[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