DOI 10.34185/1562-9945-6-137-2021-06
UDC 004.89

O.O. Linevych, T.A. Likhouzova

SEARCH ENGINES FOR VIDEO

Annotation. The article is devoted to an overview of search engines that process media files. The relevance of the work is due to the fact that more and more information in the Internet is represented not by text, but either images or video. The purpose of the study is to identify problems that affect the search engines work performance. The key points for any search engine are the form of saving the tag tags used in the search and the algorithms used for the search (these factors determine the speed of the search), as well as how accurately the tags characterize the video (this factor determines the accuracy of the search). We are primarily interested in the accuracy of constructing tags that describe video clips. To increase a system efficiency it is necessary to make a model that would simultaneously combine sound and images (i.e. combine information that was produced during analysis of both), regarding the history of sounds and images of the video and form logical connections with minimal information loss.
Key words: search engines, video analysis, tag building.

Introduction. The amount of video content in the World Wide Web actively grows, hence an ability to examine it effectively becomes severe. To find wishful information users undergo a long and exhausting search in the Internet, caused by software engines that provide more video records both that fill and unfill requirements. As a result users need to analyze those records to extract essential one.

According to web sources reports, based on society researches [1], the amount of Internet content will increase up to 80% in 2021 year. It means that video content popularity grows every day, thus video operations that ameliorate working with it is a promising direction.

Research purpose. The aim of the research is to reveal problems that affect video searching function efficiency of search engines.

The structure of a search engine comprises of three modules. Modules are responsible for:

© Linevych O.O., Likhouzova T.A., 2021

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

- video data receiving;
- data analysis;
- providing of analysis results in a structured logically correct way.

The research is dedicated to the data analysis module, which is planned to be optimized in future studies. Hence now the task is to find a prototype system which would provide users the most suitable video results according to users requirements.

System should be represented as a software system.

1. System input data: video files.
2. Files formats: MPEG, AVI, MP4.
3. File attributes that will be analyzed:
   - static;
   - dynamic.

Static attributes: format, name, size, quality that help to estimate video analysis (i.e. dynamic attributes) – speed of the process and quality of the process results. Examples: time analysis of AVI format video file will vary from WMV video, because of files structures difference; meaningful file name may describe its content and speed up its analysis; file size determines analysis time.

Dynamic attributes: sounds and images of video objects. Objects are natural, artificial or abstract (something that can be comprehended intellectually e. g. definitions or events) elements. Object is characterized by its state, behavior (that can be described as an action list) and identity [2]. People examine characteristics of objects with the help of sight, hearing, touch, smell organs. Considering that fact a video provides sound and image information, people can use sight and hearing organs without an ability to touch or smell it. From that fact it is can be inferred that sound complements image and vice versa, because:

- image analysis provides information about objects sizes/shapes/colors/origin/materials/use;
- sound analysis concomitant provides information about the actions of objects;
- analysis of sequentially connected sounds/images provides with processes (actions) information that are performed by objects and object states changes (more in [3]).

Based on the last dynamic attributes fact it was deduced a rule: to obtain complete accurate information about the content of a video (compared to human capabilities), a system should analyze each dynamic attribute of the video in the context of the whole video (thus "analysis of related images/sounds sets" describes the processes, rather than taken out of context, objects without knowledge of changes in their states and behavior). This rule is applied during the review phase of existing video analysis systems to determine its quality. A system is of a high quality if provides complete and accurate information about video content in the shortest possible time compared to existing analysis systems (including human).

Analysis of recent research and publications. Currently, there are many search engines on the market that can work with video [4]. According to the scope, they can be grouped into the following categories (for each of the categories given its typical representatives):

- any (Amazon Rekognition [5], IBM Watson Visual Recognition [6]);
- recommendation systems (film or video clips searching – YouTube video recommendation system [7]);
- manufacturing (image recognition/operational equipment – Cogniac [8]);
- sport (people recognition, events registration – Kinovea [9], SkillSpector [10]);
- distance education (proctoring systems – Dartfish [11]);
- security systems (video surveillance, security, incident analysis [12]).

According to the input information, search engines are divided into those that analyze:

- image (an overview of such systems is in [13]);
- image and sound (an overview of such systems is in [14]);
- image and text (an overview of such systems is in [15]);
- image, sound and text (an overview of such systems is in [16]).

Video elements that can be recognized by search engines in a recognition complexity order:

- sound to text conversion (subtitles creation);
- image text recognition
- person face and identity recognition;
- object recognition by its characteristics;
- object trajectory recognition;
- events identification and description.

Video recognition (both sound and image) is a relatively simple task, which is currently well developed and implemented in all considered systems at a high level.

To solve the problem of face recognition in most cases, classifiers are used, most often based on the method of k nearest neighbors (kNN and the method of reference vectors SVM), and neural networks of different types [17].

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

Recognition of objects by features is similar in principle to the recognition of faces with their subsequent identification, but requires a much broader basis for the preservation of reference samples and a more complex model for their classification. There is also a need to re-train the model to add a new recognizable object. To solve this problem, neural networks are mostly used [16, 17, 18]

Object trajectory recognition is a relatively new task that requires large computing power to be solved. Now methods for its solution are offered: the hidden Markov model, multi-object tracking on the basis of Siamese neural networks and use of heterogeneous associative graphs [16, 19, 20].

And the last – identification, a coherent description of events or processes, can be solved as a combination of previous tasks. Currently, there are few such solutions and they can be improved [16, 21].

The results of the search engine work can be in the form of:

- video files or their fragments that match the specified search keys;
- a list of tags that best describe the video;
- a list of objects and/or people present in the video;
- a list of events recorded on video;
- detailed logically-connected structured text description of properties and behavior of objects, and also directly or indirectly connected processes.

The key points for any search engine are the form of saving the tag tags used in the search and the algorithms used for the search (these factors determine the speed of the search), as well as how accurately the tags characterize the video (this factor determines the accuracy of the search) [22–26].

We are primarily interested in the accuracy of constructing tags that describe video clips. In this case, the greatest attention should be paid to the identification of events and objects involved in these events.

Conclusions. The main problem of existing systems is the lack of a logically connected detailed summary of video living/man-made/automatic systems [27], which would describe the events, behavior, properties and relationships of the video objects.

Excessive conversions of video information from data to text or image to text, use of only one information source (either sound or image) lead to the loss of relationships and other information of objects. To increase a system efficiency it is necessary to make a model that would simultaneously combine sound and images (i. e. combine information that was produced during analysis of both), regarding the history of sounds and images of the video and form logical connections with minimal information loss. For example, let's describe a moment in time: there is a sound of a shot, image with a bursting inflatable ball: first we combine sound and image – if the human can describe the event, she/he immediately names and describes it, if not, then can begin to create logical connections regarding event history, previously analyzed information about objects existed in a video and acquired for the life knowledge base.

Thus, the next step in improving video systems is to create models (such as neural networks), which can be used to build logical inferences as humans do.

REFERENCES

1. Best Free Social Media Search Tool – Talkwalker [Electronic resource] Accessed (2021-06-15): https://www.talkwalker.com/social-media-analytics-search
2. Grady Booch, Robert A. Maksimchuk, Michael W. Engle, Bobbi J. Young, Jim Conallen, Kelli A. Houston "Object-Oriented Analysis and Design with Applications". The Addison-Wesley Object Technology Series, 2007 – 717 pages, ISBN 0-201-89551-X
3. Herbert A. Simon "The Architecture of Complexity," in Proceedings of the American Philosophical Society, Vol. 106, No. 6, pp. 467-482, Dec. 12, 1962. https://www.jstor.org/stable/985254
4. Best Image Recognition Software [Electronic resource] Accessed (2020-12-25): https://www.g2.com/categories/image-recognition
5. Amazon Rekognition [Electronic resource] Accessed (2021-05-10): https://aws.amazon.com/ru/rekognition/?blog-cards.sort-by=item.additionalFields.createdDate&blog-cards.sort-order=desc
6. IBM Watson Visual Recognition [Electronic resource] Accessed (2021-05-10): https://www.ibm.com/products/maximo/remote-monitoring
7. J. Davidson, "The YouTube video recommendation system", Proc. 4th ACM Conf. Rec. Syst., pp. 293-296, 2010.
8. Cogniac [Electronic resource] Accessed (2021-05-10): https://store.sap.com/dcp/en/product/display-0000059790_live_v1
9. Kinovea [Electronic resource] Accessed (2021-05-10) https://www.kinovea.org/features.html
10. SkillSpector [Electronic resource] Accessed (2021-05-10): https://skillspector.software.informer.com/1.3/
11. Dartfish [Electronic resource] Accessed (2021-05-10): https://www.dartfish.com/
12. Video Surveillance [Electronic resource] Accessed (2021-05-10): https://www.g2.com/search?utf8=%E2%9C%93&query=Video+Surveillance

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

13. Kanrar, S., Mandal, N.K.” Video traffic analytics for large scale surveillance”, in Multimed Tools Appl 76, 13315–13342, 2017.
https://doi.org/10.1007/s11042-016-3752-0
14. B. Huurnink, C. G. M. Snoek, M. de Rijke and A. W. M. Smeulders, "Content-Based Analysis Improves Audiovisual Archive Retrieval," in IEEE Transactions on Multimedia, vol. 14, no. 4, pp. 1166-1178, Aug. 2012.
https://doi.org/10.1109/TMM.2012.2193561
15. R. Zhou, D. Xia, J. Wan and S. Zhang, "An Intelligent Video Tag Recommendation Method for Improving Video Popularity in Mobile Computing Environment," in IEEE Access, vol. 8, pp. 6954-6967, 2020.
https://doi.org/10.1109/ACCESS.2019.2961392
16. Deldjoo Y. Enhancing Video Recommendation Using Multimedia Content. In: Pernici B. (eds) Special Topics in Information Technology. SpringerBriefs in Applied Sciences and Technology. Springer, Cham, 2020.
https://doi.org/10.1007/978-3-030-32094-2_6
17. T. Schenkel, O. Ringhage, and N. Branding, ‘A Comparative Study of Facial Recognition Techniques : With focus on low computational power’, Dissertation, 2019.
https://www.diva-portal.org/smash/get/diva2:1327708/FULLTEXT01.pdf
18. Sanmoy, P., Acharya, S., & Bhuva, K. A review on facial recognition algorithms and their application in retail stores. Academy of Marketing Studies Journal, vol. 22, no 3, 7396, 2018.
https://www.abacademies.org/articles/a-review-on-providing-loyalty-discounts-to-customers-in-retail-stores-7396.html
19. Zhang, W. L., Yang, K., Xin, Y. T., & Zhao, T. S. “Multi-Object Tracking Algorithm for RGB-D Images Based on Asymmetric Dual Siamese Networks,” in Sensors (Basel, Switzerland), vol. 20, no 23, 6745, 2020.
https://doi.org/10.3390/s20236745
20. Sheng H., Zhang Y., Chen J., Xiong Z., Zhang J. Heterogeneous association graph fusion for target association in multiple object tracking. IEEE Trans. Circuits Syst. Video Technol. 2018;11:3269–3280.
https://doi.org/10.1109/TCSVT.2018.2882192
21. M. Tavassolipour, M. Karimian and S. Kasaei, "Event Detection and Summarization in Soccer Videos Using Bayesian Network and Copula," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 24, no. 2, pp. 291-304, Feb. 2014.
https://doi.org/10.1109/TCSVT.2013.2243640
22. Sedky Adly, M. S. Abdelwahab, I. Hegazy and T. Elarif, "Issues and Challenges for Content-Based Video Search Engines A Survey," 2020 21st International Arab Conference on Information Technology (ACIT), pp. 1-18, 2020.
https://doi.org/10.1109/ACIT50332.2020.9300062

23. M. Wang, R. Hong, G. Li, Z. Zha, S. Yan and T. Chua, "Event Driven Web Video Summarization by Tag Localization and Key-Shot Identification," in IEEE Transactions on Multimedia, vol. 14, no. 4, pp. 975-985, Aug. 2012.
https://doi.org/10.1109/TMM.2012.2185041
24. G. Doumenis, S. Papastefanos, V. Mateevitsi, F. Andritsopoulos, N. Achilleopoulos and A. V. Mikhalev, "Video index and search services based on content identification features," 2008 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, pp. 1-4, 2008.
https://doi.org/10.1109/ISBMSB.2008.4536659
25. J. Li and J. Z. Wang, "Real-Time Computerized Annotation of Pictures," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 6, pp. 985-1002, June 2008.
https://doi.org/10.1109/TPAMI.2007.70847
26. B. Z. Yao, X. Yang, L. Lin, M. W. Lee and S. Zhu, "I2T: Image Parsing to Text Description," in Proceedings of the IEEE, vol. 98, no. 8, pp. 1485-1508, Aug. 2010.
https://doi.org/10.1109/JPROC.2010.2050411
27. J.G. Miller "Living Systems," McGraw-Hill, 1978 – 1102 pages, ISBN 978-0070420151

Received 13.09.2021.
Accepted 18.09.2021.

Пошукові системи

Стаття присвячена огляду пошукових систем, що можуть працювати з медіафайлами. Актуальність роботи пов'язана з тим, що все більше інформації і інтернеті представлено не текстом, а зображеннями або відео. Мета дослідження – виявити проблеми, що впливають на ефективність роботи пошукових систем для роботи з медіафайлами.

Для отримання повної точної інформації про вміст відео система повинна аналізувати кожен динамічний атрибут відео в контексті цілого відео (цілого, так як "аналіз набору пов'язаних зображень/звуків" описує процеси, а не вирвані з контексту, об'єкти, без знань про зміни їх поведінки). Дане правило застосовано під час етапу огляду варіантів існуючих систем аналізу відео для визначення якості системи: система вважається якісною, коли надає максимально повну і точну інформацію про вміст відео в найкоротші строки порівняно з існуючими системами аналізу, включаючи людину.

Ключовими моментами для роботи будь-якої пошукової системи є форма збереження міток-тегів, по яким відбувається пошук, та алгоритми, що використовуються для пошуку (ці фактори обумовлюють швидкість пошуку), а також наскільки точно мітки характеризують відео (цей фактор обумовлює точність пошуку). Нас в першу чергу цікавить точність побудови тегів, що описують

ISSN 1562-9945 (Print)
ISSN 2707-7977 (Online)

відеофрагменти.

Найголовніша проблема існуючих систем - відсутність побудованої по відео логічно-зв'язаної детальної living system, яка б описувала події, поведінку і властивості об'єктів цієї системи. Від зайвих конвертацій даних звук-текст, зображення-текст та від використання при конвертації тільки одного джерела інформації (тільки звук або тільки зображення) губиться інформація про зв'язки між об'єктами. Для підвищення ефективності системи варто зробити модель, яка б одночасно поєднувала звук і зображення (проаналізовану інформацію про них), враховувала історію звуків і зображень у загальній системі всього відео, що ми аналізуємо, і утворювала б логічні зв'язки з мінімальними втратами інформації за рахунок уникнення, при можливості, конвертації.

Поисковые системы

В статье выполнен обзор поисковых систем, которые могут выполнять поиск в медиафайлах. Актуальность работы обусловлена тем, что всё больше информации в интернете представлено не в виде текста, а изображениями или видео. Цель исследования – выявить проблемы, влияющие на эффективность поисковых систем, работающих с медиафайлами.

Ліхоузова Тетяна Анатоліївна – к.т.н., доцент кафедри технічної кібернетики, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського».
Ліневич Ольга Олександрівна – аспірантка кафедри технічної кібернетики, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського».

Лихоузова Татьяна Анатольевна – к.т.н., доцент кафедры технической кибернетики, Национальный технический университет Украины «Киевский политехнический институт имени Игоря Сикорского».
Линевич Ольга Александровна – аспирантка кафедры технической кибернетики, Национальный технический университет Украины «Киевский политехнический институт имени Игоря Сикорского».

Likhouzova Tetiana – PhD, associate professor, Department of Technical Cybernetics National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute».
Linevych Olha – post-graduate student, Department of Technical Cybernetics National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute».