

An Efficient Framework for Video Documentation of Bladder Lesions for Cystoscopy: A Proof-of-Concept Study

Okyaz Eminaga (✉ okyaz.eminaga@stanford.edu)

Stanford University School of Medicine

T. Jessie Ge

Stanford University School of Medicine

Eugene Shkolyar

Stanford University School of Medicine

Mark A. Laurie

Stanford University School of Medicine

Timothy J. Lee

Stanford University School of Medicine

Lukas Hockman

Stanford University School of Medicine

Xiao Jia

Stanford University School of Medicine

Lei Xing

Stanford University School of Medicine

Joseph C. Liao

Stanford University School of Medicine

Research Article

Keywords: Video-based documentation, cystoscopy documentation, cystoscopy atlas, data management, data distribution, video for cystoscopy

Posted Date: April 20th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1556709/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The identification of relevant frame sequences in an endoscopy video is labor-intensive and complicated by the poor data integrity of endoscopic information. We designed a surgeon-guided framework to efficiently manage the visual data from commonly performed bladder endoscopy (cystoscopy) videos. Screenshots of bladder lesions were captured during cystoscopy during transurethral resection of bladder tumor, then manually labeled according to case identification, date, lesion location, imaging modality, and pathology. The framework used the screenshot to search for and extract a corresponding 10-second video clip. Each video clip included a one-second space holder with a QR barcode informing the video content. The success of the framework was measured by the secondary use of these short clips and the reduction of storage volume required for video materials. From 86 cases, the framework successfully generated 249 video clips from 230 screenshots, with 14 erroneous video clips from 8 screenshots excluded. The HIPPA-compliant barcodes provided information of video contents with a 100% data completeness. A web-based educational gallery was curated with various diagnostic categories and annotated frame sequences. Compared with the unedited videos, the informative short video clips reduced the storage volume by 99.5%. In conclusion, our framework expedites the generation of visual contents with surgeon's instruction for cystoscopy and potential incorporation of video data towards applications including clinical documentation, education, and research.

Introduction

Cystoscopy is an endoscopic procedure that visualizes the interior of the urinary bladder to facilitate diagnosis and intervention [1] and is one of the most common procedures in urology [2]. Clinical documentation of cystoscopy is traditionally limited to a written description of cystoscopic findings along with pertinent clinical history and plays a vital role in optimizing patient care [3].

Contemporary cystoscopic equipment has image and video acquisition capability that can supplement written cystoscopy report. Taking screenshots of suspicious lesions during cystoscopy is commonly performed to facilitate longitudinal surveillance, assess treatment response, and inform patients of relevant findings. Storage of cystoscopy images and videos, however, is uncommon due to lack of efficient infrastructure for storage and retrieval of large videos files. In addition to visual documentation, capturing cystoscopy images and videos could be useful for curation of educational cystoscopy atlases [4, 5] and the development of computer-aided solutions for augmented endoscopic imaging [6–11]. Furthermore, visual data is a fundamental resource for developing artificial intelligence-based solutions in medical imaging [12].

The current practice for visual documentation of cystoscopy typically involves recording the entire procedure and taking still screenshots of concerning lesions as directed by the surgeon. However, data management of visual content in the clinical setting is largely unstructured and disorganized. Particularly for full length videos, the need for high-volume data storage and time-intensive searches to find and extract the relevant frame sequences lead to inefficient data management and significant amount of

information waste. Although images are easier to store and navigate, they are static and lack spatiotemporal information. The poor data integrity of clinicopathological information and annotation is another common issue in the visual documentation of cystoscopy that limits the secondary use of cystoscopic materials for educational and research purposes [13, 14].

Given the limitations to visual documentation of cystoscopy, we have developed a framework for provider-guided visual documentation. Our framework preserves spatiotemporal information and improves data integrity by linking pathological or annotation information with the corresponding visual contents in an efficient and decentralized manner.

Material And Methods

We designed a framework for visual documentation that utilized cystoscopy videos and screenshots taken during the procedure to document clinically significant lesions. Screenshots function as a communication medium between the surgeon and the framework that will curate lesion-specific video content. Figure 1 summarizes the workflow of the framework for visual documentation of cystoscopic findings.

The study was approved by the Stanford University Institutional Review Board (Protocol #29838 and #36085) and informed consent was obtained from all study subjects. Eighty-six cystoscopy videos recorded during transurethral resection of bladder tumors (TURBT) performed at the Veterans Affairs Palo Alto Medical Center by a single surgeon (JCL) between 2016 and 2018 were reviewed. Each case had a full-length video and one or more screenshots with pathological correlation. This cohort was a subset of the dataset previously published by Shkolyar et al [10].

Videos and screenshots were recorded and captured on an integrated cystoscopy system capable for both white light and blue light cystoscopy (KARL STORZ D Light C Photodynamic Diagnostic System, Tuttlingen, Germany). In some cases, multiple video files were generated in a single procedure due to surgeon preference or interruptions in video recording. A clinical research coordinator was responsible for data repository and associating the screenshots with the video records for each case. The visual content of the videos and the screenshots did not include protected health information (PHI).

Data management

Screenshots taken during cystoscopy functioned as the data exchange medium between the clinical expertise and the framework. A standardized naming scheme was devised for the filenames of the screenshots to store information on pseudonymized case identification, date, pathology, lesion location, and the lesion number, as followed:

[case id][date]_[lesion location]_[lesion id]_[lesion description]_[image modality]

We utilized well-accepted ontologies to describe the entities in the urinary bladder[15–18]. The location was defined according to the anatomical scheme of the urinary bladder proposed by the European

Association of Urology guideline for cystoscopy reporting [1]. The lesion identification (e.g., tumor1, tumor2) documents the unique lesions visible in the screenshots. The image modality was either white light cystoscopy or blue light cystoscopy, an enhanced photodynamic imaging frequently applied for a better visualization of suspicious bladder lesions [19].

We defined a hierarchical data storage for cystoscopy videos and screenshots as shown in Fig. 2 for systematized data access and scalability. Screenshots were stored in JPEG, BMP or PNG formats and videos were stored in MP4 format. Videos were taken in either 320p or 480p resolution. The data are stored on Stanford Medicine Box, an institutional HIPPA-compliant cloud-based storage solution [20] and is accessible to all study team members using a secured two-factor authentication.

Using a fault-tolerance exhaustive search algorithm written in Python, keywords in the screenshot file names were searched to extract label information for QR barcode generation and downstream data processing. For demonstration purposes, we selected keywords that describe the lesions according to the occurrences of various entities, thereby covering the most frequent benign and cancer pathologies as provided in Table 1.

Table 1

Keywords used to manually label screenshots in their file names. Stage and grade are defined according to the American Joint Committee on Cancer (AJCC)[21] and reported together in the filename.

Key words	Description
inflammation	Acute and chronic Inflammation
benign	Benign lesions, not specified
dysplasia	Dysplasia
	<i>Bladder cancer stage</i>
Tis	Tis
Ta	Ta
T1	T1
T2	T2
	<i>Grade</i>
lg	low grade
lg_focal_hg	low-grade with focal high-grade
hg	high grade
AdenocarcinomaEntericType	Adenocarcinoma (Enteric Type)
clot	Blood clot
nobiopsy	Cystoscopic lesions with no corresponding biopsies, due to either low suspicion or fulgurated by surgeon decision.
FPBLC	False-positive blue light cystoscopy (positive blue light signal indicating suspicion for cancer, but the corresponding appearance under the white light is obviously not suspicious).

Content search algorithm

We utilized Mean-Squared Error (MSE) as a metric to measure the image similarity between screenshots and video frames[22, 23]:

$$MSE = s := f(R, F) = \frac{1}{whc} \sum_{i=0}^{w-1} \sum_{j=0}^{h-1} \sum_{k=0}^{c-1} [R(i, j, k), F(i, j, k)]^2$$

1

where R is the matrix of any screenshot, F is the matrix of any frame image in a video, w is the image width, h the image height, and c the channel number of the image. Since cases can contain multiple

screenshots from a single cystoscopy video, we generated a sequence of scores for each screenshot:

$$S_n = \{s_{1,}, s_{2,}, s_{3,}, \dots, s_l\}$$

2

where S_n is the score collection for a screenshot n for a frame sequence length l . This can also be written as $S_n = \bigcup_{r=0}^l f(R_k, F_r)$, where r is the frame index. The index frame with the lowest score (x) is the frame most similar to the screenshot, and was identified using the following formula:

$$\operatorname{argmin}_{x \in S_n} f(x) := \{x \in S_n \wedge \forall s \in S_n: s \geq x\}$$

3

Next, a window size that considers the frames spanning 3 seconds before and 7 seconds after $\text{frame}_{\operatorname{argmin}}$ was defined to increase the likelihood of including different views of the lesion.

Lesion content extraction

A Python script was written incorporating the algorithm to extract frame sequences including the most similar frame to the screenshot to generate short video clips for all screenshots. After extraction of the video clip a QR barcode containing the corresponding clinical and pathological information about the lesion, the video source, and the file source of the reference image was generated and appended to the video. This QR barcode facilitates review of the video and screenshot filenames to ensure the information was extracted correctly for quality control. The resultant video clip was stored in MP4 format and named according to the filename of the screenshot.

MSE scores were calculated after normalizing all color channels (red, green, blue) between 0 and 1. Since the same lesion can be found in different frame positions in a cystoscopy video, we ran the search algorithm through the entire video and considered the frame indices with the lowest MSE scores for the frame index selection. To prevent the selection of frames next to the index frame and to mitigate the likelihood of having video clips with duplicated visual content, we defined a lockout interval of 3 seconds before and 10 after the index frame. The script was configured to process 8 cases at the same time using parallel computing on an Intel processor with 9 cores and 32 GB RAM.

Utility

Quality control

A web portal was designed to allow expert urology reviewers to assess the content quality of the 10-second video clips securely using two-factor authentication. For each video clip, the reviewer obtained information about the original screenshot using the QR barcode [24] and verified whether the lesion seen in the screenshot was in the video clip. For quality control, the reviewer checked the video contents for its

lesion presentation in different angles and the correct lesion selection. Finally, data completeness for pathology information in the barcodes for all video clips was assessed.

Application

As an application of our framework, we examined the feasibility of creating a cystoscopy atlas by presenting the video clips arranged by pathological diagnosis. To protect PHI and preserve HIPAA compliance, the QR barcode containing the source file information was replaced with a QR barcode including only the pathology information.

Additionally, we examined the feasibility of using the video clips to store annotation information and to reveal the label distribution for each frame in these video clips. The 10-second video clips were each manually annotated to denote the range of frames meeting the description of the keyword considered by the current study. The frames were annotated according to the keywords described in Table 1 using a customized annotation tool (<https://github.com/oeminaga/VideoAnnotator>). Any frame that did not meet any of the description of the keywords was considered as background. Annotation data was then stored in json format in the barcode of the video clips in addition to the pathology information.

Results

Our framework was successfully implemented without exposing PHI and in compliance of HIPAA regulations. A total of 238 screenshots were available from 86 cystoscopy procedures on 78 patients; the full-length videos were processed (i.e., search for frames close to screenshots) at a speed of 43 ± 20 frames per second (FPS), depending on the number of the screenshots to process per video. For instance, a 51-minutes raw video at 240p and 30 FPS containing 2 lesions (screenshots) took 62.14 minutes to process by a search speed of 24.6 FPS.

Our framework extracted the video clips from the screenshots with an error rate of 3.3% (8/238 of screenshots). Through our QC process, 14 video clips derived from 8 screenshots were excluded for not accurately reflecting the content of the screenshot. In depth review of these 14 clips demonstrated that they originated from screenshots labeled with keywords 'Ta,' regardless of grade, or 'nobiopsy'. These screenshots were dark images that were taken under blue light cystoscopy, which has a darker background, or blurry images taken with the cystoscope positioned too close to the lesion. The resultant erroneous video clips for dark images occurred when a dark screenshot had an MSE similar to a video frame with the cystoscope located outside of the patient body. Blurry screenshots tended to have an MSE similar to other frames when the camera was out of focus.

The resulting 249 video clips from 230 screenshots for 143 lesions represent a wide range of cancerous and benign lesions of the bladder (Table 2). Additional 19 video clips were extracted because the additional video clips were originated from the distant frame sequences containing frames similar to the corresponding screenshot. The 230 screenshots were all identified within their corresponding videos, thereby achieving a data completeness of 100% for lesion pathology in these videos.

We associated the label information with the corresponding video clips and build the web portal for quality control as shown in Fig. 3 and **Supplementary file 1**. The internal web portal provided the web link to each clip for external sharing via email or barcode format.

Table 2
Distribution of key words for short video clips generated from screenshots.

Keyword's description	Frequency, n (%)
Benign	
Acute and chronic Inflammation	17 (6.8)
Benign lesions, not specified	39 (15.7)
Dysplasia	2 (0.8)
Bladder cancer stage	
Tis	11 (4.4)
Ta LG (low grade)	91 (36.5)
Ta LG with focal high-grade	4 (1.6)
Ta HG (high grade)	40 (16.1)
T1 HG	21 (8.4)
T2 HG	11 (4.4)
Non-urothelial Malignancies	2 (0.8)
Adenocarcinoma (Enteric Type)	
Other	1 (0.4)
Blood clot	
Lesions with no biopsy	9 (3.6)
False-positive blue light cystoscopy	1 (0.4)

For all video clips, we labeled the frame sequence for lesion and the information integrated into the video was enough to curate a cystoscopic atlas. Figure 4 shows the distribution of the cystoscopy frames containing a lesion of interest (LOI) and background over the frames of the clip videos, demonstrating that the majority of the labeled frames contain the LOI.

Moreover, we were able to use the video clip as storage for annotation data using the barcodes of the space holder as shown in Fig. 5, so that the video clip incorporated images and the corresponding image

labels ready for data distribution (An example video clip with the barcode containing the annotation data is provided in the supplementary).

Using our framework to store the lesion content of a cystoscopy video as short video clips reduced the data storage requirements by 99.5%. The full-length raw videos (cumulative length: 1,821 minutes) totaled 47.3 GB in size, whereas 249 short video clips (cumulative length: 45.7 minutes) totaled only 247.25 MB in size. A single 11-second video clip occupied 0.51 ± 0.26 MB for 240p or 1.82 ± 1.62 MB for 480p, whereas a full-length video occupied 703 ± 468 MB for 240p or $1,416 \pm 734$ MB for 480p.

Discussion

This study illustrates a semi-automated framework to improve the visual documentation of cystoscopic findings. We showed that screenshots of bladder lesions routinely taken by surgeons can be used by our framework to automatically generate concise video clips of pertinent cystoscopic findings.

Reducing data storage requirements becomes even more of a concern as cystoscopic videos are recorded at increasingly higher resolution, as a 1080p video is approximately 4 times the size of a 480p video. More importantly, this framework allows the surgeon to review cystoscopic videos in a shorter amount of time, by focusing data on the lesions of interest. Cystoscopic videos can often contain a large amount of clinically irrelevant information (e.g., time that the instrument is out of the body for cleaning or equipment changes), or negative cystoscopic findings (e.g., no lesion of interest in frame). This framework can help integrate video data, which is typically cumbersome to navigate, into routine clinical workflow to augment the amount of cystoscopic data at hand or use them for physician-patient communication.

The current framework of cystoscopy documentation is text-based (within the surgery or procedure report and the pathology report). However, textual description is limited and influenced by the perception bias of the surgeon, as supported by previous studies on the limits of human perception in describing image content [25, 26]. Additionally, it has been shown that non-visual categorical word descriptions of object shape (e.g., 'papillary', 'sessile') influence how a person visually processes the visual medium [25, 27]. Accordingly, to reduce the potential bias on visual processing, we labelled the video content with objective information verified by pathology (e.g., pathologic grade and stage) or obvious visual appearance (e.g., blood clot), while we omitted subjective descriptors of the lesion appearance (e.g., papillary, flat, texture descriptions) in the video labelling.

The proposed framework further facilitates data decentralization and preservation of data integrity for documentation and data sharing by storing the label information inside the video clip as a QR code. Studies in the medical imaging domain generally use metadata sheets to label and associate the annotation information with the corresponding visual contents. However, this approach inherits a higher risk of losing data integrity when the metadata sheets are no longer available. In contrast, our solution manages each data point (video clip) as an independent storage unit to support decentralized data management and eliminate the need for metadata sheets to utilize a data point. Decentralized data

facilitates better framework scalability and data distribution compared to rigid and centralized data collection, be integral part of de-centralized data solution [28].

Prior studies of deep learning for cystoscopy images utilized still images, either captured directly by surgeons or extracted from cystoscopy videos to curate a cystoscopy atlas [6–11]. In these studies, the collection of still images for cystoscopy data followed a predefined selection criteria of frames (e.g., a single frame with the best representation of the lesion, every 10th frame, the first 30 frames) to reduce data dimensionality and redundancy; however, this approach considers cystoscopy data to be static, which is not reflective of the real-world condition and the dynamic streaming content, and therefore limits the ability to generate generalizable and relevant data.

There are certain limitations that should be considered. First, the implementation of the framework is dependent on the existing technical infrastructure and the workflow for data management, the availability of videos and screenshots and therefore requires to be adjusted according to the existing infrastructure. We applied retrospective data that may reflect the limitations of any retrospective study [29] including the selection bias. Nevertheless, we examined our framework for screenshots representing a rare histology type of bladder cancer, the putative precursor of bladder cancer (dysplasia), and other clinical situations (i.e., the existence of clot in the bladder, lesions not biopsied) for generalization. Our work represents a proof-of-concept study and the framework concept as well as the tools developed as part of this study are prototypes and therefore subject to continuous improvement. Finally, we chose the MP4 format for the video clips, as this can run on all multimedia players and operating systems, while medical imaging systems use DICOM format as their standard. Nevertheless, it is simple to convert video clips to DICOM format [30, 31].

Overall, our work provides insight into the challenges of video-based clinical documentation and proposes a solution to manage the visual documentation of cystoscopy. Our approach can be used to curate cystoscopic atlas and datasets for research and clinical documentation. Future works will focus on developing a video player tool that reads the annotation data from the QR barcode and overlays it for clip presentation and merging artificial intelligence and the proposed documentation framework to improve the overall documentation workflow for cystoscopy.

Conclusion

The current work introduces and validates as proof-of-concept a framework for surgeon-instructed video documentation of cystoscopy which maintains spatiotemporal information and supports de-centralized data management.

Declarations

Acknowledgement

We thank Kathy Mach for critical review of the manuscript. The research was supported in part by National Institutes of Health R01 CA260426 to JCL.

Conflict of interest statement

No conflict of interests.

Role of funding source

The funding source has no role in the study design and execution.

Ethical Approval

The study was approved by the Stanford University Institutional Review Board (Protocol #29838 and #36085).

Informed Consent

informed consent was obtained from all study subjects.

Authors' Contributions

O.E. and J.C.L. designed the study. Data collected by O.E., E.S., T.J.L, and X.J.; O.E. wrote the script and analyzed the data. O.E. and M.A.L. implemented the framework. O.E., J.G., L.H., and J.C.L. performed the quality control, O.E, J.G., J.C.L, and E.S. wrote the main manuscript text and O.E. prepared figures and tables. The study was supervised by L.X. and J.C.L. All authors reviewed the manuscript.

References

1. Witjes JA, Bruins HM, Cathomas R, Comperat EM, Cowan NC, Gakis G, et al. European Association of Urology Guidelines on Muscle-invasive and Metastatic Bladder Cancer: Summary of the 2020 Guidelines. *Eur Urol.* 2021;79(1):82-104. doi: 10.1016/j.eururo.2020.03.055.
2. Engelsjerd JS, Deibert CM. Cystoscopy. *StatPearls [Internet].* 2020.
3. Douglas-Moore J, Lewis R, Patrick J. The importance of clinical documentation. *The Bulletin of the Royal College of Surgeons of England.* 2015.
4. Lenherr SM, Crosby EC, Cameron AP. Cystoscopic findings: a video tutorial. *Int Urogynecol J.* 2015;26(6):921-3. doi: 10.1007/s00192-014-2614-4.

5. Ronstrom C, Lai HH. Presenting an atlas of Hunner lesions in interstitial cystitis which can be identified with office cystoscopy. *Neurourol Urodyn*. 2020;39(8):2394-400. doi: 10.1002/nau.24500.
6. Eminaga O, Eminaga N, Semjonow A, Breil B. Diagnostic Classification of Cystoscopic Images Using Deep Convolutional Neural Networks. *JCO Clin Cancer Inform*. 2018;2:1-8. doi: 10.1200/CCI.17.00126.
7. Ikeda A, Nosato H, Kochi Y, Kojima T, Kawai K, Sakanashi H, et al. Support System of Cystoscopic Diagnosis for Bladder Cancer Based on Artificial Intelligence. *J Endourol*. 2020;34(3):352-8. doi: 10.1089/end.2019.0509.
8. Ikeda A, Nosato H, Kochi Y, Negoro H, Kojima T, Sakanashi H, et al. Cystoscopic Imaging for Bladder Cancer Detection Based on Stepwise Organic Transfer Learning with a Pretrained Convolutional Neural Network. *J Endourol*. 2021;35(7):1030-5. doi: 10.1089/end.2020.0919.
9. Mutaguchi J, Morooka KI, Kobayashi S, Umehara A, Miyauchi S, Kinoshita F, et al. Artificial intelligence for segmentation of bladder tumor cystoscopic images performed by U-Net with dilated convolution. *J Endourol*. 2022. doi: 10.1089/end.2021.0483.
10. Shkolyar E, Jia X, Chang TC, Trivedi D, Mach KE, Meng MQ, et al. Augmented Bladder Tumor Detection Using Deep Learning. *Eur Urol*. 2019;76(6):714-8. doi: 10.1016/j.eururo.2019.08.032.
11. Suarez-Ibarrola R, Kriegmair M, Waldbillig F, Grune B, Negassi M, Parupalli U, et al. A novel endoimaging system for endoscopic 3D reconstruction in bladder cancer patients. *Minim Invasive Ther Allied Technol*. 2020:1-8. doi: 10.1080/13645706.2020.1761833.
12. O'Leary DE. Artificial intelligence and big data. *IEEE intelligent systems*. 2013;28(2):96-9.
13. Danciu I, Cowan JD, Basford M, Wang X, Saip A, Osgood S, et al. Secondary use of clinical data: the Vanderbilt approach. *Journal of biomedical informatics*. 2014;52:28-35.
14. Mingers J. The paucity of multimethod research: a review of the information systems literature. *Information systems journal*. 2003;13(3):233-49.
15. Coleman JF, Hansel DE. Benign Diseases of the Bladder. *Surg Pathol Clin*. 2008;1(1):129-58. doi: 10.1016/j.path.2008.07.001.
16. Salvadores M, Alexander PR, Musen MA, Noy NF. BioPortal as a Dataset of Linked Biomedical Ontologies and Terminologies in RDF. *Semant Web*. 2013;4(3):277-84.
17. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol*. 2010;17(6):1471-4. doi: 10.1245/s10434-010-0985-4.
18. Comperat EM, Burger M, Gontero P, Mostafid AH, Palou J, Roupret M, et al. Grading of Urothelial Carcinoma and The New "World Health Organisation Classification of Tumours of the Urinary System and Male Genital Organs 2016". *Eur Urol Focus*. 2019;5(3):457-66. doi: 10.1016/j.euf.2018.01.003.
19. Daneshmand S, Bazargani ST, Bivalacqua TJ, Holzbeierlein JM, Willard B, Taylor JM, et al. Blue light cystoscopy for the diagnosis of bladder cancer: Results from the US prospective multicenter registry. *Urol Oncol*. 2018;36(8):361 e1- e6. doi: 10.1016/j.urolonc.2018.04.013.

20. Patridge EF, Bardyn TP. Research Electronic Data Capture (REDCap). *J Med Libr Assoc.* 2018;106(1):142-4. doi: 10.5195/jmla.2018.319.
21. Amin MB, Greene FL, Edge SB, Compton CC, Gershewald JE, Brookland RK, et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J Clin.* 2017;67(2):93-9. doi: 10.3322/caac.21388.
22. Palubinskas G. Image similarity/distance measures: what is really behind MSE and SSIM? *International Journal of Image and Data Fusion.* 2017;8(1):32-53.
23. Palubinskas G. Mystery behind similarity measures MSE and SSIM. 2014 IEEE International Conference on Image Processing (ICIP): IEEE; 2014. p. 575-9.
24. Stark MM. *QR Codes: The Technical Guide.* A. K. Peters, Ltd.; 2013.
25. Carbon CC. Understanding human perception by human-made illusions. *Front Hum Neurosci.* 2014;8:566. doi: 10.3389/fnhum.2014.00566.
26. Pearson J, Naselaris T, Holmes EA, Kosslyn SM. Mental Imagery: Functional Mechanisms and Clinical Applications. *Trends Cogn Sci.* 2015;19(10):590-602. doi: 10.1016/j.tics.2015.08.003.
27. Noorman S, Neville DA, Simanova I. Words affect visual perception by activating object shape representations. *Sci Rep.* 2018;8(1):14156. doi: 10.1038/s41598-018-32483-2.
28. Ayoade G, Karande V, Khan L, Hamlen K. Decentralized IoT data management using blockchain and trusted execution environment. 2018 IEEE International Conference on Information Reuse and Integration (IRI): IEEE; 2018. p. 15-22.
29. Norvell DC. Study types and bias-Don't judge a study by the abstract's conclusion alone. *Evid Based Spine Care J.* 2010;1(2):7-10. doi: 10.1055/s-0028-1100908.
30. Liu F, Hernandez-Cabronero M, Sanchez V, Marcellin MW, Bilgin A. The current role of image compression standards in medical imaging. *Information.* 2017;8(4):131.
31. Pianykh OS. *Digital imaging and communications in medicine (DICOM): a practical introduction and survival guide.* Springer; 2012.

Figures

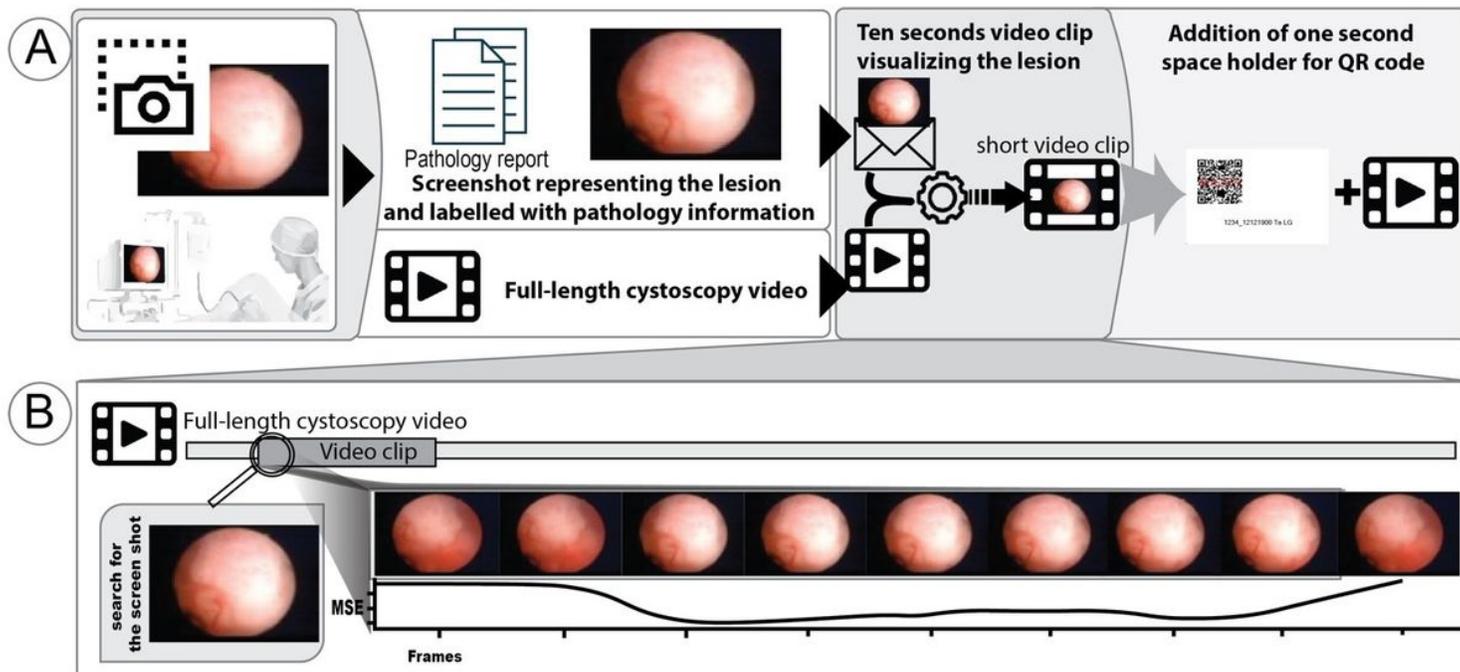


Figure 1

Workflow of the framework for visual documentation of cystoscopy. (A) A screenshot of a lesion of interest taken during cystoscopy. After the procedure, the full-length cystoscopy video is saved, and the screenshot file is labeled with the corresponding location of the lesion in the bladder. Once the pathological diagnosis is available, the information is added to the screenshot filename. (B) The search algorithm is applied to the videos to generate short video clips containing the identified lesion by estimating the mean square error (MSE). The MSE measures the similarity between the corresponding screenshot and the video frames. Then, we extracted the frame sequence 3 seconds before and 7 seconds after the frame with the lowest MSE that is most similar to the corresponding screenshot. Finally, a one-second space-holder containing a QR code is added to the beginning of each video clip. The QR code contains pathologic information, frame annotation, and the screenshot and video resources of these video clips for quality control.



Figure 2

Hierarchical file storage for screenshots and videos. PHI is blacked out. A root folder of the case was named with the case identification number and the date of transurethral resection of bladder tumors. In each case folder, we have one folder “Images” to store screenshots and one folder “Video(s)” to keep video records. The pathology and surgery reports are optionally stored in the case folder. The required information was extracted from the file name of the screenshots using the keyword search.

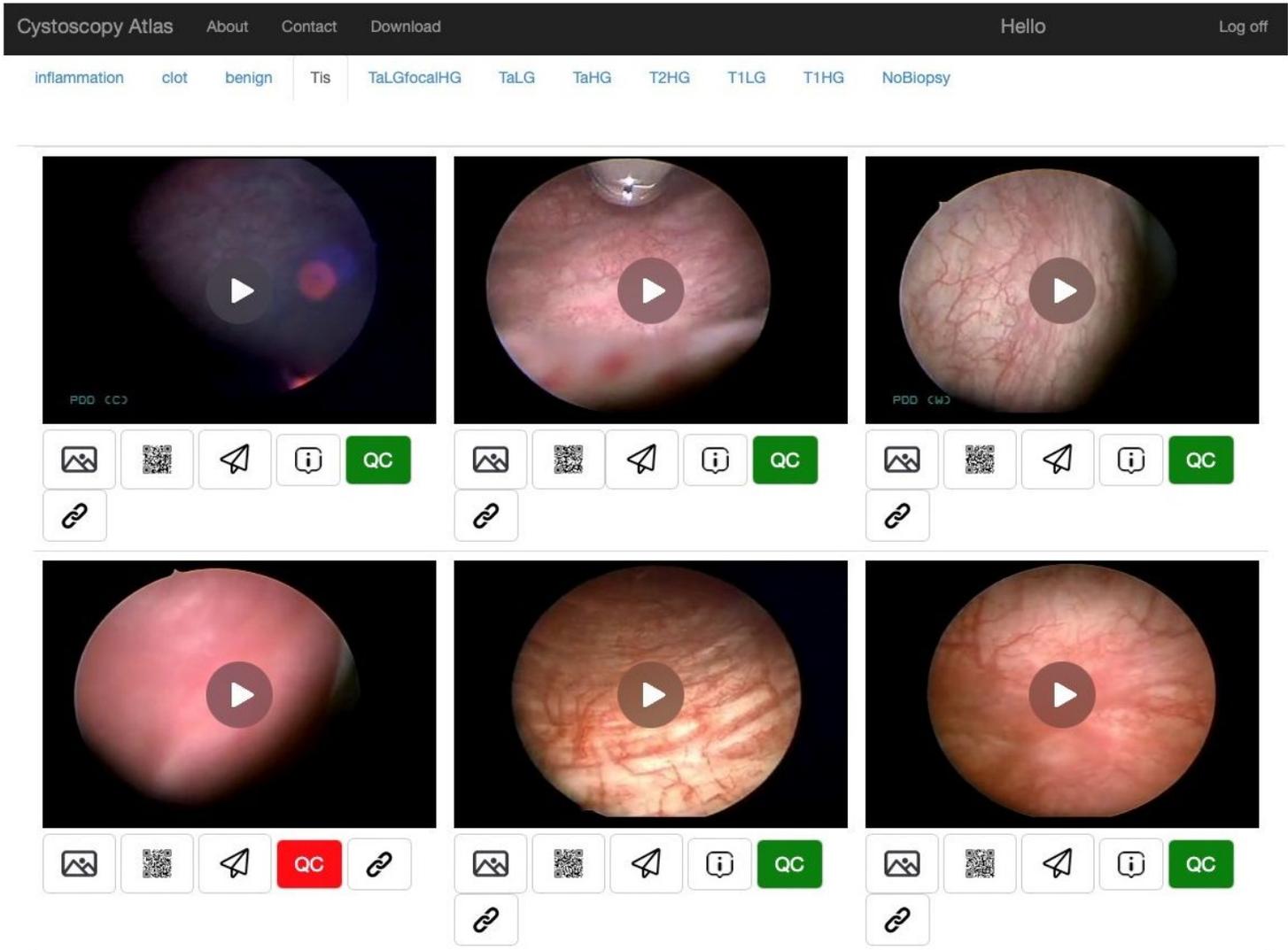


Figure 3

Web portal created to perform quality control of generated video clips. Video clips are categorized according to their key words. Here, video clips with key word 'Tis' are shown. From left to right, the buttons under each video clip are options to: a) preview the index frame of the clip; b) view a QR barcode linking to the video clip c) share the video clip via email, d) access external contents, e) toggle quality control, where green means that the video clip passed quality control, while red means that the video clip failed quality control; f) configure the links to the external contents. Links to additional information were not added to video clips that failed QC (See the first left image of the second row).

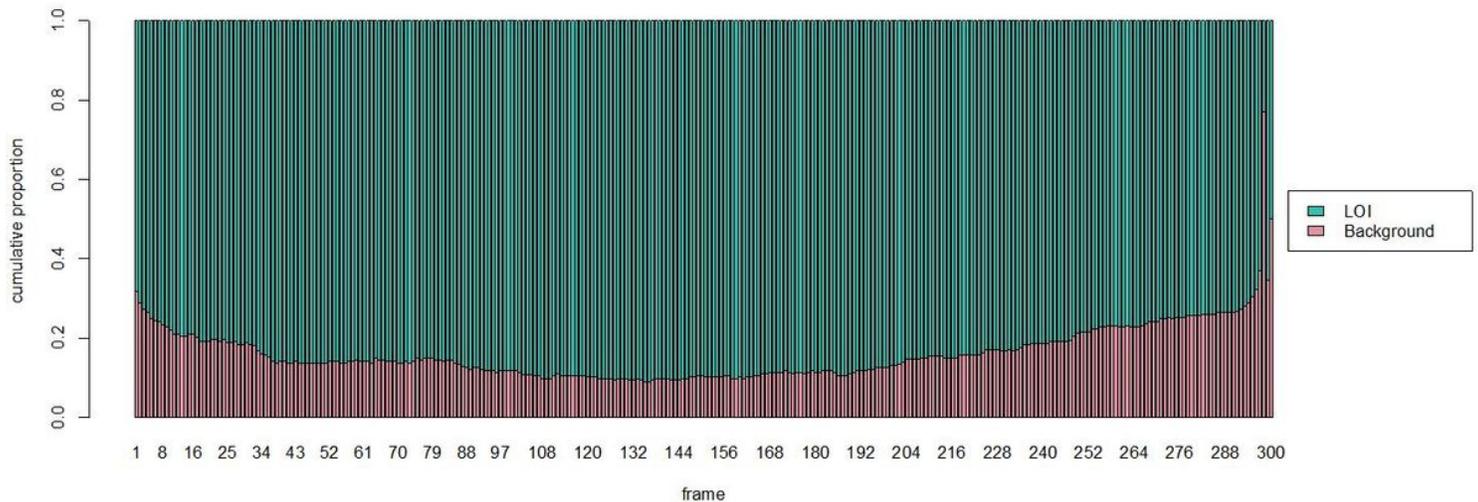


Figure 4

shows the overall distribution of lesions of interest (LOI) in 300 frames on 249 video clips after excluding the first second space holder for barcode. This figure suggest that 10 seconds are sufficient to document the lesion of the interest given the overall distribution. The frames for all clips were labelled according to the key words (Table 1) using the customized annotation tool. Background means the frames not meeting any of the categories of Table 1. One second corresponds to 30 frames. The change in the background distribution over the frames indicates the variation of video contents over the time, thereby reflecting the dynamic of the camera view during the 10 seconds.

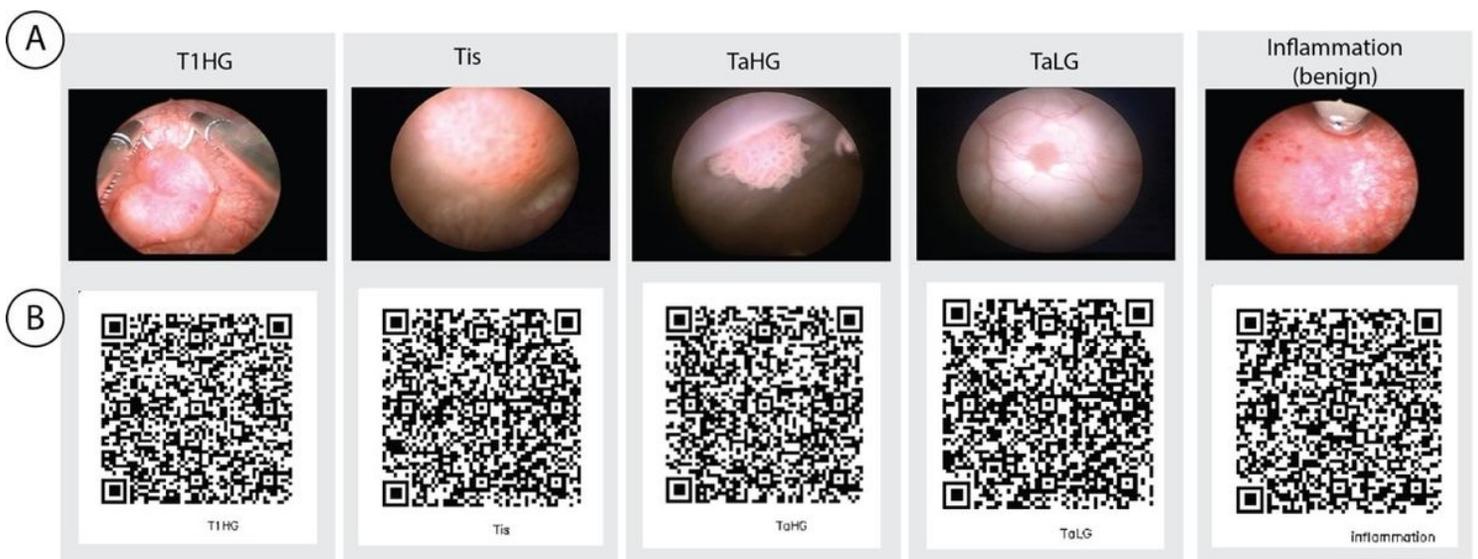


Figure 5

Representative examples of (A) lesions representing the major categories of bladder cancer and (B) barcodes including the corresponding annotation data for each lesion generated by the framework. The

barcodes are also displayed in the first second of each video clip to facilitate decentralized data annotation for data sharing that stores time interval of lesion appearance in the short video clip (attributes: StartTime and EndTime; time unit: seconds). Please use the camera function of your smart phone to unravel the annotation content of the barcode. The annotation data are stored in json format and generated by our custom annotation tool (A source of the annotation tool is available in <https://github.com/oeminaga/VideoAnnotator>). The supplementary section provides an example video that includes the annotation data in the barcode.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryfile1.mp4](#)