



HAL
open science

Utilisation de Tags pour le résumé des vidéos chirurgicales

Alexandre Haddad, Gilles Bailly, Rabah Taouachi, Olivier Choussy, Ignacio Avellino

► **To cite this version:**

Alexandre Haddad, Gilles Bailly, Rabah Taouachi, Olivier Choussy, Ignacio Avellino. Utilisation de Tags pour le résumé des vidéos chirurgicales. IHM'24 - 35e Conférence Internationale Francophone sur l'Interaction Humain-Machine, AFIHM; Sorbonne Université, Mar 2024, Paris, France. hal-04487266v1

HAL Id: hal-04487266

<https://hal.science/hal-04487266v1>

Submitted on 3 Mar 2024 (v1), last revised 21 Mar 2024 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Utilisation de Tags pour le résumé des vidéos chirurgicales

Understanding how Tagging Supports Video Summarization

ALEXANDRE HADDAD, Institut Curie - Service de chirurgie cervico-faciale, Sorbonne Université, Institut des Systèmes Intelligents et de Robotique, ISIR, France

RABAH TAOUACHI, Institut Curie - Service de chirurgie cervico-faciale, France

OLIVIER CHOussy, Institut Curie - Service de chirurgie cervico-faciale, Université Paris Cité, France

IGNACIO AVELLINO, Sorbonne Université, CNRS, INSERM, Institut des Systèmes Intelligents et de Robotique, ISIR, France

Les vidéos chirurgicales sont utiles pour des usages pédagogiques ou scientifiques. Pour autant, elles sont difficiles à produire. Ce travail explore une nouvelle approche de montage de vidéos chirurgicales, consistant à ajouter des *tags* pour indiquer les moments à conserver dans une vidéo. Nous étudions cette approche à l'aide de trois ateliers où des paires de chirurgiens-internes doivent réaliser un résumé vidéo en utilisant un prototype de recherche. Les résultats montrent des dynamiques de collaboration différentes en fonction du niveau d'expérience des internes, les limites de la notion d'addition de séquences par les *tags* dans les pratiques de montage de vidéos chirurgicales et la flexibilité permise par l'utilisation de *tags* dont la possibilité de créer des résumés différents à partir d'une même vidéo. Il montre une bonne acceptation du concept par les participants. Cette étude suggère plusieurs pistes d'études pour prouver son efficacité.

Surgical videos serve various purposes, including educational and scientific applications, however, there are several obstacles to their production. This work explores a novel approach to surgical video editing, introducing the concepts of sequence addition through tags and collaborative work. We conducted three workshops where surgeons edited their own videos collaboratively with residents, using a software prototype. The results show that the collaboration dynamics vary based on the interns' level of experience, the limits of the tagged sequence addition concept in surgical video editing practices, and the flexibility to create multiple summaries from the same video. These results demonstrate a positive acceptance of the concept and suggest various avenues for further study to establish its effectiveness.

CCS Concepts: • **Human-centered computing** → **Collaborative and social computing; Interaction design process and methods.**

Mots Clés et Phrases Supplémentaires: chirurgie, résumés vidéo, chirurgie endoscopique, montage vidéo collaborative, tags vidéo

Additional Key Words and Phrases: surgery, video summarization, endoscopic video, collaborative video editing, video tags

Reference:

Alexandre Haddad, Rabah Taouachi, Olivier Choussy, and Ignacio Avellino. 2024. Utilisation de Tags pour le résumé des vidéos chirurgicales. *IHM '24 : Actes étendus de la 35^e conférence Francophone sur l'Interaction Humain-Machine, March 25–29, 2024, Paris, France*

1 INTRODUCTION

Les vidéos sont une ressource utile dans le domaine de la chirurgie, elles sont notamment utilisées pour apprendre, discuter des cas, diffuser des connaissances scientifique [4, 5]. Elles sont particulièrement intéressantes car de nombreuses procédures sont déjà vidéo-assistées, par exemple en chirurgie laparoscopique ou endonasale où le chirurgien opère grâce à une optique connectée à une caméra, ainsi, la capture de vidéos chirurgicales brutes est simple [27]. Néanmoins,

pour que les vidéos chirurgicales soient utiles, elles doivent être résumées, car elles peuvent s'étendre sur plusieurs heures et inclure des gestes répétitifs (p. ex., sutures ou dissection) ainsi que des séquences inutilisables (p. ex., images floues ou surexposées). Des études empiriques ont étudié les pratiques actuelles sur le résumé des vidéos chirurgicales : le montage est réalisé par des experts du domaine, des chirurgiens, de façon manuelle et collaborative avec des logiciels grand public du commerce (p. ex., iMovie) [5]. Bien que des méthodes automatiques de traitement d'image existent, par exemple pour identifier des segments flous [28], elles ne permettent pas aujourd'hui de produire des résumés vidéos qui répondent aux attentes des chirurgiens [1].

Avellino et al. [5] proposent l'utilisation de *tags* pour surmonter des défis qu'ils identifient liés aux outils d'édition classiques pour le résumé manuel, notamment 1) l'approche dit *soustractive*, où l'utilisateur doit éliminer des segments pour résumer la vidéo, et 2) le manque de support à la collaboration. Le chirurgien peut créer des *tags* durant l'enregistrement de la vidéo, qui représentent les segments à conserver sous une approche *additive*, et qui peuvent fournir un support au travail d'articulation nécessaire à la collaboration. Cependant, ces potentiels bénéfiques n'ont pas été démontrés.

Notre objectif à long terme, est de permettre aux chirurgiens de résumer des vidéos chirurgicales, en étiquetant les étapes importantes et les moments clés. Dans cet article, nous faisons un premier pas vers cet objectif, en étudiant :

- Q1 : Comment une approche additive reposant sur l'utilisation de *tags* permet de réaliser une vidéo chirurgicale ?
- Q2 : Comment les *tags* sont utilisés lors de la collaboration ?

Pour répondre à ces questions, nous avons mené trois ateliers avec chacun un binôme composé d'un chirurgien expert ORL (oto-rhino-laryngologie) et d'un interne. Le binôme a utilisé un prototype qui permet d'utiliser des *tags* pour résumer des vidéos chirurgicales. Cette évaluation montre que l'approche répond aux besoins des chirurgiens, avec certaines limitations. La sélection de segments est utilisée pour délimiter les étapes intéressantes de la chirurgie, mais ensuite l'étude montre le besoin d'une approche soustractive pour le travail d'affinage. Les *tags* ont également été identifiés comme bénéfiques pour modifier rapidement un résumé, et créer des nouveaux résumés pour un autre objectif. Concernant la collaboration, l'étude montre que le niveau d'expérience de l'interne influence la division des rôles. Enfin, les participants ont donné de nombreuses suggestions sur l'application future au bloc opératoire. Ces résultats informent le design d'une étude subséquente, pour comparer les deux approches de façon contrôlée.

2 TRAVAUX ANTÉRIEURS

2.1 Pratiques autour des vidéos chirurgicales

Le principe de chirurgie *mini invasive* consiste à insérer les instruments chirurgicaux, ainsi qu'une caméra appelée endoscope, dans le patient, au travers de petites incisions. La chirurgie s'effectue ensuite en observant un flux vidéo ininterrompu sur un écran 2D. Dans cette configuration, l'enregistrement vidéo est trivial, car il suffit d'appuyer sur un bouton pour enregistrer. Cependant, les vidéos produites, font la durée d'une opération, c'est-à-dire, souvent plus d'une heure et parfois dix heures ou plus [27]. Pour être exploitables, les vidéos enregistrées doivent être montées, pour n'inclure que les informations pertinentes et réduites à des durées généralement inférieures à vingt minutes [10]. Cependant, de nombreuses vidéos ne sont pas montées (et donc non utilisées) à cause de l'effort nécessaire à leur production. Ceci a un impact sur la quantité de contenus utiles, notamment pour l'apprentissage de la chirurgie [5].

2.2 Résumé vidéo

Les approches de montage vidéo peuvent être classées en fonction de leur degré d'automatisation :

2.2.1 Montage manuel. La méthode manuelle est la plus utilisée dans le domaine de la chirurgie. Avellino et al. [5] montrent que la création d'un résumé vidéo se déroule en plusieurs étapes, où les chirurgiens identifient d'abord des séquences d'intérêt de façon « grossière » en éliminant les séquences autour. Puis ils affinent, en éliminant les séquences inutiles ou redondantes à l'intérieur du résultat obtenu [10]. Ce processus est réalisé de façon collaborative : les chirurgiens experts délèguent des étapes à des chirurgiens junior ou des internes [2], qui peuvent avoir ou non participé à l'intervention chirurgicale. Ceci nécessite souvent plusieurs itérations entre chirurgiens pour que le résultat final soit validé. Cette approche est chronophage, le temps d'édition pouvant durer plus longtemps que la vidéo d'origine. De plus, le délai — parfois plusieurs semaines — entre l'opération chirurgicale et le montage de la vidéo est problématique. Le chirurgien doit passer du temps à revoir la vidéo d'origine pour se rappeler des moments importants oubliés [5]. Une autre difficulté est liée aux outils d'édition, tels qu'iMovie (Apple), peu adaptés aux chirurgiens [5]. Ces outils de traitement de vidéos décrits notamment par Pea and Hoffert [31] s'appuient sur une approche *soustractive*, où le chirurgien doit identifier toutes les séquences à éliminer, qui sont majoritaires dans la vidéo d'origine, plutôt que celles qu'il doit conserver. De plus, les séquences éliminées sont difficiles à récupérer. Enfin, ces outils manquent de support pour éditer une vidéo de manière collaborative.

2.2.2 Montage (semi-)automatique. Les méthodes automatiques s'appuient sur des informations externes ou internes pour réaliser des résumés vidéo. Les méthodes qui s'appuient sur des informations externes, telles que la piste audio [21], les annotations [11] ou les données de localisation, ne sont pas pertinentes en chirurgie car la vidéo ne contient pas ces informations. Les méthodes qui s'appuient sur des informations internes, telles que la détection de gestes [20], de couleurs [9] ou d'activités de mouvement [6] ne sont pas non plus adéquates actuellement, car la vidéo chirurgicale est une prise de vue unique avec des nuances de rouge et n'inclue pas de changements de prise de vue détectables (comme dans les films), ou de différences de couleur. Certaines solutions, permettent de détecter des images inutilisables car surexposées ou floues [28], des images où la caméra est occultée [23] ou de flouter automatiquement les visages [17]. Certaines solutions peuvent aussi identifier des séquences intéressantes en reconnaissant dans la vidéo des gestes pertinents, ou des instruments [24, 25]. Des solutions, telles que *Surch*, [16] détectent également des étapes de chirurgie dans une vidéo, ce qui pourrait faciliter le montage. D'autres solutions laissent l'utilisateur étiqueter un petit ensemble de segments afin d'entraîner un système à produire un montage [8]. Cependant ces systèmes n'ont généralement pas été évalués, à l'exception de *Magisto* [1], une IA qui crée des résumés de vidéos, le résultat n'est toutefois probant que dans 60% des cas. En dehors de certains contextes spécifiques, p. ex., vidéo de caméra de surveillance, le résumé automatique de vidéo reste un défi, en particulier pour le cas des vidéos chirurgicales. Ceci explique probablement pourquoi ces méthodes ne sont pas intégrées dans les pratiques courantes, et l'édition laissée au soin d'un chirurgien.

2.3 Étiquetage sémantique de vidéos

L'utilisation de balises, ou *tags*, dans des vidéos a été déjà exploré dans le domaine de l'IHM. Par exemple, Mackay [26] a étudié l'ajout de *tags* par un chercheur lors de l'enregistrement de passations d'expérience, pour naviguer dans les moments clés, Huron et al. [13] l'usage de *tags* participatifs via Twitter pour faciliter l'analyse de vidéos de conférences scientifiques, et ?] a étudié l'annotation vidéo dans le contexte sportif, révélant que les spectateurs étaient capables d'annoter de manière précise et détaillée des informations clés. En chirurgie particulièrement, les captures d'écran [22], ou la voix ont été proposés comme des moyens d'annoter, ou *tagguer*, les segments vidéo [11]. Même si l'utilisation d'annotations permet donc d'identifier efficacement les segments importants d'une vidéo, dans les travaux cités, ils

Workshop	Participant	Rôle	Expérience	Sexe
W1	P1	Chirurgien senior	Expert	Masculin
	P2	Interne	Moyenne	Féminin
W2	P3	Chirurgien senior	Expert	Masculin
	P4	Interne	Faible	Féminin
W3	P5	Chirurgien senior	Expert	Masculin
	P6	Interne	Élevée	Féminin

TABLE 1 – Participants

proviennent d’une source externe (p. ex., un chercheur ou le public) et pas la personne réalisant l’action. De façon plus importante, ces travaux n’ont pas exploré l’utilité des *tags* pour le montage vidéo.

Finalement, peu de travaux ont exploré la collaboration dans la création de résumés vidéo. *VidCrit* [29] est conçu pour qu’une personne qui annote des vidéos puisse interagir avec une personne qui donne un retour sur les annotations. Malgré le développement technique de fonctionnalités collaboratives, ce travail n’étudie pas l’usage des annotations pour la collaboration. *Polemictweet* [13], repose sur la collaboration des spectateurs de conférences scientifiques mais les messages sont à sens unique, limitant finalement l’aspect purement collaboratif. Le projet *DIVER* [30] propose une interface permettant d’annoter des vidéos de manière collaborative, associant des *tags* à une zone précise sur la vidéo. Bien que cela permette de créer des liste d’extrait étiquetés, ce projet n’a pas pour but principal le montage vidéo et garde une fonction analytique.

En résumé, l’approche la plus couramment utilisée pour le résumé de vidéos reste le montage manuel, malgré les diverses difficultés mentionnées. Bien que l’annotation de vidéos a été étudiée, ces études ne se concentrent pas sur la réalisation de résumés vidéo ou la collaboration.

3 MÉTHODE

Ce travail vise à comprendre comment l’utilisation de *tags* contribue à la réalisation collaborative de résumés vidéo. Nous avons développé un prototype qui permet aux chirurgiens de marquer des segments vidéo pour résumer la vidéo de manière collaborative. Nous l’avons testé durant des ateliers avec des chirurgiens experts, ou nous nous intéressons à mieux comprendre les besoins et les usages (la qualité des résumés produits n’était pas l’objet de cette étude). Le format atelier nous semblait approprié, au vu de la difficulté d’introduire une technologie nouvelle dans un environnement complexe comme un bloc opératoire, sans présenter le concept aux utilisateurs dans un cadre neutre au préalable. Par souci de transparence, nous avons pré-enregistré le protocole de l’atelier et les questionnaires : <https://osf.io/uh7bq>.

3.1 Participants

6 chirurgiens ont participé à 3 ateliers. Chaque atelier comprenait un expert, un chirurgien senior ayant plus de 5 ans d’expérience post-internat et un novice, un interne en chirurgie. Les participants ont été recrutés grâce au réseau personnel des auteurs et par le bouche à oreille. Le rôle et le numéro des participants sont présentés dans le tableau 1.

3.2 Prototype logiciel

Nous avons nommé le prototype « peroscope » (pour endoscopie peropératoire). Il s’agit d’une interface web fonctionnant sur les plate-formes de recherche expérimentale Webstrates [18] et Videostrates [19]. Le prototype permet

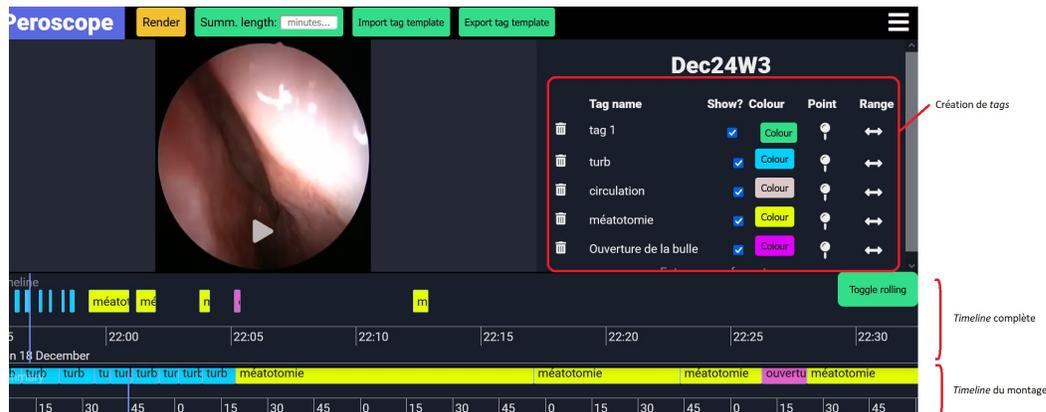


FIG. 1 – Interface du prototype « peroscope »

aux utilisateurs de créer des *tags* sur une *timeline* vidéo qui servent à sélectionner les séquences à conserver dans un résumé, le logiciel peut être utilisé en temps réel par plusieurs utilisateurs. Les utilisateurs peuvent d'abord créer des **catégories de tag**, en spécifiant un nom et une couleur, puis démarrer l'enregistrement d'une vidéo en direct. La vidéo est affichée au centre de l'interface, la partie inférieure montre deux *timelines*, une **timeline complète** qui représente toutes les séquences enregistrées, et la **timeline récapitulative**, qui additionne tous les passages *tagués*. L'utilisateur peut *taguer* les passages de la vidéo comme une « séquence », en marquant de longs segments, soit comme un « événement », avec une durée prédéfinie de 5 secondes pour marquer des segments courts. Une fois l'enregistrement en direct arrêté, les balises peuvent être redimensionnées, renommées, déplacées, supprimées, désactivées temporairement, et d'autres peuvent être ajoutés directement sur la *timeline* complète. Ces actions ont un effet direct sur la timeline récapitulative. Il est également possible de définir un objectif de durée, qui colore la *timeline* récapitulative jusqu'à cette durée. La vidéo enregistrée peut être parcourue en survolant les deux *timelines*, ce qui affiche le moment correspondant sur la vidéo. Si plusieurs utilisateurs travaillent sur le même projet, leurs modifications sont instantanément synchronisées (figure 1).

3.3 Procédure

Avant l'atelier, nous avons demandé aux chirurgiens experts d'envoyer une ou deux vidéos d'interventions chirurgicales enregistrées, d'environ une heure. Nous avons reçu une vidéo par chirurgien senior, que nous avons utilisée pour le projet. Le jour de l'atelier, nous avons accueilli les participants, expliqué le but de l'étude et présenté les formulaires de consentement. Nous avons présenté le prototype, en montrant un exemple de projet pour montrer les différentes fonctions, comment créer des *tags* : définir un nom, changer la couleur, désactiver et créer un *tag* dans une position spécifique sur la *timeline* — ainsi que comment modifier les *tags* — les repositionner, les redimensionner — ainsi que définir la longueur souhaitée. L'atelier comportait trois parties et était animé par le premier et le dernier auteur.

3.3.1 Partie I : taguer pour résumer. Le but de cette première partie était de demander aux participants de *taguer* la vidéo pendant 10 à 15 minutes. Nous avons d'abord demandé aux participants de définir plusieurs objectifs pour le résumé vidéo, par exemple créer un cours, présenter une technique pour une conférence ou un cas pour un staff chirurgical. Nous avons sélectionné un objectif pour la première tâche et informé les deux participants que cet objectif serait modifié dans la deuxième partie. Le chirurgien senior a travaillé seul sur la vidéo pour cette première partie, tandis que l'interne était observateur. Nous avons utilisé un protocole de « réflexion à haute voix » pour comprendre les actions du chirurgien.

3.3.2 Partie II : Travail collaboratif. Pour la deuxième partie, nous avons initialement prévu de supprimer des *tags* de manière aléatoire (moins de 5%) pour illustrer un fait courant dans le résumé vidéo, à savoir que le contexte (bloc opératoire, sécurité des patients, entre autres) peut entraîner l'omission de certains *tags* importants lors de la chirurgie. Nous n'avons finalement pas procédé à cette suppression car le nombre de *tags* créés était inférieur à celui prévu. Nous avons placé les participants sur des ordinateurs différents, avec pour instruction de se parler uniquement si nécessaire lors du montage collaboratif de la vidéo. Notre objectif était d'observer la communication à travers les *tag* entre les participants, par exemple des conseils ou instructions du chirurgien senior et des questions ou des suggestions de l'interne, la division potentielle du travail entre les participants, comment ils évitaient les redondances de tâches ou les conflits sur la *timeline*. Nous souhaitions également surveiller les besoins latents, les conflits et les points de frustration dus à des problèmes de collaboration.

3.3.3 Partie III : Réutilisation des tags. Pour la dernière partie, nous avons demandé aux participants de modifier l'objectif du montage. Par exemple, si l'objectif initial était de créer une vidéo pour une conférence (destiné aux experts), nous leur demandions de créer un résumé pour produire une vidéo pédagogique (destiné aux chirurgiens débutants). Notre objectif était d'observer comment le binôme crée ses *tags*, et s'approprie différentes fonctions (noms, couleurs, possibilité de désactivation) afin de servir la modification.

Après ces trois parties, nous avons débriefé avec les participants, pour préciser leurs impressions sur l'utilité de « taguer » pour résumer, les défis auxquels ils ont été confrontés et les stratégies qu'ils ont utilisées pour collaborer.

3.4 Collecte et analyse de données

Les ateliers ont été filmés à l'aide de deux caméras (une pour chaque participant) et le premier auteur a pris des notes écrites. Les participants ont rempli un court questionnaire (Annexe A) après l'atelier, portant sur l'utilisation des *tags*, (combien ils en ont créés, avant la vidéo, pendant la vidéo, au stade collaboratif) et les moyens de collaborer. Pour l'analyse, nous avons créé une liste des manières dont les *tags* ont été utilisés et appropriés, en identifiant les stratégies utilisées pour résumer par addition via les *tags* et en identifiant les défis rencontrés. Nous avons recherché des schémas qui reflètent la manière dont les différentes contraintes affectent ces usages, par exemple, le fait que l'objectif change, la relation entre les *tags* et le contenu de la vidéo, ainsi que la manière de collaborer.

4 RÉSULTATS

L'étude met en évidence les avantages et les limites de l'utilisation des *tags* et de l'approche additive pour la création de résumés vidéo. Parmi les avantages, le fait de pouvoir réaliser facilement plusieurs résumés (différentes longueurs pour différentes audiences) à partir de la même vidéo d'origine est particulièrement apprécié. L'étude montre également les changements de dynamiques de collaboration en fonction de l'expérience des internes avec plus de travail

synchrone avec le chirurgiens seniors lorsque l'interne a moins d'expérience. Enfin, l'étude interroge sur le besoin de plus d'*awareness* des actions des autres utilisateurs. Nous développons dans cette section chacun de ces résultats.

4.1 Montage par ajout de séquences

Tous les participants ont trouvé le concept pertinent. P6 a par exemple répondu « *Oui, surtout en live au bloc* » à la question sur l'utilisation de cet outil dans leur pratique quotidienne. Concernant le processus de montage, nous avons observé deux phases principales : le montage « grossier » et le montage « fin » (ou étape d'affinage). Dans la phase de montage grossier, les participants ont *tagué* les principales étapes de la chirurgie, en marquant le début de chacune de ces étapes, et prolongent ensuite le *tag* jusqu'à la fin de l'événement. Enfin, ils donnent un nom à ce *tag*. Nous observons qu'en délimitant les étapes chirurgicales, une grande partie des séquences inutiles sont automatiquement éliminées. Dans la deuxième phase, les participants ont repris leurs *tags* pour supprimer les gestes répétitifs, tels que les sutures ou les gestes longs et sans mouvements comme les gestes de cautérisation. Alors que la première phase a été réalisée par tous les participants, seuls deux des trois participants seniors ont effectué le montage « fin ». Ils ont indiqué que c'est la partie la plus fastidieuse du montage, et qu'elle ne nécessite pas toujours d'expérience du domaine. Le besoin de modifier (p. ex., changer la fin, diviser) les *tags* existants, afin de supprimer des passages en leur sein, n'a pas été facilité par notre prototype, en revanche, une partie supprimée d'un *tag* peut être récupérée facilement. Ce comportement contraste avec l'approche proposée qui consiste à ajouter/étendre les *tags* plutôt qu'à les supprimer/couper. P1 a exprimé qu'« *il serait intéressant de pouvoir effacer une partie des tags à éditer* », faisant référence à une véritable gomme physique, mimant de la main son besoin d'effacer une partie d'un long *tag*.

4.2 Appropriation des tags

À la fin de la première partie de l'atelier, les participants ont créé entre 10 et 15 *tags*. Tous ont choisi un nom pour les différents *tags*, mais P1 et P5 n'ont pas utilisé la couleur pour définir le type de *tag*. Les *tags* ont parfois été utilisés de manière inattendue : la première consiste à déléguer du travail. Les chirurgiens indiquent à l'interne sur quelle partie de la vidéo il doit travailler. La seconde consiste à créer plusieurs montages à partir d'une même vidéo. Par exemple, créer une synthèse pédagogique pour des internes inexpérimentés en gardant tous les gestes répétitifs d'une part, et créer une vidéo de conférence pour les chirurgiens seniors en ne gardant que la première occurrence du geste d'autre part. Ainsi, une catégorie de *tag* indique la séquence incluant tous les gestes répétitifs. Un *tag* supplémentaire indique uniquement la première occurrence du geste. Il est donc possible de produire différents montages pour différents objectifs en affichant/cachant différents types de *tags*. Cette notion est apparue particulièrement intéressante à tous les participants, les logiciels de montage classiques ne prenant pas en charge ce besoin. Enfin, P1 a suggéré que les *tags* pourraient, en plus de servir à réaliser le montage de la vidéo, être intégrés à la vidéo, pour servir de sous-titres ou d'annotations.

4.3 Montage collaboratif

Nous avons observé que la nature de la collaboration dépend de l'expertise des internes. P2 et P4 (les internes les moins expérimentés), ont immédiatement demandé des instructions à l'oral, tandis que P6 (l'interne le plus expérimenté) a commencé seul le montage « fin », sans les consignes du chirurgien senior. Lors de l'atelier avec le plus jeune interne (P4), le chirurgien senior a défini précisément les tâches à effectuer, puisqu'il lui a demandé de réaliser le montage « fin » d'une étape chirurgicale particulière, après lui avoir montré où cette étape se trouvait sur la *timeline* avec un *tag*. P4 avait également plus de questions à poser au chirurgien senior, car il peinait à différencier deux structures

anatomiques similaires et à trouver la fin de l'étape chirurgicale en question. P2, plus expérimenté, a reçu l'instruction de monter les étapes suivantes de la chirurgie, il a réalisé à la fois le séquençage « grossier » et « fin », il n'a pas eu besoin de beaucoup de conseils et la communication orale est restée rare. Le dernier interne (P6), le plus expérimenté de tous, fut rapidement laissé seul au montage, P5 observant principalement et ne donnant que très peu de conseils.

Ces observations montrent qu'un interne inexpérimenté ne peut pas effectuer le montage grossier, car il a du mal à identifier clairement les étapes chirurgicales, il a besoin de conseils pour s'orienter dans la vidéo, alors qu'un interne expérimenté peut effectuer le montage « grossier » ainsi qu'une partie du montage « fin », ne nécessitant qu'une validation. Par exemple, dans le troisième atelier, P5 n'a pas eu besoin de modifier le montage, et s'est contenté de vérifier le travail de son interne, ce qui aurait pu être fait de manière asynchrone. À l'inverse, dans le deuxième atelier, P4 n'aurait pas pu continuer à travailler sans l'assistance immédiate de P3. La communication via les *tags* est apparue dans l'un des ateliers mais n'était pas le moyen de communication le plus utilisé. La nécessité de pointer des éléments sur la *timeline* est apparue dans les trois workshops, la visualisation des fenêtres de travail des autres utilisateurs était suffisante pour éviter les conflits mais insuffisante pour que l'interne ou le chirurgien attirent l'attention de l'autre sur une partie précise. Par exemple, P3 a dit « *vous devez travailler ici* » en montrant avec son pointeur le moment dont il parlait sur la chronologie.

4.4 Usage des *tags* au bloc opératoire

Notre dernier résultat concerne des impressions sur l'enregistrement de *tags* au bloc opératoire. Dans les questionnaires, tous les participants étaient positifs quant à la mise en place du montage grâce aux *tags* au bloc opératoire. Lors du dernier atelier, P5 a insisté sur l'importance d'« *une interface simple* », soulignant que plus elle sera simple, plus elle sera utilisée. P1 a insisté sur l'importance pour le chirurgien de savoir qu'un *tag* est en cours, afin d'éviter des marquages sans fin et des omissions de marquage avec un minimum de distraction. Il a toutefois été admis que des erreurs et des omissions étaient inévitables. Tous les chirurgiens ont émis l'hypothèse que le *tagging* per-opératoire resterait grossier et pourrait ensuite être corrigé. La question de l'interaction pour démarrer un *tag* au bloc opératoire a été systématiquement posée, mais sans être explorée car ce n'était pas l'objet de notre travail.

5 DISCUSSION

Les résultats de cette étude montrent (1) des étapes similaires à celles identifiées dans la littérature pour la création de résumés vidéo [5]; et, (2) que le processus de création est chronophage et difficile [27]. Nous contribuons à cette littérature, en démontrant l'avantage des *tags* pour la création collaborative de résumés vidéo.

5.1 Complémentarité des approches additives et soustractives

Notre étude suggère qu'une seule approche n'est pas suffisante pour la création de résumés vidéo. L'approche additive est efficace pour l'étape grossière consistant à identifier les segments pertinents. En revanche, l'approche soustractive apparaît plus pertinente pour l'étape d'affinage pour éliminer des séquences superflues. Une fonctionnalité de découpage et de soustraction reste donc nécessaire. L'usage de *tags*, pour gagner en efficacité [7] ou pour favoriser la collaboration [26] est déjà admis. Notre étude suggère en revanche la pertinence des *tags* pour la création de résumés vidéo car ils servent, en plus de leurs fonctions initiales, d'outil de découpage vidéo. Nous pensons que l'usage de *tags* est complémentaire avec les méthodes (semi-)automatiques d'aide au montage vidéo. Par exemple, lors des étapes de *séquençage fin*, les séquences avec des gestes répétitifs, sans mouvement ou floues peuvent être *taguées* ou supprimées automatiquement. D'un autre côté, *taguer* les étapes principales resterait à la discrétion du chirurgien.

5.2 Travail collaboratif : complément entre décision et main d'oeuvre

Nous interprétons la délégation des chirurgiens seniors vers les internes de certaines séquences comme le souhait de faire gagner du temps aux seniors plutôt que le souhait d'améliorer la qualité du résultat final en croisant les expertises [2]. En fonction de son expertise, plus ou moins de tâches sont déléguées à l'interne. De la même façon, la quantité de travail synchrone dépend de l'expertise de l'interne : moins l'interne est expérimenté, plus il a besoin de travailler côte à côte avec le chirurgien senior pour recevoir de l'aide pour reconnaître les structures anatomiques et les différentes étapes théoriques dans la vidéo. Les résultats suggèrent donc que la possibilité de travail synchrone, moins importante chez les utilisateurs expérimentés, devient obligatoire si un des utilisateurs a peu d'expérience. La répartition du travail, ainsi que de l'utilité du travail synchrone restent à interpréter avec prudence car les vidéos chirurgicales sont par essence souvent longues de plusieurs heures [5, 10]. Le travail synchrone peut permettre d'accélérer la création de résumés chirurgicaux en favorisant le travail à plusieurs sur de longs projets.

Finalement, nous soulignons, comme mentionné par 2/3 des participants et la littérature [2, 12], que les aspects collaboratifs du concept peuvent améliorer la fonction pédagogique / réflexive. En marquant les étapes, et gestes importants, l'interne prend conscience de son niveau de compréhension sur une chirurgie particulière, et, montre quelles notions lui sont manquantes pour réaliser le montage final. Enfin, alors que l'apport de l'apprentissage numérique via la réalité virtuelle et les vidéos pédagogiques est à l'origine de multiples études [14, 15], notre étude est la première à explorer l'amélioration de l'efficacité du montage de vidéo chirurgicale en appliquant des principes existant en CSCW.

5.3 Limitations et travail futur

Notre prototype avait plusieurs limitations. Il présentait des bugs et un temps de mise en mémoire tampon important, provoquant des problèmes de synchronisation entre la position du curseur sur la *timeline* et l'image vidéo correspondante, ce qui finissait par altérer l'expérience utilisateur. Cependant, le fait que les participants soient tout de même d'accord sur la pertinence du concept renforce la validité de nos résultats. De plus, l'applicabilité au bloc opératoire n'a pas été testée dans ce travail préliminaire. La prochaine étape est l'introduction du concept dans la pratique quotidienne des chirurgiens, notamment au bloc opératoire, ce qui comporte de nombreux défis comme la prise en compte de la distraction propre à cet environnement [3] ainsi que de trouver des participants pour tester le concept en pratique réelle. Enfin, il s'agira également de prouver l'efficacité du concept en comparant le prototype aux logiciels de montages existants en considérant à la fois la qualité des résumés produits et le temps pour les produire.

REMERCIEMENTS

Ce travail a été financé par le Labex CAMI (projet ANR-11-LABX-0004), programme Investissement d'Avenir.

RÉFÉRENCES

- [1] Lingga Adidharma, Zixin Yang, Christopher Young, Yangming Li, Blake Hannaford, Ian Humphreys, Waleed M. Abuzeid, Manuel Ferreira, Kristen S. Moe, and Randall A. Bly. 2021. Semiautomated Method for Editing Surgical Videos. In *Journal of Neurological Surgery Part B : Skull Base*, Vol. 82. Georg Thieme Verlag KG, P057. <https://doi.org/10.1055/s-0041-1725452> ISSN : 2193-6331, 2193-634X Issue : S 2 Journal Abbreviation : J Neurol Surg B Skull Base.
- [2] Amr I. Al Abbas, Jae P. Jung, MaryJoe K. Rice, Amer H. Zureikat, Herbert J. Zeh, and Melissa E. Hogg. 2019. Methodology for Developing an Educational and Research Video Library in Minimally Invasive Surgery. *Journal of Surgical Education* 76, 3 (May 2019), 745–755. <https://doi.org/10.1016/j.jsurg.2018.10.011>
- [3] Sophia Antoniadis, Stefanie Passauer-Baierl, Heiko Baschnegger, and Matthias Weigl. 2014. Identification and interference of intraoperative distractions and interruptions in operating rooms. *Journal of Surgical Research* 188, 1 (May 2014), 21–29. <https://doi.org/10.1016/j.jss.2013.12.002>

- [4] Knut Magne Augestad, Khayam Butt, Dejan Ignjatovic, Deborah S. Keller, and Ravi Kiran. 2020. Video-based coaching in surgical education : a systematic review and meta-analysis. *Surgical Endoscopy* 34, 2 (Feb. 2020), 521–535. <https://doi.org/10.1007/s00464-019-07265-0>
- [5] Ignacio Avellino, Sheida Nozari, Geoffroy Canlorbe, and Yvonne Jansen. 2021. Surgical Video Summarization : Multifarious Uses, Summarization Process and Ad-Hoc Coordination. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 140 :1–140 :23. <https://doi.org/10.1145/3449214>
- [6] Ajay Divakaran, Kadir A. Peker, and Huifang Sun. 2001. Video summarization using motion descriptors. In *Storage and Retrieval for Media Databases 2001*, Vol. 4315. International Society for Optics and Photonics, 517–522. <https://doi.org/10.1117/12.410963>
- [7] Paul Dourish and Victoria Bellotti. 1992. Awareness and coordination in shared workspaces. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work (CSCW '92)*. Association for Computing Machinery, New York, NY, USA, 107–114. <https://doi.org/10.1145/143457.143468>
- [8] Zhiwei Fan, Congliang Chen, and Tingting Jiang. 2017. A semi-automatic editing method for surgery videos. In *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. 115–120. <https://doi.org/10.1109/ICMEW.2017.8026213>
- [9] K. Fujimura, K. Honda, and K. Uehara. 2002. Automatic video summarization by using color and utterance information. In *Proceedings. IEEE International Conference on Multimedia and Expo*, Vol. 1. IEEE, New York, NY, USA, 49–52 vol.1. <https://doi.org/10.1109/ICME.2002.1035715>
- [10] Robert S. Gilder. 1988. Editing surgical videotapes. *Journal of Audiovisual Media in Medicine* 11, 4 (Jan. 1988), 117–120. <https://doi.org/10.3109/17453058809021622> Publisher : Taylor & Francis _eprint : <https://doi.org/10.3109/17453058809021622>.
- [11] Mario Guggenberger, Michael Riegler, Mathias Lux, and Pål Halvorsen. 2014. Event understanding in endoscopic surgery videos. In *Proceedings of the 1st ACM International Workshop on Human Centered Event Understanding from Multimedia*. ACM, Association for Computing Machinery, New York, NY, USA, 17–22. <https://doi.org/10.1145/2660505.2660509>
- [12] Giselle G. Hamad, Matthew T. Brown, and Julio A. Clavijo-Alvarez. 2007. Postoperative video debriefing reduces technical errors in laparoscopic surgery. *The American Journal of Surgery* 194, 1 (July 2007), 110–114. <https://doi.org/10.1016/j.amjsurg.2006.10.027> Publisher : Elsevier.
- [13] Samuel Huron, Petra Isenberg, and Jean Daniel Fekete. 2013. PolemicTweet : Video Annotation and Analysis through Tagged Tweets. In *Human-Computer Interaction – INTERACT 2013*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Paula Kotzé, Gary Marsden, Gitte Lindgaard, Janet Wesson, and Marco Winckler (Eds.). Vol. 8118. Springer Berlin Heidelberg, Berlin, Heidelberg, 135–152. https://doi.org/10.1007/978-3-642-40480-1_9 Series Title : Lecture Notes in Computer Science.
- [14] Santiago González Izard, Juan A. Juanes, Francisco J. García Peñalvo, Jesús M^o Gonçalves Estella, M^o José Sánchez Ledesma, and Pablo Ruisoto. 2018. Virtual Reality as an Educational and Training Tool for Medicine. *Journal of Medical Systems* 42, 3 (Feb. 2018), 50. <https://doi.org/10.1007/s10916-018-0900-2>
- [15] Rishad Khan, Joanne Plahouras, Bradley C. Johnston, Michael A. Scaffidi, Samir C. Grover, and Catharine M. Walsh. 2019. Virtual reality simulation training in endoscopy : a Cochrane review and meta-analysis. *Endoscopy* 51, 7 (July 2019), 653–664. <https://doi.org/10.1055/a-0894-4400>
- [16] Jeongyeon Kim, Daeun Choi, Nicole Lee, Matt Beane, and Juho Kim. 2023. Surch : Enabling Structural Search and Comparison for Surgical Videos. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–17. <https://doi.org/10.1145/3544548.3580772>
- [17] I. Kitahara, K. Kogure, and N. Hagita. 2004. Stealth vision for protecting privacy. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, Vol. 4. 404–407 Vol.4. <https://doi.org/10.1109/ICPR.2004.1333788> ISSN : 1051-4651.
- [18] Clemens N. Klokose, James R. Eagan, Siemen Baader, Wendy Mackay, and Michel Beaudouin-Lafon. 2015. Webstrates : Shareable Dynamic Media. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM, New York, NY, USA, 280–290. <https://doi.org/10.1145/2807442.2807446>
- [19] Clemens N. Klokose, Christian Remy, Janus Bager Kristensen, Rolf Bagge, Michel Beaudouin-Lafon, and Wendy Mackay. 2019. Videostrates : Collaborative, Distributed and Programmable Video Manipulation. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (UIST '19)*. Association for Computing Machinery, New York, NY, USA, 233–247. <https://doi.org/10.1145/3332165.3347912>
- [20] Jarrod Knibbe, Sue Ann Seah, and Mike Fraser. 2015. Videohandles : Searching through action camera videos by replicating hand gestures. *Computers & Graphics* 48 (2015), 99–106.
- [21] Mackenzie Leake, Abe Davis, Anh Truong, and Maneesh Agrawala. 2017. Computational video editing for dialogue-driven scenes. *ACM Transactions on Graphics* 36, 4 (Aug. 2017), 1–14. <https://doi.org/10.1145/3072959.3073653>
- [22] Andreas Leibetseder, Bernd Munzer, Klaus Schoeffmann, and Jorg Keckstein. 2017. Endometriosis Annotation in Endoscopic Videos. In *2017 IEEE International Symposium on Multimedia (ISM)*. IEEE, Taichung, 364–365. <https://doi.org/10.1109/ISM.2017.69>
- [23] Junhua Liao, Haihan Duan, Xin Li, Haoran Xu, Yanbing Yang, Wei Cai, Yanru Chen, and Liangyin Chen. 2020. Occlusion Detection for Automatic Video Editing. In *Proceedings of the 28th ACM International Conference on Multimedia (MM '20)*. Association for Computing Machinery, New York, NY, USA, 2255–2263. <https://doi.org/10.1145/3394171.3413725>
- [24] Constantinos Loukas. 2018. Video content analysis of surgical procedures. *Surgical Endoscopy* 32, 2 (Feb. 2018), 553–568. <https://doi.org/10.1007/s00464-017-5878-1>
- [25] Mathias Lux, Oge Marques, Klaus Schöffmann, Laszlo Böszörményi, and Georg Lajtai. 2010. A novel tool for summarization of arthroscopic videos. *Multimedia Tools and Applications* 46, 2 (Jan. 2010), 521–544. <https://doi.org/10.1007/s11042-009-0353-1>
- [26] W. E. Mackay. 1989. EVA : an experimental video annotator for symbolic analysis of video data. *ACM SIGCHI Bulletin* 21, 2 (Oct. 1989), 68–71. <https://doi.org/10.1145/70609.70617>

- [27] Gaël Nana, Charisse Colvin, and Omer Karim. [n.d.]. Surgical video – part 1 : intraoperative video recording and storage. 19, 3 ([n. d.]).
- [28] JungHwan Oh, Sae Hwang, Wallapak Tavanapong, Piet C. de Groen, and Johnny Wong. 2003. Blurry-frame detection and shot segmentation in colonoscopy videos. In *Storage and Retrieval Methods and Applications for Multimedia 2004*, Vol. 5307. SPIE, 531–542. <https://doi.org/10.1117/12.527108>
- [29] Amy Pavel, Dan B. Goldman, Björn Hartmann, and Maneesh Agrawala. 2016. VidCrit : Video-based Asynchronous Video Review. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. Association for Computing Machinery, New York, NY, USA, 517–528. <https://doi.org/10.1145/2984511.2984552>
- [30] Roy Pea, Robb Lindgren, and Joseph Rosen. 2008. Cognitive technologies for establishing, sharing and comparing perspectives on video over computer networks. *Social Science Information* 47, 3 (2008), 353–370. <https://doi.org/10.1177/0539018408092577> _eprint : <https://doi.org/10.1177/0539018408092577>.
- [31] Roy D. Pea and Eric Hoffert. 2007. Video workflow in the learning sciences : Prospect of Emerging technologies for argmenting work practices. Lawrence Erlbaum Associates, 427. <https://telearn.hal.science/hal-00190028>

A WORKSHOP QUESTIONNAIRES

Post Workshop Questionnaire

Please answer the following question by reflecting on the concept of summarizing through tagging, and summarizing through collaborative work. We acknowledge that our prototype has delays and minor bugs that we can fix, please keep this feedback for the second part of the questionnaire.

In what way are tags useful for making video summaries ?

Please describe one specific moment where summarizing through tags was particularly useful

Please describe one specific moment where tags made it more difficult to create the video summary

How did you collaborate with your colleague to summarize the video?

Please describe one specific moment where the collaboration was particularly useful.

Please describe one specific moment where the collaboration made it more difficult to create the video summary

Do you see yourself using this concept in your daily practice, including in the OR and after? Why?

For the next set of questions, please answer regarding the implementation of the property.

Did you find the prototype software easy to use ? Why ?

What should we improve before testing the prototype in the OR ?
