set of criteria helps to choose the best candidates. For example, secondary articles, e.g., surveys, are only reviews of previous articles and do not offer a new method or idea; thus, they offer no value. In addition, articles that did not meet the qualitative scores were removed.

2.3. Inclusion and exclusion criteria

A specific criterion is needed to select articles in the review because many of them are not of sufficient scientific quality and meanwhile, reading all of them is time-consuming and cannot provide accurate results. Therefore, selecting articles according to a set of criteria helps to choose the best candidates. For example, secondary articles, e.g., surveys, are only reviews of previous articles and do not offer a new method or idea; thus, they offer no value. In addition, articles that did not meet the qualitative scores were removed. Two criteria were considered for inclusion and exclusion which are presented in Table 2.

Table 2. Inclusion/exclusion criteria

#	Inclusion criterion
1	Primary studies
2	Peer-reviewed studies
3	Satisfying the minimum quality threshold
#	Exclusion criterion
4	Secondary studies
5	Short papers (B5 pages)
6	Non-peer-reviewed studies
7	Duplicated studies
8	Non-English written papers
9	Studies that do not use software technology to summarize
10	Grey literature
11	Redundant paper of the same authorship
12	Studies that are not about video summarization

2.4. Sources selection and search

This paper only used electronic databases with the help of a search string, the following electronic databases were automatically searched: ACM Digital Library, IEEE Xplore, Scopus, Science Direct, Springer Link, and MathSciNet. The main goal of this step is to find the best papers that can fit the main aspect of summarization. While there are a considerable number of papers that have worked on summarization in different forms of data, e.g., video, image, text, etc., there should be a better selection of papers to avoid redundancy or choosing the wrong papers.

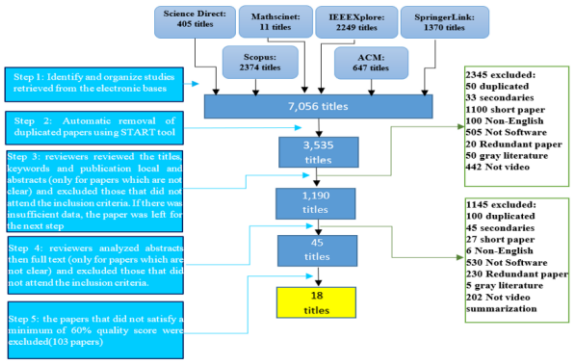

Figure 1. Study search and selection process

In step 1, a list of keywords was chosen to search through electronic databases. Therefore, the following terms were searched:

- "Summarization"
- "Video summarization"
- "Video skimming"
- "Video abstraction"
- "Video summary"

The search terms for different types of software engineering articles were combined as follows:

(1 and 2) OR 3 OR 4 OR 5

The terms were selected by reading lots of papers in this field and then extracting the main keywords. The search results (7,056 papers) were downloaded automatically and then organized and grouped with the help of the StArt tool. Different steps of the process are shown in Figure 1. After 5 different steps, finally, 18 papers remained to be considered in this review.

2.5. Quality assessment

With the help of a scoring technique, the quality assessment (QA) of the final extracted papers was done. All papers were evaluated based on a set of 14 quality criteria. Five questions were proposed according to the scope and research questions of this SLR. The assessment instrument used is presented in Table A1 (See Appendix). Questions Q7, Q8, Q9, Q10, and Q13 were adopted from the literature. The scores of questions Q2, Q4, Q8, Q9, Q13, and Q14 were determined using a two-grade scale score (Yes/No). If the answer was yes, the paper received 1 point in this question; otherwise, it received 0 points. In addition to these alternatives, the remaining questions (except Q12) allowed a third one. If the contribution was not so strong, the study received 0.5 points, consisting of a three-grade scale score to these questions. Q12 gets 1 point if the study also added value to the industrial community and 0.5 points if it did not. These questions assessed the paper's accuracy and scientific accuracy as much as possible. It has two

 | DOI: 10.20533/ijmip.2042.4647.2022.0064

aspects. The first one assesses a study as a standard paper, the second aspect examines a paper as a video summarization article. The score of quality assessment for all articles, and the number of times which studies were cited by other papers are tabulated in Table A2 (See Appendix).

2.6. Data extraction and synthesis

To extract data from articles, at different steps, the titles, abstracts, introduction, and at last the full text of a paper were assessed. Data were extracted from the selected papers at the final step. To guide this data extraction process, the data collection approach suggested by Kitchnham and Charters [4] was adopted. While reading these papers, data were extracted with the help of a predefined form (see Table A3 in Appendix).

3. Results and analysis

The quality rate of these papers and other important details are described in the following sub-sections. After the inclusion and exclusion step, 18 papers were selected.

3.1. Quality assessment results

The extracted articles may be used as a reference in future studies, these articles must have sufficient quality. Therefore, qualitative evaluation was executed in this study to assess and select the best articles. It was done to achieve better and more accurate results. The average quality percentage of each question (in quality form) for all 18 papers is demonstrated in Figure 2, and the results are shown in Table A2, according to the assessment questions described in Table A1.

The score of any study was not less than 65%; the average score was 75%. The minimum quality of 75% was chosen to establish an acceptable quality threshold for the articles. Regarding the averages of specific quality criteria, Q4, Q8, and Q11 Questions received the lowest average scores (<0.2), Q3, Q12, Q13, and Q14 received, on average, intermediary scores between 0.75 and 0.87, and the remaining questions received the highest average scores (>0.9). The lowest average score of Q4 and Q8 (i.e., 0.37) indicates that a considerable number of the studies have not been concerned with the reuse of the main ideas of other authors. The overall quality of the selected studies was found acceptable.

3.2. Overview of the studies

In the following, the general characteristics of studies such as publication year, publication nationality, type of source, etc. are discussed in the subsections below.

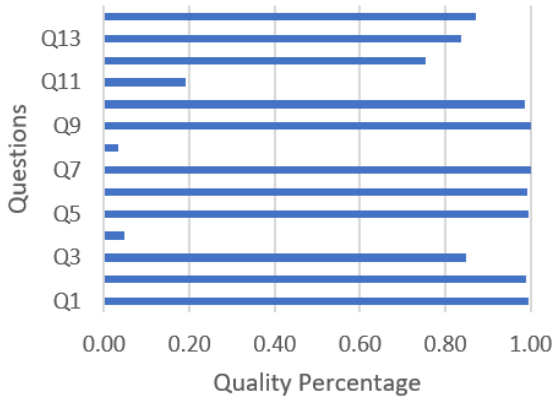

Figure 2. Quality assessment of the studies

3.2.1. Publication nationality. The nationality of articles was also determined in this research. More specifically, the nationality of the university or organization of the first author was considered. If the papers were submitted individually, only the nationality of the first author was considered. Figure 3 illustrates the authorship distribution per country.

Figure 3. Authorship distribution per country

China has the highest number of articles (50%) published among other countries, which was followed by India (22%). 6 different countries have worked in this area, which demonstrates the importance of video summarizing.

3.2.2. Keyword Cloud. To understand the importance of words in video summarization, the keywords of papers were identified. Figure 4 demonstrates the cloud of keywords. Video summarization, Skimming, neural network, and video
analysis have the highest frequency among other keywords, respectively.

3.2.3. Article's evaluations. The growth of published papers has been more tremendous from 2019 to 2021 than in previous years, which is apparently due to the emergence of new evaluations in methods and the existence of more public datasets. Video summarization studies are moving towards more accurate evaluations. The number of videos used in

 | DOI: 10.20533/ijmip.2042.4647.2022.0064

the assessments is increasing every year, and this increase has been significant compared to the early years. The number of videos used in 2021 has increased 22 times compared to 2010. There is a group of datasets that have been repeated in most of the journals. These common datasets were used from 2014, and 2020 has the most significant number of common datasets. The same thing happened to the evaluation metrics, and the most common metrics were used in 2020.

Almost all articles have compared their results with others, but this effort reached its peak in 2021. All these statistics indicate growing progress in this field every year, and the field of video summarization is moving toward a more accurate, logical, and formal evaluation. The complete details are tabulated in Table 3.

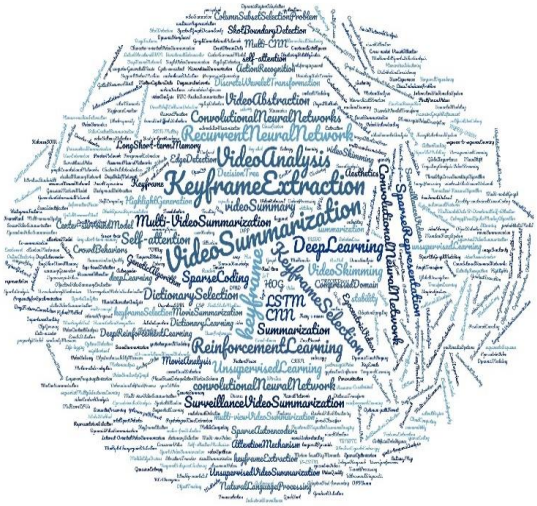

Figure 4. Word clouds of studies keywords

Table 3. Details of datasets and metrics

Year	Num of papers	Compare with others	Num of videos	Common datasets	Common metrics
2010	2	0.75	15	0	0
2011	2	0	90	0	0
2012	2	0	13	0	0
2013	5	0.2	132	0	0
2014	10	0.7	828	5	5
2015	13	0.77	422	2	6
2016	15	0.8	791	4	8
2017	11	0.86	1054	1	7
2018	17	0.75	707	5	11
2019	25	0.88	3404	12	19
2020	31	0.88	3390	18	25
2021	46	0.92	29087	8	4

3.2.4. Application context. The study settings were categorized into industry and academic contexts. The majority of the papers (97%) were placed in the academic category. Although video mining and summarization products are generally observed in our daily lives, the number of articles working directly on the industrial results is small, perhaps due to the lack of interest of software developers in this field to enter the academic field. The distribution of articles on academy and industry is displayed in Figure 5. All studies are tabulated in Table 4.

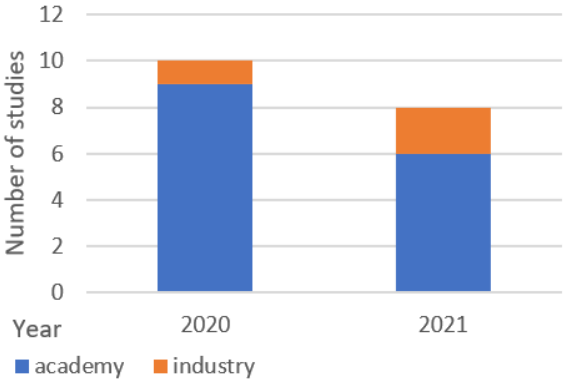

Figure 5. Defining academy and industry studies

Table 4. The studied articles

ID	Author	ID	Author
S001	Khan et al. [3]	S010	Claborne et al. [12]
S002	Ji et al. [4]	S011	Zhaho et al. [13]
S003	Ji et al. [5]	S012	Zhu et al. [14]
S004	Muhammad et al. [6]	S013	Li et al. [15]
S005	Sahu et al. [7]	S014	Shingrakhia & Patel [16]
S006	Zhang et al. [8]	S015	Elharrouss et al. [17]
S007	Xu et al. [9]	S016	Haruyama et al. [18]
S008	Shingrakhia et al. [10]	S017	Li et al. [19]
S009	Archana et al. [11]	S018	Psallidas et al. [20]

3.3. RQ1. What is the domain of a Video?

3.3.1. Results. Domain means different genres, types, or usage of video summarization. This question distinguishes different kinds of using video theoretically or practically. Most of the articles (16%) did not mention any specific domain; they were considered to be used generally. 16 different types of domains were introduced in the papers and the most frequent ones are browsing and indexing (10%), sports (7%), user-generated videos, internet, and movies (6%), data storage and compression, and egocentric (5%).

3.3.2. Analysis and discussion. The question addresses various areas covered by the video summary. By doing this, one can understand the multiple applications of video summarization in industry, commerce, science, etc. It can also be used to come up with new ways to use video summarization. Due to the increase of video data in today's world, video summarization applications are constantly increasing, and each time, a new method for more creative use of it is presented in the literature. Many articles have not specified a certain scope for

applying their proposed method. Several more commonly-used fields frequently repeated in most articles are:

- Browsing videos
- User-generated videos
- Egocentric videos
- Sport videos

Covid19 made lots of industries establish lots of virtual conferences or lecture videos, now more than ever; finding ways to summarize these data would be in great need.

3.4. RQ2. Which features are used to summarize a video?

3.4.1. Results. A video is defined by its features. This question attempts to explore video features and their use in the video summarization process. As tabulated in Table 5, most of the papers (61%) have used only one feature for video summarization, fewer articles used more features. Table A4 shows the number of features used in summarization methods.

Table 5. Number of features used in various studies

Number of used features	*count*	*%*
1 feature	11	61
2 features	1	6
3 features	2	11
4 features	1	6
Not Defined	3	16

Each video has several features that can be used to summarize the video with a variety of summarizing methods. according to the study, 17 types of features were introduced.

Nearly all articles used one feature, followed by the ones that used 2 up to 4 features; however, it is clear that most articles preferred to use fewer features. The following features have been used the most:

- Visual features
- High-level deep features
- Colour features

Generally speaking, features like colour features or visual features were used from the very beginning of video summarizing till now. But since then, many new features were presented that help to achieve better results. By exploring the development of features, it seems in the last five years, researchers have tried to use new kinds of features which help them to achieve better semantical results. Therefore, semantic level features like deep visual features, SIFT, and FC features are used vastly, and spatiotemporal, Convolutional Neural Network (CNN) features and HOG are favoured more than ever in recent years and that is because of the algorithms that are used recently.

3.5. RQ3. What are the odds, limitations, and challenges of the method?

3.5.1. Results. The purpose of this research question is to identify different challenges in video summarization. After understanding the challenges in such articles, a suitable solution could be found to deal with them. General challenges in the process of video summarization are as follow:

- The Idea of users:
  - Bridging the semantic gap has been the most difficult challenge.
  - The best summary of a video differs among different people due to its subjective nature.
  - The ability to adapt the summary concerning the application and user expectations.
- The difficulties in learning complicated semantic structural relations between videos and summaries.
- The gigantic volume of video data whose indexing, and management are problematic issues. It is highly challenging because of the enormous size of the data.
- When video collections become huge, how to explore both within and across videos efficiently is a challenging task that is often performed slowly with considerable duration and high-quality video data.
- Delimiting shot boundaries to extract a representative keyframe from each shot is not trivial as most shot boundary techniques are heuristic and sensitive to the types of video transitions.
- Browsing and editing videos is a tedious job as many videos are usually extremely unstructured, and long-running performed operations are essential elements. Therefore, the complicated video structure cannot be sufficiently exploited.
- Many of the existing video abstraction approaches have high computational requirements, which complicates the integration and exploitation of current technologies in natural environments.

Other than these challenges, there are a group of domain-specific challenges, which are tabulated in Table A5 (See Appendix).

3.5.2. Analysis and discussion. By exploring the challenges in general, most of them are related to how users see the results of the video summary. The preponderance of articles has considered this challenge to be a semantic gap between the result of automated video summarizing and the human aspect. It is almost difficult to eliminate or reduce this distance. The second reason is the high volume of videos (which usually does not have a specific structure), which complicates any search on them,

 | DOI: 10.20533/ijmip.2042.4647.2022.0064

mainly when summaries are provided for online uses, which in this case, they should be compatible with network bandwidth and many other factors that can be problematic.

3.6. RQ4. Which methods are used?

3.6.1. Results. To provide a good video summary, researchers must use a variety of methods. In the articles reviewed in this study, a variety of ways are offered, the most used methods in this context are as follows:

- Neural network
- Machin learning methods
- Using classifier

The most used algorithms are:

- LSTM
- Neural network
- CNN

3.6.2. Analysis and discussion. In 2018 studies tried a more meaningful summarization with the help of artificial algorithms and clustering. Graph-based methods, neural networks, and greedy optimization algorithms were in great use in 2019, it was the start of paying more attention to LSTM, CNN, and similar methods. These methods evolve in 2020 graph-based methods, clustering, CNN, and most summarization network algorithm were used more than ever. In 2021, studies tried to find new ways to improve and develop the most recently used algorithms in the last two years; for algorithms like LSTM, CNN, graph network, etc.

3.7. RQ5. Which datasets are used?

3.7.1. Results. All data mining methods require datasets to work with and assess their results. 13 different public datasets and 2 self-made datasets have been used in various studies. Eight papers have not used any datasets. Most of the studies used 75 videos in their assessments. The minimum number of videos was eight, and the maximum was 164 videos. Most studies used TvSum, SumMe, or video projects, or two of them together. Table 6 shows the number of datasets used in the articles.

Table 6. Number and percentage of datasets

Number of datasets used	*count*	*%*
1 dataset	4	22
2 datasets	3	17
3 datasets	4	22
4 datasets	2	11
5 datasets	2	11
Not Defined	3	17

Table A6 (See Appendix) shows all the datasets used in different articles, along with the number of videos that exist in each of these datasets. A series of datasets are constantly reused in various articles. The maximum number of used datasets belong to the:

- SumMe
- Tvsum
- YouTube
- Open Video Project (OVP)
- VSUMM

3.7.2. Analysis and discussion. Some articles have not used any data to evaluate their methods (6% of all the articles). Many articles pointed to the lack of good quality or usable datasets in a specific field; therefore, 12% of the articles created new ones. While there are datasets that support the different genres of video data, there is a great need for a more consistent dataset that covers most of the situations. A dataset that can be updated by users and then explored and analyzed by experts.

3.8. RQ6. What are the evaluation methods?

3.8.1. Results. Figure 6 shows all the evaluation methods used in the considered articles. The majority of articles evaluated the papers objectively, and the remaining ones evaluated their study subjectively or used a mixture of subjective and objective evaluation methods. It means that the subjective evaluation of papers is complicated, which is due to the challenges between what users want and what the automated results of summarizations are.

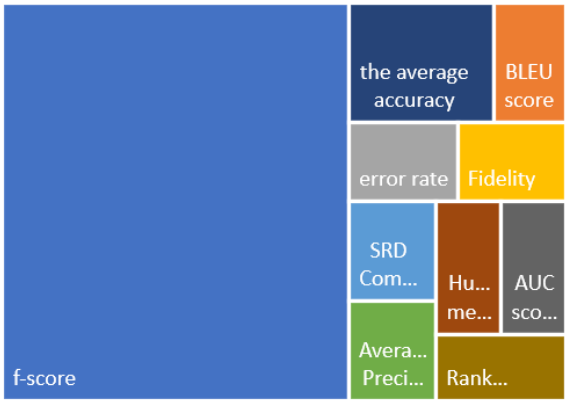

Figure 6. Different metrics for evaluating the results

3.8.2. Analysis and discussion. Comparing the proposed method of an article with that of other valid articles shows its accuracy level. In a qualitative evaluation, the quality of the final results is essential. In this method, the opinion of the human factor is the criterion. Therefore, in most cases, there is no convenient and automated evaluation method, and only 12% of articles presented qualitative comparisons or used a combination of them. Many

 | DOI: 10.20533/ijmip.2042.4647.2022.0064

articles have attempted to create an automated evaluation method, but it is so hard. Because video summarization is a relatively new science, at first, just a low number of evaluation methods were introduced or used, and there were no public datasets to compare the final results with, but over time, interest in conducting research in this field increased, and with the continuous progress in accessing videos (software or hardware) and the increase of articles and datasets, the possibility of evaluation in this field grew to a great extent. Therefore, nearly 87% of the studied articles have compared their results with others, and only 13% of them have not made this assessment due to the uniqueness of their work or other reasons. It seems there is still a great lack of qualitative evaluation, while it is more favoured but because of difficulties in these kinds of evaluation, the number of articles that worked on it is a little low.

4. Discussion

This section starts with a discussion about the scope of this review. Next, some related studies are summarized and discussed, and also the threats to the validity of this work are addressed.

4.1. Scope of this systematic review

The SLR focused on video skimming and attempted to explore video mining, video summarization, and its application in different ways. The present research tried hard to select the best articles from past decades and focus on exploring this field accurately.

4.2. Related work

It was noted that some studies have investigated different types of video summarization methods; however, there is still a lack of a systematic review that can easily cover all articles. Most of the studies in this regard are domain-specific surveys. Tanveer Hussain et al. [21] presented a comprehensive study of multi-view video summarization. Kalaivani and Roomi [22] reviewed event detection and video summarization. They investigated unusual event detection and multi-camera video surveillance systems.

4.3. Threats to validity

This section describes concerns that must be improved in future replications of this study and other aspects that must be taken into account to generalize the results of the SLR performed in this work. Construct validity: This paper focused on video skimming, but while there is a considerable number of papers that worked on data mining and text or multimedia summarization, distinguishing between them is too hard.

4.4. Challenges

By reviewing the selected articles, it seems the most important challenges are:

- Many articles have worked on video summarization and bring out new ideas every day, but still, this is a need for more practical methods. Some general ways of using video summarization (such as sports, news, and rushes videos) are presented, but no further effort was made. Most of the articles worked on improving the basic algorithms, which is good, but it still lacks creative ideas for using video summarization. Presenting more creative methods would help the increase of using video summarization at the industry level.
- It is certain that despite the existing advances in the presented articles, the ability of these algorithms to summarize video semantically is not still so good, and there is a great need for further improvement in this field.
- One of the most important issues in video summarization algorithms is the evaluation of the results. It is hard to have a correct and qualitative semantic assessment. Various methods have been proposed that evaluate the result formally, and there is an essential need for new methods to evaluate the results semantically or evaluate them as human agents.
- One of the main problems of video summarizing or video mining, in general, is that they require strong software or hardware systems to process the data. It is necessary to propose new methods which can process the algorithms more easily and accurately.

4.5. Further research

The SLR generated several promising research directions that are of high importance but under-explored in current research and practice: How can we apply video summarization to the industry sector and how many researchers are eager to work on this idea?

1. RQ1. What is the domain of a Video? (News, sports, rushes, movies, surveillance, user/consumer, egocentric, etc.) The majority of papers did not mention any specific domain for their proposed method. It may suggest that researchers don't need to declare the particular applicants of their work and they preferred a general method. For further research, exploring the reason for this aspect would be interesting.

2. RQ2. Which features are used to summarize a video? Many articles used visual features or high-level features. For further research, exploring these kinds of features would be a good idea.

 | DOI: 10.20533/ijmip.2042.4647.2022.0064

3. RQ3. Which methods are used for video summarisation? Many different methods and algorithms exist in the literature for video mining and summarization. A more detailed study of these methods and their characteristics would be of great interest. While a lot of studies worked on artificial methods, exploring these methods, and trying to use them with other useful methods would be a great idea.

4. RQ4. Which datasets are used? Summarizing user-generated videos or egocentric videos are being more and more popular, yet there is not a rich database in the field that can support all the cases, working on this database would help studies to gain a better evaluation.

5. RQ5. What are the evaluation methods? One of the shortcomings of evaluation in video summarization is the lack of an appropriate quality evaluation method, which is usually obtained through human measurement. Creating an automatic evaluation that is judging just like humans, would be a good solution.

5. Conclusion

The present study conducted an SLR to investigate the use of video skimming. The main goal was to improve the understanding of its advancements and their usage and also to identify their evidence in the literature.

A group of 7056 articles from 2020 up to 2021 on video skimming was examined in this research; after a restricted citation filter and quality assessment, 18 studies were finally included. nine research questions were selected, and the extracted studies were investigated according to these questions. The number of articles working on this area is increasing, and new evaluation methods or datasets are being introduced each year and yet there is a great need for new methods to improve the results. The findings showed that industrial researchers are not eager to participate in academic studies and there should be a way to reduce this gap. At last, it can be said that video summarization is nearly a new field that is evolving each year.

In future work, we intend to further investigate some of the research directions presented in this paper.

6. References

[1] Ba, K., and Charters. S. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering. 2.

[2] LAPES. (2004). Start-state of the art through systematic review tool. http://lapes.dc.ufscar.br/tools/start_tool. (Access Date: 13 November 2021).

[3] Khan, A.A., et al., Content-Aware Summarization of Broadcast Sports Videos: An Audio-Visual Feature Extraction Approach. Neural Processing Letters, 2020. 52(3): p. 1945-1968.

[4] Ji, Z., et al., (2004). Deep attentive and semantic preserving video summarization. Neurocomputing. 405: p. 200-207.

[5] Ji, Z., et al., (2020).Deep Attentive Video Summarization With Distribution Consistency Learning. IEEE Transactions on Neural Networks and Learning Systems. p. 1-11.

[6] Muhammad, K., Hussain, T., and Baik, S.W. (2020). Efficient CNN-based summarization of surveillance videos for resource-constrained devices. Pattern Recognition Letters, 2020. 130: p. 370-375.

[7] Sahu, A., and A. S. Chowdhury, (2020). Multiscale summarization and action ranking in egocentric videos. Pattern Recognition Letters. 133: p. 256-263.

[8] Zhang, Y., et al., (2020). Unsupervised object-level video summarization with online motion auto-encoder. Pattern Recognition Letters. 130: p. 376-385.

[9] Xu, J., Sun, Z., and Ma, C. (2020). Crowd aware summarization of surveillance videos by deep reinforcement learning. Multimedia Tools and Applications, 2020.

[10] Shingrakhia, H., and Patel, H. (2020). Emperor Penguin optimized event recognition and summarization for cricket highlight generation. Multimedia Systems, 2020. 26(6): p. 745-759.

[11] Nanjaiyan, A., and Malmurugan, N. (2020). Multi-edge optimized LSTM RNN for video summarization. Journal of Ambient Intelligence and Humanized Computing.

[12] Daniel, M., Claborne, K.T.P., Rysavy, S. J., Henry, M. J. (2020). Video Summarization Using Deep Action Recognition Features and Robust Principal Component Analysis. Systemics, Cybernetics, and Informatics, 2020. 8: p. pp 50-58.

[13] Zhao, B., Li, X., and Lu, X. TTH-RNN: Tensor-Train Hierarchical Recurrent Neural Network for Video Summarization. IEEE Transactions on Industrial Electronics, 2021. 68(4): p. 3629-3637.

[14] Zhu, W., et al., (2021). DSNet: A Flexible Detect-to-Summarize Network for Video Summarization. IEEE Transactions on Image Processing. 30: p. 948-962.

[15] Li, P., et al., (2021). Exploring global diverse attention via pairwise temporal relation for video summarization. Pattern Recognition. 111: p. 107677.

[16] Shingrakhia, H. and Patel, H. (2021). SGRNN-AM and HRF-DBN: a hybrid machine learning model for cricket video summarization. The Visual Computer.

[17] Elharrouss, O., et al., (2021). A combined multiple action recognition and summarization for surveillance video sequences. Applied Intelligence. 51(2): p. 690-712.

 | DOI: 10.20533/ijmip.2042.4647.2022.0064

[18] Haruyama, T., et al., (2021). User-selectable Event Summarization in Unedited Raw Soccer Video via Multimodal Bidirectional LSTM. ITE Transactions on Media Technology and Applications. 9(1): p. 42-53.

[19] Li, W., et al., (2020). From Coarse to Fine: Hierarchical Structure-Aware Video Summarization. p. 75-87.

[20] Psallidas, T., et al., (2021). Multimodal Summarization of User-Generated Videos. Applied Sciences. 11(11).

[21] Bradski, G., Keahler, A., and Pisarevsky, V. (2005). Learning-based computer vision with Intel's open-source computer vision library. Intel. Technology Journal. 9: p. 119-130.

[22] Poleg, Y., Arora, C., and Peleg, S. (2014). Temporal Segmentation of Egocentric Videos. in 2014 IEEE Conference on Computer Vision and Pattern Recognition.

7. Appendix

For more information, some important tables are presented in this section, which would help further research in the video summarizing field.

Table A1. Assessment criteria for the quality of studies

#	Questions	Possible answers
1	Is there a rationale for why the study was undertaken?	Y=1, N=0, P=0.5
2	Is the paper based on research (or is it merely a ''lessons learned'' report based on expert opinions)?	Y=1, N=0
3	Is there a clear statement of the goals of the research?	Y=1, N=0, P=0.5
4	Does the study reuse an existing ontology?	Y=1, N=0
5	Is the proposed technique clearly described?	Y=1, N=0, P=0.5
6	Is there an adequate description of the context (industry, laboratory setting, products used, etc.) in which the research was carried out?	Y=1, N=0, P=0.5
7	Does the study bring out a new method in the video summarizing or just use this approach in a case study?	Y=1, N=0, P=0.5
8	Is the study supported by a tool?	Y=1, N=0
9	Is the study empirically evaluated?	Y=1, N=0
10	Is there a discussion about the results of the study?	Y=1, N=0, P=0.5
11	Are the limitations of this study explicitly discussed?	Y=1, N=0, P=0.5
12	Does the research also add value to the industrial community?	Y=1, P=0.5
13	Is the proposed method evaluated on a public dataset?	Y=1, N=0
14	Is the proposed method compared with those proposed in similar papers?	Y=1, N=0

Table A2. Quality and citation of papers

ID	Num of authors	cited	Qual (%)	ID	Num of authors	cited	Qual (%)	ID	Num of authors	cited	Qual (%)
S001	4	13	78.57	S007	3	1	75.0	S013	6	11	78.6
S002	4	8	78.57	S008	2	6	82.1	S014	2	3	85.7
S003	5	22	89.29	S009	2	2	75.0	S015	5	24	85.7
S004	4	5	78.57	S010	4	0	75.0	S016	3	0	71.4
S005	2	6	75.00	S011	2	17	78.6	S017	7	0	85.7
`S006	5	45	78.57	S012	2	4	78.6	S018	4	0	78.6

Table A3. Description forms of articles

#	Study data	Description	Relevant RQ
1	Study identifier	Unique id for each study	Study overview
2	Authors, years, title, citation, country		Study overview
3	Article source		Study overview
4	Type of article	Journal, Conference, Workshop, book chapter	Study overview
5	Research method	Experiment, Case study, survey, not applicable	Study overview
6	Video domain	What is the domain of a Video? (News, sports, etc.)	RQ1
7	Features	Which features are used to summarize a video?	RQ2
8	Content or context	What are the odds, limitations, and challenges of the method?	RQ3
9	Methods	Which datasets are used?	RQ4
10	Datasets	What are the evaluation methods?	RQ5
11	Video applicant	What is the domain of a Video? (News, sports, rushes, movies, surveillance, user/consumer, egocentric, etc.)	RQ6

Table A4. Different features of studies

Feature	%	Feature	%	Feature	%	Feature	%
Visual features	17	Not defined	8	high-level deep features	8	global frames	4
Frame-specific	4	low-level features	4	context	4	Optical flow	4
histogram of oriented gradient (HOG)	4	audio features	4	dominant green-color pixel ratio (DGPR)	4	Total Emotion Count (TEC)	4
sift	4	MSER features	4	audio features	4	Total Emotion Intensity	4
Max Emotion Intensity	4	local features	4				

Table A5. Domain-specific challenges

Domain	Challenges
Movie	Summarization of a movie becomes a substring mining problem. Substrings are mined hierarchically from scene level to storyline level, which keeps continuity and completeness of the skims.
User-Generated Videos	Different users might take videos with different styles, view angles, and depth of fields for the same scenery, resulting in several unaligned videos with partially different semantics. These videos are usually captured when the photographers are moving (e.g., walking, running, or bicycling) and thus result in shakiness. The existing methods have two critical issues for summarizing user-created videos: 1) information distortion and 2) high redundancy among keyframes. Most of the videos are taken by amateur users, many of which are less aesthetically pleasing. Videos are captured with a moving camera that constantly changes its viewing direction. Finding the significant and valuable portion of the video one needs to understand the content present in it. Moreover, the categories of videos over the broad web are very diverse, like home videos, documentaries, sports videos, etc. So, it makes video summarization more brutal because of the unavailability of prior knowledge. Unfortunately, most user-generated videos lack any tags or comments to indicate their categories.
Lecture	Lecture videos are generally recorded indoors, low illuminated, in noisy environment conditions, and the contents of the scene rarely change much.
Sports	Automatic generation of highlights from a sports video is a challenging task as different sports games have different rules and situations.
Surveillance Videos	The main difficulty of large-scale surveillance video summarization arises from the contradiction between the high-degree spatiotemporal redundancies and the limited storage budget. A quick view of such crowd surveillance video in a constrained time is in increasing demand because it always contains a huge number of redundancy frames
Egocentric Videos	Egocentric videos are very shaky and contain abrupt changes. Egocentric videos are highly redundant.

Table A6. Different datasets and their distribution in papers

dataset	%	dataset	%	dataset	%	dataset	%
SumMe	24	TvSum	22	YouTube	12	Open Video Project	7
VSUMM	7	UMN	5	PETS	5	WorldExpo'10	5
MED	5	OrangeVille	2	CoSum	2	Vtw	2

 | DOI: 10.20533/ijmip.2042.4647.2022.0064