A SYSTEMATIC REVIEW ON CONTENT-BASED MEDICAL VIDEO RETRIEVAL

Vagner Mendonça Gonçalves[1] and Fátima de Lourdes dos Santos Nunes[1]

[1]School of Arts, Sciences and Humanities/University of São Paulo – USP, São Paulo, Brazil

Resumo: Objetivo: Identificar o estado da arte com relação à aplicação de recuperação de vídeos baseada em conteúdo no contexto de vídeos médicos. Nossa contribuição é fornecer uma visão geral sobre estudos anteriores, bem como vantagens e desafios a serem explorados no tema por pesquisadores da área de aplicações de informática em saúde. Métodos: Nós conduzimos uma revisão sistemática nas principais bases de trabalhos científicos das áreas da computação e medicina. Depois de definir um protocolo, nós conduzimos uma busca por artigos e tecemos uma discussão crítica sobre os estudos encontrados. Resultados: Os resultados mostraram que a aplicação do conceito de recuperação de vídeos baseada em conteúdo na área médica é recente e ainda pouco explorada. Conclusões: Nos trabalhos analisados, as aplicações de interesse consistem na identificação e descrição de estruturas representadas nos vídeos para posterior indexação e recuperação. Os objetivos mais comuns verificados são auxílio ao disgnóstico, análise de procedimentos e tomada de decisão.

Palavras-chave: Armazenamento e Recuperação da Informação; Diagnóstico por Computador; Literatura de Revisão como Assunto.

***Abstract**: Objective: Identify the state of the art about the application of content-based video retrieval in the context of medical videos. Our contribution is to provide an overview of previous studies as well as advantages and challenges to be explored on this subject to the researchers interested in the applied health informatics field. Methods: We conduct a systematic review in the main scientific databases of the computing and medical fields. After defining a protocol, we conducted a search and, then, a critical discussion about the studies. Results: The results showed that the application of the concept of content-based video retrieval in the medical field is recent and has been little explored. Conclusions: Applications of interest in the analyzed works consist in the identification and description of structures depicted in the videos for later indexing and retrieval. The most common goals are to aid diagnosis, procedures analysis, and decision making.*

***Keywords**: Information Storage and Retrieval; Diagnosis, Computer-Assisted; Review Literature as Topic.*

Introduction

Digital videos have an important role in multimedia applications. In the medical field, different modalities of videos have been employed in tasks ranging from diagnosis (ultrasound and endoscopy, for example), through the monitoring of procedures and surgeries, to computer-aided medical education.

Content-Based Video Retrieval (CBVR) refers to the extraction and comparison of features automatically extracted from videos to allow the query on a database providing a video or video sequence as parameter. Then, the application must retrieve the most similar videos or video sequences indexed

at the database. The feature extraction is generally related to colors, textures, shapes, movement and sound.

Although there is a considerable number of techniques used in CBVR considering the medical field, there is no studies that group them in a categorized way. Thus, it is necessary to analyze these works to describe efficient automated techniques for analysis, interpretation and retrieval of such data.

In this paper, we present the conduction and results of a Systematic Review (SR) whose aim was identifying the state of the art about the application of CBVR in the context of medical videos. Our contribution is to provide an overview of previous studies as well as advantages and challenges to be explored on this subject to the researchers interested in the Applied Health Informatics field.

The paper is organized as follows. In the Section "Materials and Methods", we present concepts of SR, the protocol used and the conduction process. We present and discuss the results in the Section "Results and Discussion". Finally, we present our final considerations in the Section "Conclusion".

Materials and Methods

The SR is a rigorous methodology of bibliographic research that aims to identify primary and secondary studies related to a particular research topic[1]. An SR is carried out in three well defined phases: planning, conduction and analysis of results[1,2]. In the planning phase a protocol is defined specifying the research questions and the methodology to be employed in the conduction of the review, among other specific points that will guide the search. In the conduction phase, the bibliographic research itself is carried out. It is in this phase that the selection of the studies occurs, according to the criteria for inclusion and exclusion defined. Finally, in the analysis of results phase the data extraction is performed allowing to compare the results in order to obtain the conclusions about them.

Planning – Our interests in this SR are applications, techniques and possible challenges related to the subject. In this context, the following research questions were defined:

- What are the current interests in the application of CBVR in the context of medical videos?
- What techniques and methodologies have been applied to CBVR in the context of medical videos?

An exploratory analysis on the subject was previously conducted using the Google Scholar tool. This first analysis provided evidences that the research on this subject by the scientific community is recent. We retrieved works which investigated the processing of medical videos throughout the 2000s, but papers dealing with CBVR were observed only in the end of the 2000s and early 2010s. Then, we do not limit the time period in our queries for searching studies of interest.

This first exploration of related works guided the selection of the reference sources and the definition of the keywords used in the SR. We have consulted bases which traditionally publish articles on the subject, according to the experience of previous works of our research group and the information collected in the preliminary exploratory analysis.

The following databases, listed in the order in which they were consulted, were selected:

- PubMed;
- PubMed Central;
- IEEE Xplore Digital Library;
- ACM Digital Library;
- Scopus.

To search in the selected databases, we used the following composition of terms: ("video retrieval" OR CBVR) AND ("computer-aided diagnosis" OR "computer-assisted diagnosis" OR "medical image" OR "medical imaging" OR surgery). The terms had to be present in the title, abstract, keywords

or indexation topics in order to make a paper be retrieved. These options were defined by means of advanced search tools available in the databases.

Only in PubMed and PubMed Central, which do not offer advanced features in their query tools, a simple search was performed using the term "video retrieval". Thus, the retrieval of all works of these bases which mention the key term was possible.

Table 1 presents the compositions of terms translated for each of the search engines of the consulted databases.

Table 1: Composition of terms used in the researches.

Source	Search tool	Composition of terms
PubMed	Search	"video retrieval"
PubMed Central	Search	"video retrieval"
IEEE	Command Search	("Document Title":"video retrieval" OR "Abstract":"video retrieval" OR "Author Keywords":"video retrieval" OR "Index Terms":"video retrieval" OR Topic:"video retrieval" OR "Document Title":CBVR OR "Abstract":CBVR OR "Author Keywords":CBVR OR "Index Terms":CBVR OR Topic:CBVR) AND ("Document Title":"computer-aided diagnosis" OR "Abstract":"computer-aided diagnosis" OR "Author Keywords":"computer-aided diagnosis" OR "Index Terms":"computer-aided diagnosis" OR Topic:"computer-aided diagnosis" OR "Document Title":"computer-assisted diagnosis" OR "Abstract":"computer-assisted diagnosis" OR "Author Keywords":"computer-assisted diagnosis" OR "Index Terms":"computer-assisted diagnosis" OR Topic:"computer-assisted diagnosis" OR "Document Title":"medical image" OR "Abstract":"medical image" OR "Author Keywords":"medical image" OR "Index Terms":"medical image" OR Topic:"medical image" OR "Document Title":"medical imaging" OR "Abstract":"medical imaging" OR "Author Keywords":"medical imaging" OR "Index Terms":"medical imaging" OR Topic:"medical imaging" OR "Document Title":surgery OR "Abstract":surgery OR "Author Keywords":surgery OR "Index Terms":surgery OR Topic:surgery)
ACM	Advanced Search	("video retrieval" or Keywords:"video retrieval" or CBVR or Keywords:CBVR) and ("computer-aided diagnosis" or Keywords:"computer-aided diagnosis" or "computer-assisted diagnosis" or Keywords:"computer-assisted diagnosis" or "medical image" or Keywords:"medical image" or "medical imaging" or Keywords:"medical imaging" or surgery or Keywords:surgery)
Scopus	Search	("video retrieval" OR CBVR) AND ("computer-aided diagnosis" OR "computer-assisted diagnosis" OR "medical image" OR "medical imaging" OR surgery)

In order to select only relevant works, we defined inclusion and exclusion criteria. We included only studies that met the inclusion criteria and any of the exclusion criteria. The defined inclusion criteria were:

(a) present or address applications and methodologies for the use of CBVR in the context of medical videos.

The exclusion criteria were as follows:

(b) not address applications or methodologies for the use of CBVR in the context of medical videos;

(c) be similar, in content and results, with other study(ies), by the same authors, that may also come to be retrieved in any of the consulted databases;

(d) not be fully available in the consulted databases or in any other database accessible on the Internet.

Conduction – The researches were carried out between March and July 2014. In total, 154 studies were retrieved. As a whole, just 8 studies (5.19%) were included. Every conduction stage of the SR was duly documented based on the models proposed in reference works[1,2]. Figure 1 shows a flow diagram[3] which summarizes the selection of the studies.

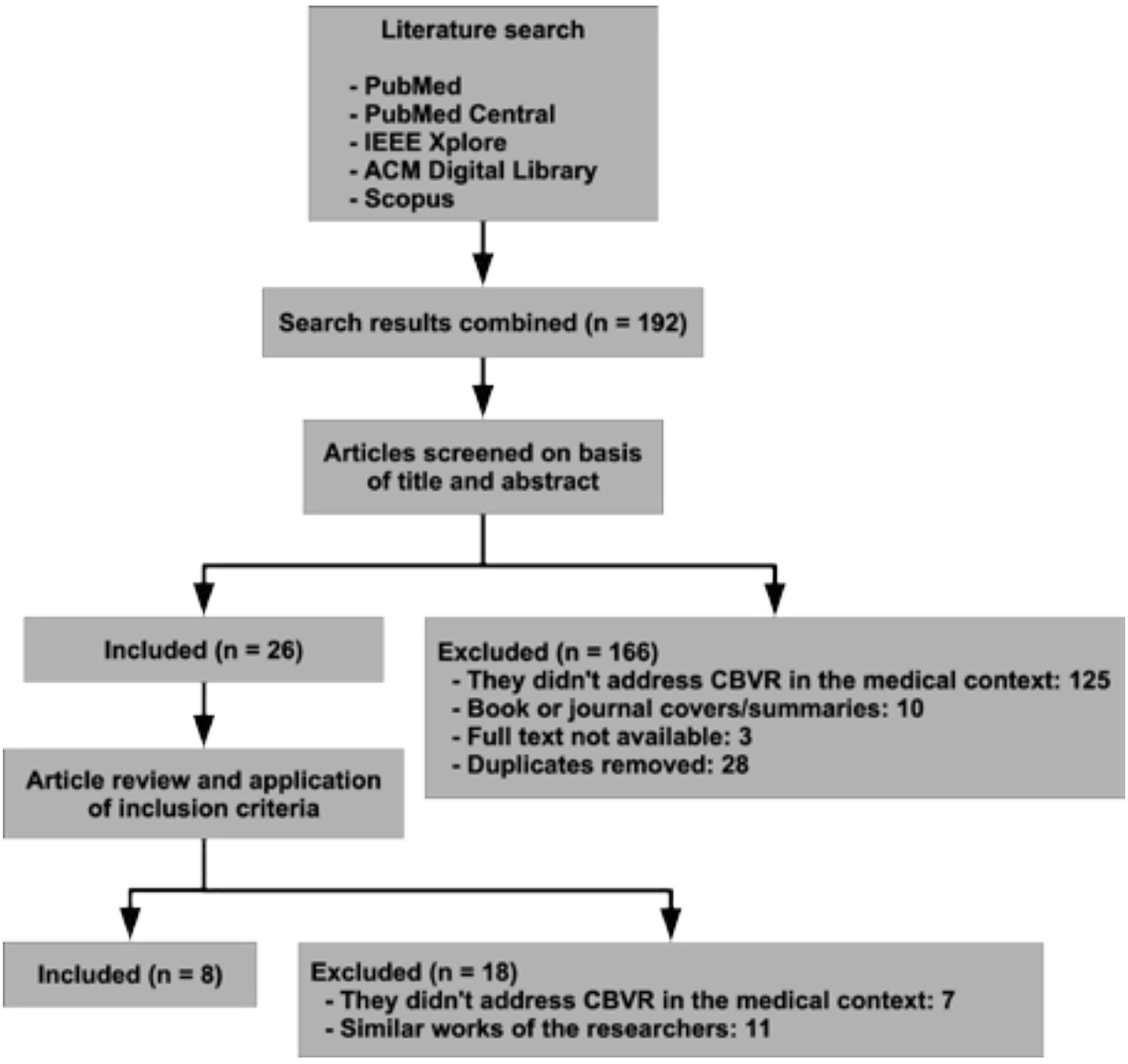

Figure 1: Flow diagram summarizing the selection studies stage.

The next section presents and discusses the results obtained through this SR.

Results and Discussion

Included works and considerations about the results – Table 2 presents the included works as well as the topics extracted from each one that are of most interest to this SR.

Table 2: Included works and topics of interest to the review.

Ref.	Modality	Body region / procedure	Application	Feature extraction	Retrieval approach
4	Monitoring of surgeries	Eyes / Cataract and epiretinal membrane surgeries	System for analysis of video sequences and identification of surgical tasks in real-time.	Texture and color through Wavelet analysis; motion by analyzing the optical flow between frames.	Search of nearest neighbors using a variation of k-d tree.
5	Confocal endomicroscopy	Colorectal region	Classification of colorectal polyps.	Visual signature characteristics obtained by an adaptation of the Bag-of-Visual-Words method.	Feature extraction; measurement of the similarity between depicted objects; training of a classifier k-nearest neighbors.
6	Confocal endomicroscopy	Colorectal region	Video retrieval system that combines visual characteristics with semantic concepts in order to reduce the gap.	Visual signature characteristics obtained by an adaptation of the Bag-of-Visual-Words method.	Extraction of visual characteristics; using a probabilistic approach (Fisher) to estimate, given a vector of visual characteristics, is a semantic concept is present in the video segment.
7	Confocal endomicroscopy	Colorectal region	System (Atlas) of endomicroscopy videos with support to retrieve by visual content.	Visual signature characteristics obtained by an adaptation of the Bag-of-Visual-Words method.	Extraction of visual characteristics; using a k-nearest neighbors classifier to identify the most similar video.

8	Echocardiography	Heart	Retrieval by extracting features derived from texts written in the video. The texts indicate measurements performed during the exam.	Detection of only textual frames; removal of any non-text content; usage of an OCR engine to extract words in the textual regions; usage of the measures identified to compose feature vectors.	Comparison by similarity between the vector of a video model and the vectors of the videos in database.
9	Laparoscopic surgical video	Intra-abdominal region / Laparoscopy	Content-based retrieval approach in which, by means of a query image, are retrieved the video frames that contain the instrument depicted in the image model.	Feature extraction using color histogram, Hough Transform (linearity of the border) and compactness level of the object of interest in the image.	Measurement of similarity between the model image characteristics and the characteristics of each frame. A threshold indicates when the similarity is such that the frame should be retrieved.
10	Videos for medical education	Monitoring of surgeries	Automatic annotation of videos focused on medical education.	Identification of objects over the frames of the videos; characterization of the objects detected by means of low-level features (color and texture).	Usage of Support Vector Machine to classify the identified objects in a hierarchical way; assignment of semantic concepts the videos according to the result of the classification; users search by key terms that refer to classes/concepts annotated.

11	Colonoscopy	Colorectal region	Content-based retrieval approach for detection of biopsy or therapy interventions in the video.	Automated identification of interventions over video by identifying instruments used by the physician (described by Fourier descriptors based characteristics); identification of words that define the region of the colon under analysis and the actions taken during the examination (usage of software that, given and audio segment, generates the text pronounced in the speech).	Users select a single image that represents the intervention, a selected scene of the intervention or a speech snippet identified at the time of the intervention. The choice of the user refers to the video sequence that contains the intervention.

As shown, the amount of retrieved works is small against the vast scientific literature involving, for example, computer-aided diagnosis. In this scenario, the hypothesis previously raised that the concept of CBVR has been little explored in the medical context is enhanced. In addition, it is a recent issue since most of the studies included were published in the end of 2000s and early 2010s. This result demonstrates that the exploitation of techniques for processing, analysis and retrieval of medical videos is a research opportunity that can generate significant contributions in the coming years.

Endoscopic techniques – From the included studies, we can observe that the most frequent interest is the analysis, description and retrieval structures from videos of endoscopic techniques. Computer-aided diagnosis in videos from Confocal Endomicroscopy is a topic of interest[5,6,7]. Confocal Endomicroscopy is a technique which allows real-time analysis with high resolution (microscopic level) of epithelial tissue of the gastrointestinal tract[5].

Works have presented studies aiming to develop a tool that uses CBVR to aid the diagnosis of colorectal cancer. They propose different goals that together tend to develop the tool, such as application of video retrieval techniques based on visual characteristics[7], combination between content-based retrieval and semantic annotation[6], and an approach that uses endomicroscopy video retrieval for classification of colorectal polyps[5].

All these works involve extraction of low-level features to describe videos. However, a concern of the researchers is the semantic gap, i.e. the difference between the meaning attributed to a video through its low-level features (colors, textures, shapes, movements) and the real meaning of the situation represented by it. All approaches presented propose adding semantic criteria in CBVR to reduce the mentioned problem[6].

A content-based retrieval approach applied to videos from another endoscopy technique, the laparoscopy, also was presented[9]. Laparoscopy is a minimally invasive surgical technique[12]. It is applied to display and operate intra-abdominal structures and organs through an incision and insertion of a device called laparoscope. The interest of the researchers was to retrieval the video frames containing the instrument used in the procedure and represented in an image provided by reference. According

to the researchers, such task is of interest to manufacturers of the instruments used in the procedures. They can analyze how their products are used, avoiding having to scan the entire video to find snippets of interest.

A content-based retrieval approach to detect therapy or biopsy interventions in videos from colonoscopy also was presented[11]. Colonoscopy is an endoscopic technique that allows the inspection of the entire colon and performing therapeutic operations such as removal of polyps[11]. The objectives of the approach range from facilitate further analysis of complications from procedures performed by colonoscopy, to the development of a CBVR system capable of supporting education and endoscopic research.

Monitoring of surgeries – In this context, an approach which aimed to develop a real time search tool for video sequences similar to a video recorded during an ophthalmic surgical procedure was presented[4]. The researchers investigated the specific cases of epiretinal membrane and cataract surgeries. The technique involves identifying the surgical task being performed in the processed video sequence. According to the researchers, this approach can warn the surgeon as well as aids the professional to make a decision when an atypical or risk situation during occurs during the procedure.

The semantic gap was also a concern[4]. The researchers used a feature weighting approach to improve the correlation between low-level features and semantic concepts related to surgical tasks.

An approach aiming to automatic annotation of videos from monitoring of surgeries focused on medical education also was presented[10]. This is a mixed approach of CBVR and keyword-based retrieval, in which the description of the video contents through extractors is used in assigning labels to them. The user uses keywords that refer to the labels assigned to find videos of interest.

Echocardiography – An approach to retrieve videos from echocardiography exams was presented[8]. According to the researchers, this kind of video is an important source of information to aid cardiac diagnosis, being able to show shape and movements of the heart from different angles.

However, the feature extraction technique presented by the researchers is quite different from other mentioned in this paper. Features are extracted from the interpretation of texts featured throughout the video segment. The texts indicate measurements performed during the exam. An Optical Character Recognition (OCR) engine is used. The measures identified from the texts are used in the composition of feature vectors, which in turn allow measuring the similarity between videos.

Methods and techniques applied in the included works – The included works in this SR show different approaches for feature extraction to describe video content as well as training approaches for classifiers to obtain useful models to retrieve similar videos and different retrieval approaches.

Works have applied approaches to define visual signatures for describing confocal endomicroscopy videos[7]. This is an adaptation of the method originally presented in another referenced work[13], called Bag-of-Visual-Words. To calculate these signatures techniques to split video frames in regions and grouping these regions are used in order to identify concepts represented in the frames[7]. The description of these concepts using mathematical and statistical tools allow users to compare the signatures of different videos.

Thinking about this approach for identification of regions linked to concepts in pictures or videos, we can observe that it is an interesting approach to aid diagnosis because it can help to identify anomalies in the image or video under analysis.

Color and texture characteristics extracted using the Wavelet Transform, as well as motion characteristics extracted using the concept of optical flow, are other techniques applied[4]. In the case of surgical monitoring, descriptive characteristics of movements are very interesting, perhaps even more than the traditional characteristics such as color, shape, and texture. Description of movements in

videos of this modality can be very useful to identify actions that occurred in the recorded procedure and deserve the attention of the specialist or surgeon.

Most studies presented including approaches to increase the semantic significance of the retrieval results, thus reducing the problem of semantic gap. For these approaches usually the researchers used classification techniques such as k-Nearest Neighbors and Support Vector Machine in order to obtain semantically efficient models using videos labelled by experts as training data.

Conclusion

In this paper we presented the results of a Systematic Review conducted to identify the state of the art about the application of CBVR in the context of medical videos. Eight works that developed techniques and approaches in this context were analyzed. The works were retrieved through systematic searches in five important databases of scientific papers.

Information systems have increasingly included multimedia data aiming to realize several goals: solve complex problems, find information in complex data, improve interaction with users, among others. Most of these challenges is due to the large growth of available digital data, due to the advancement of information and communication technologies.

In the medical field, the large volume of available digital data has potential to require the development of technologies that can help health professionals in tasks such as analysis and interpretation, computer-aided diagnosis and medical education. Therefore, efficient methods for description, storage and retrieval of such data are needed.

Currently we have as extensive amount of mathematical and statistical tools for processing and interpretation of digital images and videos. These tools can be employed in CBVR approaches for different imaging modalities and many different tasks such as computer-aided diagnosis, surgical monitoring, and medical education.

With regard to the challenges of the area, the most cited by researchers are reducing the semantic gap at the different levels of CBVR applications: data modelling, description, classification and retrieval. Another challenge is the selection and reduction of features for efficient description of videos.

The results of this SR showed that the application of the concept of CBVR in the medical field is recent and is still underexplored. Therefore, processing, analysis and retrieval of medical videos are a research opportunity that can generate significant contributions in the coming years.

Applications of interest in the works analyzed consist in the identification and description of structures depicted in the videos for later indexing and retrieval. The most common goals are to aid diagnosis, aid procedures analysis, aid the surgeon in decision making and provide easily retrievable content for medical education.

References

[1] Kitchenham BA. Procedures for Performing Systematic Reviews. Computer Science Department, Keele University and National ICT; 2004. TR/SE-0401.

[2] Biolchini JCA, Mian PG, Natali ACC, Conte TU, Travassos GH. Scientific research ontology to support systematic review in software engineering. Adv Eng Inform. 2007;21(2):133-151.

[3] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JPA, et al. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: explanation and Elaboration. PLoS Med. 2009 jul;6(7):e1000100-1-e1000100-28.

[4] Quellec G, Charrière K, Lamard M, Droueche Z, Roux C, Cochener B, et al. Real-time recognition of surgical tasks in eye surgery videos. Med Image Anal. 2014;18(3):579-590.

[5] André B, Vercauteren T, Buchner AM, Krishna M, Ayache N, Wallace MB. Software for automated classification of probe-based confocal laser endomicroscopy videos of colorectal polyps. World J Gastroentero. 2012;18(39):5560-5569.
[6] André B, Vercauteren T, Buchner AM, Wallace MB, Ayache N. Learning Semantic and Visual Similarity for Endomicroscopy Video Retrieval. IEEE T Med Imaging. 2012;31(6):1276-1288.
[7] André B, Vercauteren T, Buchner AM, Wallace MB, Ayache N. A smart atlas for endomicroscopy using automated video retrieval. Med Image Anal. 2011b;15(4):460-476.
[8] Syeda-Mahmood T, Beymer D, Amir A. Disease-Specific Extraction of Text from Cardiac Echo Videos for Decision Support. In: *Proceedings of the 10th International Conference on Document Analysis and Recognition*; 2009. p. 1290-1294.
[9] Chattopadhyay T, Chaki A, Bhowmick B, Pal A. An application for retrieval of frames from a laparoscopic surgical video based on image of query instrument. In: *Proceedings of the TENCON - IEEE Region 10 Conference*; 2008. p. 1-5.
[11] Cao Y, Li D, Tavanapong W, Oh J, Wong J, de Groen PC. Parsing and Browsing Tools for Colonoscopy Videos. In: *Proceedings of the 12th Annual ACM International Conference on Multimedia*. New York: ACM; 2004. p. 844-851.
[12] Schuessler WW, Schulam PG, Clayman RV, Kavoussi LR. Laparoscopic radical prostatectomy: Initial short-term experience. Urology. 1997;50(6):854-857.
[13] Sivic J, Zisserman A. Video Google: Efficient Visual Search of Videos. In: Ponce J, Hebert M, Schmid C, Zisserman A, editors. Toward Category-Level Object Recognition. vol. 4170 of Lecture Notes in Computer Science. Springer Berlin Heidelberg; 2006. p. 127-144.

Contact

Corresponding author: **Vagner Mendonça Gonçalves**. Student on the Master of Science Graduate Program in Information Systems at the University of São Paulo.
Phone number: +55 11 99699-4158.
Email: vagner.goncalves@usp.br.