Évaluation de la qualité et transmission en temps-réel de vidéos médicales compressées : application à la télé-chirurgie robotisée

Nedia Nouri

École doctorale IAEM Lorraine

Institut National Polytechnique de Lorraine

Évaluation de la qualité et transmission en temps réel de vidéos médicales compressées : application à la télé-chirurgie robotisée

THÈSE

présentée et soutenue à huis clos le 9 Septembre 2011

pour l'obtention du

Doctorat de l'Institut National Polytechnique de Lorraine

Spécialité Automatique, Traitement du signal et des Images, Génie informatique

par

Nedia NOURI

Composition du jury

Rapporteurs :	M. Benoît Macq	Professeur, TELE, Université Catholique de Louvain (Belgique)
	M. Marc Antonini	DR CNRS, I3S, Université de Nice-Sophia Antipolis
Examinateurs :	M. Daniel Negru	MCF, LaBRI, Université de Bordeaux
	M. Philippe Letellier	Responsable Innovation, Institut Télécom, Paris
	M. Michel Dufaut	Professeur, CRAN, Nancy Université - Directeur
	M. Jean-Marie Moureaux	Professeur, CRAN, Nancy Université - Co-directeur
	M. Denis Abraham	Chercheur, CRAN, Nancy Université - Co-encadrant
	M. Jacques Hubert	PUPH, CHU Nancy, Université Henri Poincaré

Centre de Recherche en Automatique de Nancy CNRS – UMR 7039

Nancy-Université

Mis en page avec la classe thloria.

Remerciements

Les travaux de recherche présentés dans ce mémoire ont été effectués au Centre de Recherche en Automatique de Nancy (CRAN), Unité Mixte de Recherche Nancy-Université, CNRS (UMR 7039), dans le groupe thématique Ingénierie Pour la Santé (IPS). Je remercie Monsieur Alain Richard, directeur du CRAN, pour son accueil et sa disponibilité.

Je tiens à remercier les membres du jury qui m'ont fait l'honneur de participer à l'examen de ce travail.

Monsieur Benoît Macq, Professeur à l'Université Catholique de Louvain-La-Neuve, pour l'intérêt qu'il a porté à mes travaux. Je lui sais gré d'avoir accepté de rapporter mon travail et de participer à mon jury de thèse.

Monsieur Marc Antonini, Directeur de Recherche au CNRS, pour le temps et l'attention qu'il a accordés à mon manuscrit en tant que rapporteur. Je lui exprime toute ma reconnaissance.

Monsieur Philippe Lettellier, Responsable Innovation à l'institut Télécom, pour ses nombreux commentaires constructifs.

Monsieur Daniel Négru, Maître de Conférences à l'Université de Bordeaux, pour sa participation à mon jury.

Monsieur Jacques Hubert, PUPH à l'Université Henri Poincaré, initiateur du projet de chirurgie à distance, pour les échanges que nous avons eus et qui ont contribué à orienter la réflexion pour mener à bien ces travaux.

Ce travail a été effectué sous la direction de Monsieur Michel Dufaut, Monsieur Jean-Marie Moureaux et Monsieur Denis Abraham. Je les remercie de m'avoir fait bénéficier de leurs compétences et d'avoir été patients notamment pendant l'étape de rédaction.

Je remercie tout particulièrement Jean-Marie de m'avoir accompagnée dans ce parcours semé d'embûches et qui a su, avec sa persévérance légendaire, redresser le navire malgré les diverses tempêtes. Un grand Merci Jean-Marie pour ta confiance, ta patience et ton humanité !

Je remercie tous les chirurgiens qui ont participé à l'étude de qualité : J. Siat, R. Frisoni, C. Laurain, C. Perrenot, A. Germain, M. Durand, L. Robert, J.M Tortuyeux, M. Fau, T. Serradori, T. Fouquet, C. Erb, V. Anne, M.L Scherrer, N. Reibel, N. Berte.

Merci au Docteur Manuela Perez d'avoir motivé ses confrères et consœurs pour participer à cette étude.

Je n'oublie évidemment pas toutes les personnes qui m'ont aidée durant ce parcours.

Je pense tout particulièrement à Céline Fournier, qui m'a accompagnée dans les méandres de l'administration hospitalière, toujours avec le sourire.

Nicolas Visona et Arnaud Antonelli, ingénieurs réseau, recpectivement au CHU et à la Faculté de Médecine de Nancy, pour leur aide et leur disponibilité.

Toute l'équipe de l'école de chirurgie de Nancy pour leur accueil et leur disponibilité.

Monsieur Edouard Yvroud, pour sa gentillesse et ses encouragements réguliers.

Mes collègues du CRAN : Ricardo, Christophe, Julie et mes précieux amis Rebeca et Hugo (ainsi que leur petit André) pour tout leur soutien et leur amitié sincère.

Tous ceux qui ont contribué à ces travaux : Vincent Guillemot et Christophe Burlot.

Toutes les personnes qui m'ont accompagnée pour faire mes premiers pas dans l'enseignement : Valérie, Radu, Samia et Nicolae.

Je remercie la Région Lorraine et la Communauté Urbaine du Grand Nancy pour avoir co-financé mes travaux pendant 3 ans.

Enfin, un immense merci à tous ceux qui ont refusé de m'aider sous différents prétextes. Grâce à vous j'ai appris à puiser en moi la plus grande des énergies pour continuer à faire avancer les choses et j'ai surtout découvert qu'il n'est pas d'impasse sans sortie, à condition de bien vouloir marcher dans la bonne direction. Encore une fois, MERCI !

Bien évidemment, je ne pourrais terminer sans adresser un immense merci à tous ceux qui comptent pour moi.

A toute ma famille, ma belle famille et mes amis pour leur soutien et leur intérêt.

Sans oublier Isabelle, pour nos longues discussions sur les « guignols de la vie » et Rebeca, toujours volontaire pour déguster mes plats issus de la gastronomie mondiale.

A Little N. : celle à qui je dois beaucoup ; même si elle m'a parfois joué des tours.

Enfin, mes plus grands remerciements reviennent à Hervé, pour son soutien indéfectible. Merci pour ta joie de vivre et ton enthousiasme. Merci mille fois pour notre complicité, pour nos efforts communs et surtout pour ta compréhension.

« Vous ne donnez que peu lorsque vous donnez vos biens. C'est lorsque vous donnez de vous-mêmes que vous donnez réellement. »

Khalil Gibran - Le prophète

À Mon grand-père

À Hervé

Table des matières

Introduction générale

De nos jours, les images et les vidéos numériques sont omniprésentes et les données associées sont gigantesques. Ce phénomène a nécessité la naissance de plusieurs méthodes de traitement des images et des vidéos et en particulier les techniques de compression permettant de réduire considérablement la quantité de données. Bien que ces techniques soient complètement transparentes à l'utilisateur final, elles sont présentes dans plusieurs applications de la vie courante. Les ordinateurs, les téléphones portables, les appareils photos, les caméras vidéo, la télévision utilisent des encodeurs/décodeurs (Codec). Cependant, il existe un domaine, longtemps réticent à ces techniques malgré la « révolution numérique » qu'il connaît, où de grandes quantités de données numériques sont générées quotidiennement : le domaine médical.

Les informations médicales, composées de données cliniques, d'images, de signaux physiologiques, sont stockées et souvent transmises à travers les réseaux de communication. Le développement du PACS (Picture Archiving and Communication System) dans les hôpitaux a conduit à l'amélioration de la circulation de l'information radiologique à travers les réseaux : depuis la demande d'examens, jusqu'à la mise à disposition du résultat (intégrant image et compte rendu) dans le cadre d'un dossier patient multimédia. L'information radiologique est mise à disposition des bons acteurs (radiologues et cliniciens de manière simultanée), dans l'hôpital, dans un réseau de services, voire à terme dans un réseau de soins ville - hôpital. Il s'agit, donc, de connecter l'ensemble des modalités d'imagerie numérique (TDM, IRM, Angiographie, Radiologie Conventionnelle, Médecine Nucléaire, Echographie et Mammographie), archiver les résultats et les redistribuer à l'ensemble des acteurs de l'hôpital. La contrepartie de cette numérisation réside dans

des volumes de données considérables nécessitant des capacités de transmission et de stockage elles-mêmes très importantes, rendant la **compression avec perte** en amont inévitable. Celle-ci constitue un défi majeur dans un contexte aussi sensible que le contexte médical, celui de l'impact des pertes sur la qualité des données et leur exploitation.

La chirurgie fait appel également aux nouvelles technologies afin d'améliorer l'efficacité des traitements mais aussi le rétablissement de patients. En effet, il existe plusieurs systèmes développés dans le monde médical, tels que les robots de chirurgie, qui donnent le moyen au chirurgien d'opérer à distance son patient. Ces systèmes très perfectionnés permettent non seulement d'augmenter la précision du geste du chirurgien mais également de réduire la taille des cicatrices post-opératoires et la durée du séjour du patient à l'hôpital. Actuellement, dans les hôpitaux équipés de robots de chirurgie, un chirurgien peut opérer en étant à quelque mètres de son patient, voire dans une autre salle. Les données numérisées issues du robot de chirurgie ne sont pas compressées car elles transitent sur un réseau dédié. Cependant, ce développement ouvre des perspectives d'interventions chirurgicales à distance, comme l'a montré la célèbre expérimentation « Opération Lindbergh » en 2001 [Marescaux *et al.*, 2002]. La transmission de vidéos médicales sur de grandes distances est un sujet en plein développement, du fait de la « révolution numérique » que connaît le monde médical actuellement notamment dans le domaine de la chirurgie robotique. La perspective de réaliser des opérations chirurgicales à longues distances est d'un grand intérêt pour les patients mais aussi pour l'hôpital. Par exemple, la possibilité d'accès à la même qualité de soins, quelque soit l'emplacement géographique du patient, peut permettre de réaliser des économies pour le patient et l'hôpital. La chirurgie à distance peut s'avèrer d'une grande utilité dans des zones sinistrées ou dans de petits hôpitaux ne disposant pas d'équipes de chirurgiens expérimentés. Les services ne sont pas de même qualité suivant les zones géographiques. Dans une zone rurale, la chirurgie dispose de plus faibles moyens ce qui peut limiter la qualité des soins. Cependant, le coût de telles manipulations reste important (coût du robot et coût du réseau de communication). En ayant recours à la chirurgie à distance, l'échelle des distances entre chirurgiens et patients serait diminuée, ce qui augmenterait le rendement des professionnels de la santé tout en mettant à leur disposition des aides à la décision. En couplant

plusieurs techniques telles que l'imagerie et la robotique médicale, les intervenants disposeraient d'outils permettant d'augmenter la qualité des interventions en diminuant la durée de celles-ci, ce qui est bénéfique pour le patient. Dans le domaine militaire, la chirurgie à distance permettrait des soins spécifiques aux soldats blessés sur un champ de bataille sans pour autant nécessiter le déplacement de spécialistes.

La transmission en temps réel de vidéos chirurgicales afin d'opérer à distance un patient, nécessite une étude préalable de la qualité visuelle des vidéos compressées. En effet, la relation entre la qualité perçue des vidéos et leur débit après compression pemettra de déterminer le caractéristiques du réseau de transmission des données (type, contraintes, bande passante, latence). Pour la détermination de cette relation, deux approches méthodologiques sont possibles : l'une à caractère subjectif et l'autre à caractère objectif.

Dans cette thèse, nous nous intéressons à la fois à l'évaluation subjective de la qualité des vidéos chirurgicales compressées et à la mesure objective de leur qualité à travers une métrique avec référence complète. Le travail présenté ici concerne l'évaluation subjective, par un panel de chirurgiens, de séquences vidéo issues d'une application de télé-robotique chirurgicale et compressées aux formats MPEG2 et H.264. L'évaluation s'appuie sur des recommandations internationales à travers des protocoles de test normalisés, l'objectif étant de déterminer un seuil de compression permettant d'avoir une qualité perceptuelle de la vidéo chirurgicale compressée irréprochable, pour le bon déroulement de l'opération et pour assurer la sécurité des patients. Nous montrons qu'il existe un seuil de tolérance à la compression avec pertes de type MPEG2 autour de 3 Mbits/s, ce qui équivaut à un taux de compression d'environ 90 :1 du flux vidéo initialement à 270 Mbits/s ! Nous démontrons que ce seuil est situé autour de 2 Mbits/s si les vidéos sont compressées avec le standard H.264. Nous mettons ainsi en exergue la possibilité de compresser fortement des flux vidéo dans le contexte de la chirurgie robotisée. A partir de ces résultats, nous prouvons la faisabilité d'une transmission temps-réel de ce type de vidéo depuis le robot installé au bloc opératoire du CHU jusqu'à un site distant (l'Ecole de Chirurgie), et ce à travers un réseau IP partagé avec plusieurs autres applications, en l'occurrence, celui de la faculté de Médecine de Nancy. Ces résultats permettent d'envisager un jour la possibilité d'effectuer des opérations chirurgicales à distance dans

des conditions réalistes.

Ce manuscrit est organisé en cinq chapitres et structuré de la manière suivante.

Le chapitre 1 décrit le contexte médical de nos travaux ainsi que les problématiques et objectifs scientifiques.

Le chapitre 2 est dédié à la compression vidéo et notamment aux principes des standards vidéos existants et aux distorsions de la vidéo inhérentes à la compression.

Le chapitre 3 est un état de l'art des méthodes d'avaluation subjective de la qualité des vidéos et des métriques de qualité objectives.

Le chapitre 4 précise le cadre des études de qualité réalisées, des méthodologies utilisées et leur adaptation au contexte de recherche. Il présente les résultats expérimentaux obtenus après une camapgne de tests subjectifs de la qualité des vidéos issues du robot de chirurgie et l'objectivation de ces résultats à travers deux métriques objectives.

Le chapitre 5 est consacré à la transmission en temps réel de flux vidéos entre le bloc opératoire du CHU de Nancy et la Faculté de Médecine de Nancy. Cette transmission permet de valider les résultats expérimentaux de la qualité des vidéos chirurgicales.

Chapitre 1

Problématiques et contraintes de la chirurgie robotisée

La première partie de ce chapitre présente le contexte médical : la chirugie mini-invasive et les différents verrous technologiques inhérents à la chirurgie à distance. La deuxième partie présente les problématiques techniques et les objectifs scientifiques de ce travail.

1.1 Problématique Générale

1.1.1 Introduction

L'évolution de certaines techniques chirurgicales, par l'utilisation de robots de chirurgie, permet aujourd'hui des interventions mini-invasives avec une précision de geste très importante. Le second avantage de cette évolution est qu'elle ouvre des perspectives d'interventions chirurgicales à distance, comme l'a montré la célèbre expérimentation « Opération Lindbergh » en 2001 [Marescaux *et al.*, 2002]. La généralisation de ce type d'expérience n'est cependant pas encore acquise du fait des ressources réseau qu'elle nécessite, en particulier en termes de bande passante. Dans un contexte d'opération à grande distance, de télé-enseignement ou encore de télé-conseil (Telementoring), la transmission à travers les réseaux de communication de volumineux flux vidéo nécessite des capacités très importantes en termes de bande passante, rendant la compression avec perte de ces

flux incontournable. Cependant, lorsque le taux de compression augmente, les encodeurs introduisent dans la vidéo compressée des artefacts pouvant affecter sa qualité visuelle. Dans le contexte de l'imagerie médicale et en particulier celui de la télé-chirurgie, cette altération de la qualité peut être rédhibitoire. Dans des applications aussi sensibles que les applications médicales, il est indispensable de se référer au jugement humain par des essais subjectifs de la qualité perçue [ITU-R, 2000].

1.1.2 Chirurgie mini-invasive et chirurgie robotique

La pratique de la chirurgie mini-invasive (ou chirurgie par trou de serrure) a permis au chirurgien, via un écran et une caméra endoscopique, d'opérer son patient sans pratiquer une incision de 40 centimètres sur l'abdomen ou le thorax du patient. Contrairement à la chirurgie classique, dite chirurgie ouverte, les seules cicatrices mesurent 10 millimètres de diamètre et sont celles occasionnées par l'introduction de deux outils et d'un endoscope. La chirurgie mini-invasive par voie abdominale (appelée cœliochirurgie ou chirurgie laparoscopique) est actuellement le *Gold Standard* [Heinzelmann *et al.*, 1995], [Soper *et al.*, 1992] pour de nombreuses interventions réalisées très couramment : cholécystectomie, appendicectomie, cure de hernie hiatale. Le principe consiste à opérer les malades en faisant 3 à 5 petites incisions de l'ordre de 1 cm. Par ces incisions, on introduit des trocarts qui permettent le passage d'une caméra et d'outils de cœlioscopie, comme des pinces, des ciseaux ou des porte-aiguilles. La vision de l'intervention se fait sur un moniteur en deux dimensions. Il s'agit d'une technique plus difficilement applicable à certaines interventions comme la chirurgie cardiaque avec pontages à cœur battant, la chirurgie fonctionnelle (hernie hiatale) ou la chirurgie cancérologique (prostatectomies radicales). En effet, dans ces interventions souvent longues, la précision du geste est primordiale (sutures vasculaires) et la dissection est étendue (curage ganglionnaire). L'évolution de la chirurgie laparoscopique a permis une seconde avancée importante qui est l'utilisation des robots de chirurgie. Ils apportent une fiabilité et une efficacité au geste du chirurgien et permettent d'améliorer son confort et celui de son patient notamment en réduisant les risques post-opératoires. La plateforme d'un robot de chirurgie se compose de deux parties. La première partie est une machine constituée de trois ou quatre bras

mobiles qui portent les instruments cœlioscopiques (voir figure 1.2) et la caméra binoculaire endoscopique. La caméra binoculaire permet d'acquérir simultanément deux vues légèrement décalées du champ opératoire. La deuxième partie, positionnée à quelques mètres de la table d'opération ou dans une autre pièce, est une console (figures 1.3 et 1.4) où le chirurgien dispose de manettes (pour piloter à distance les instruments) et d'un pédalier (pour sélectionner soit les bras soit la caméra). Un viseur binoculaire (figure1.1) permet au chirurgien de visualiser le champ opératoire en 3D (en le reconstruisant naturellement). Le robot da Vinci (Intuitive Surgical) est utilisé quotidiennement au CHU de Nancy Brabois pour effectuer des opérations comme l'ablation de la prostate ou certaines tumeurs ORL mais aussi pour la chirurgie de l'obésité ou la chirurgie cardiaque. Le fabricant américain Intuitive Surgical, a un quasi-monopole sur le marché de la robotique chirugicale ce qui pose problème pour l'accessibilité aux données dans un cadre de recherche.

FIGURE 1.1. *Système de vision binoculaire (Intuitive Surgical ©)*

Cette nouvelle technologie apporte une aide pour le chirurgien pour effectuer avec précision ces gestes complexes car il bénéficie :

- d'une vision en 3D du champ opératoire,
- de la possibilité d'agrandir et de démultiplier le geste,
- d'une précision du geste opératoire avec des instruments articulés,
- d'une position ergonomique devant la console binoculaire

Cette technologie offre aujourd'hui de nombreuses perspectives à court terme, parmi lesquelles le *Telementoring* (que l'on pourrait traduire par Télé-conseil) et le *télé-enseignement*

FIGURE 1.2. *Instruments du robot (Intuitive Surgical ©)*

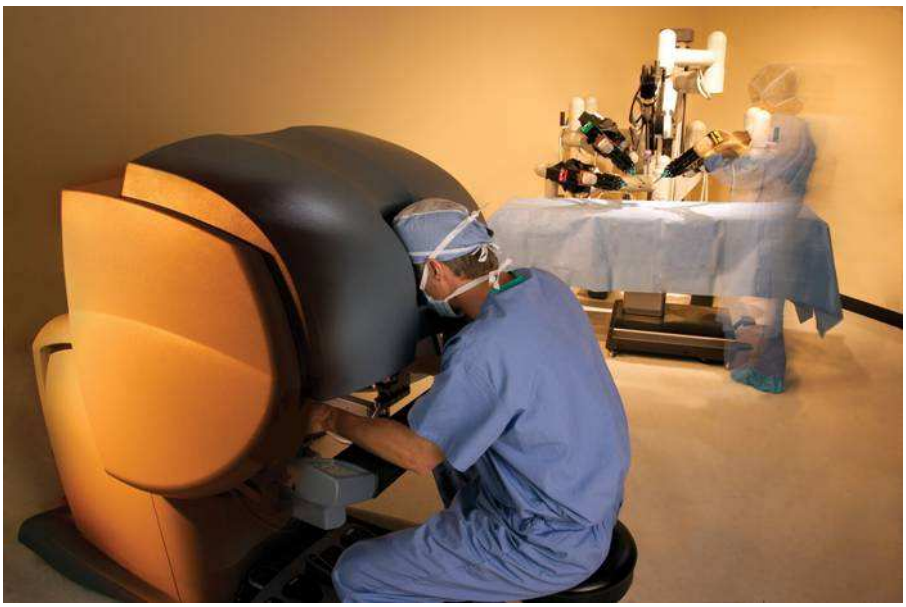


FIGURE 1.3. *Console du robot (Intuitive Surgical ©)*

mais aussi une perspective à plus long terme qui est celle de rendre la distance plus grande entre le chirurgien et le patient. En effet, grâce au développement des réseaux de télécommunications, de nouvelles applications voient le jour. L'idée du Telementoring est d'apporter l'aide d'un chirurgien expert à un chirurgien distant moins expérimenté lors d'une intervention par robot. Tous deux sont équipés de moyens de visualisation similaires. L'opérateur est le chirurgien local, l'expert reste à distance, mais supervise

FIGURE 1.4. *Robot Da Vinci (Intuitive Surgical ©)*

l'intervention en cours. Cela implique une grande compétence du chirurgien local et permet à l'expert de diminuer le temps qu'il passe pour former ses confrères. Cette solution est principalement utile pour disposer de compétences élevées à la demande. Le *telementoring* a également sa place dans la médecine militaire dans la perspective d'amener une chirurgie performante sur les lieux de conflits. Par ailleurs, la visualisation à distance en temps réel du champ opératoire et du geste du chirurgien, lors de séances de formation dans une faculté de médecine avec la possibilité de dialoguer avec le chirurgien en direct, est d'une grande utilité pour l'apprentissage de la chirurgie. Enfin, l'un des objectifs de la chirurgie à distance, est l'accès aux meilleures compétences pour tous, entre villes ou entre pays. En résumé, les développements prévisibles de cette technique concernent :

- l'enseignement chirurgical puisqu'elle permet d'imaginer le maintien du contact entre un jeune chirurgien et une équipe chirurgicale plus experte,
- la possibilité, pour d'autres pays, de bénéficier de l'expertise d'équipes renommées permettant d'élever le niveau des soins,
- la perspective de réaliser des actes chirurgicaux dans des hôpitaux ne bénéficiant pas de beaucoup de moyens humains et matériels en mutualisant les moyens,
- la possibilité d'opérer des patients même dans des conditions difficiles (conditions climatiques, zones de conflits ou sinistrées).

Toutes ces perspecives donnent lieu à de nombreux sujets de recherche et de développement. C'est dans ce contexte que se situent les travaux de cette thèse.

1.1.3 Contexte des travaux

Nos travaux s'inscrivent dans le cadre du projet RALTT (Robotic Assisted Laparoscopic Telesurgery ans Telementoring). Ce projet a été initié en 2005 et réunit une équipe pluridisciplinaire du CHU de Nancy et du Centre de Recherche en Automatique de Nancy (CRAN, Nancy-Université) [1]. Il vise à étudier la faisabilité de certaines techniques chirurgicales par robot piloté manuellement à distance et les contraintes techniques occasionnées par une opération de chirurgie assistée par un robot, piloté à distance par un chirurgien à travers les réseaux. La problématique du projet se situe donc dans un contexte de **téléopération** et de **transmission de données**. Les principaux verrous technologiques se situent au niveau des temps de traitement des données vidéo et de leur transmission. La première intervention pratiquée, à distance, sur un patient, a été *l'opération Lindbergh* le 7 septembre 2001 (Figure1.5). Cette intervention sur un patient situé dans un bloc opératoire à Strasbourg a été réalisée par un chirurgien français, le Professeur Marescaux, depuis un bloc opératoire à New York (USA). Cette première a ainsi démontré la faisabilité technique de la chirurgie à distance. Cependant, cette cholécystectomie transatlantique a mobilisé d'importants moyens, tant technologiques que financiers, qui ne sont pas reproductibles au quotidien à l'heure actuelle. Dans cette opération, France Télécom avait dédié un réseau de fibres optiques doublé de deux répéteurs Satellite et 20 ingénieurs ont participé à la mise en œuvre de la transmission du flux vidéo de l'intervention. Outre le coût particulièrement élevé de cette opération, le problème technologique majeur est le délai entre l'acquisition vidéo du champ opératoire et l'affichage pour le chirurgien sur la console du robot. Ce délai enregistré au cours l'opération Lindbergh était estimé à 155 ms [Butner et Ghodoussi, 2003] sur un réseau de fibres optiques dédié et sur une distance parcourue par les données de 15000 kilomètres. Le temps de latence dépend du réseau sur lequel on transmet les données, sa valeur sera plus grande dans le cas d'une trasmission par satellite que dans le cas d'un réseau local. Les travaux de [Fabrlzio *et al.*, 2000] ont permis de quantifier et de démontrer les effets de l'augmentation du temps de latence sur la performance du chirurgien (en chirurgie laparoscopique). En effet, les erreurs commises par les opérateurs augmentent et la précision requise pour

1. Le projet RALTT est cofinancé par la Région Lorraine et la Communauté Urbaine du Grand Nancy

effectuer un geste chirurgical à distance diminue avec l'augmentation du délai (temps de latence). Dans un contexte d'opération à grande distance, de télé-enseignement ou encore de télé-conseil (Telementoring), la transmission à travers les réseaux de communication des volumineux flux vidéo nécessite des capacités très importantes en termes de bande passante, rendant la compression avec perte de ces flux incontournable.

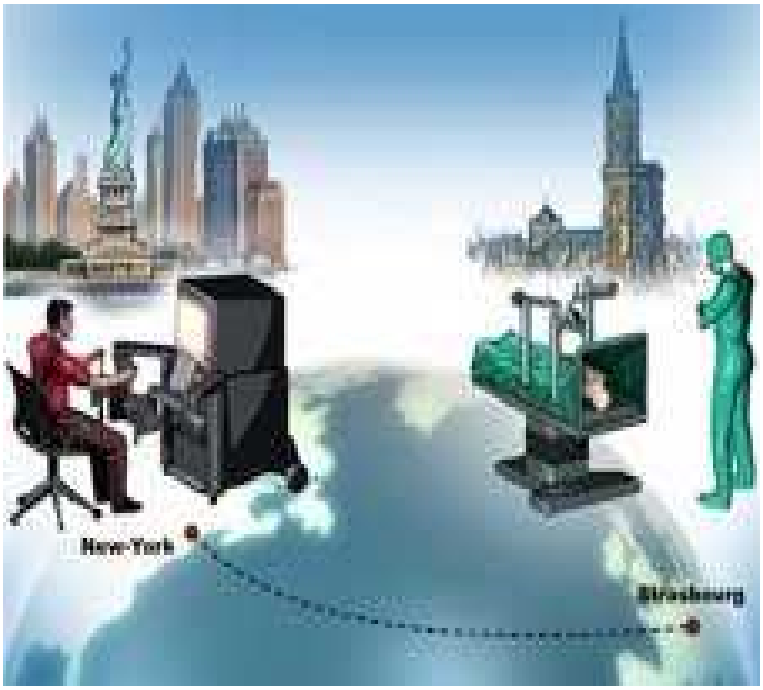


FIGURE 1.5. *Première opération assistée par robot à distance : De Strasbourg à New York*

1.1.4 Nécessité de la compression de données

Actuellement, un chirurgien opère son patient en étant à quelques mètres du robot, ou parfois dans une autre pièce. En effet, sa console de travail est reliée au robot par un réseau local dédié permettant une transmission de volumes importants de données. Cependant, malgré le développement des télécommunications à très haut débit, notamment grâce à la fibre optique, et les performances accrues des technologies de transmission sur les réseaux, la compression **avec pertes** des données est inévitable ici, en particulier pour la transmission de vidéos du fait des volumes générés. A titre d'exemple, un débit de 270 Mbits/s est généré pour la transmission d'une vidéo de résolution SD (Standard De-

finition : 720 x 576) et 1,5 Gbits/s pour une vidéo de résolution HD (High Definition : 1280 x 720). Lorsque le taux de compression augmente, les encodeurs introduisent dans la vidéo compressée des artefacts pouvant affecter sa qualité visuelle. Dans le contexte de l'imagerie médicale et en particulier celui de la télé-chirurgie, cette altération de la qualité peut être rédhibitoire. Pour les images fixes, la communauté médicale a longtemps préféré la compression **sans perte** plutôt que la compression **avec pertes** qui permet pourtant d'atteindre des taux de compression bien plus importants. Cependant, l'Association Canadienne de Radiologie (CAR) a publié une norme qui valide l'utilisation de la compression **avec pertes irréversibles** dans certaines circonstances définies et pour des types d'examens médicaux précis. Des recommandations précises sur la mise en œuvre de la compression avec pertes sont définies notamment concernant la supervision et la validation par des radiologistes experts des taux de compression utilisés [Car, 2010]. A notre connaissance, la question de la compression avec pertes ne se pose pas pour les vidéos médicales, moins utilisées dans la pratique quotidienne mais les pertes restent toujours un sujet sensible dans un domaine où la perte de la qualité constitue une gêne pour le praticien et pose également des problèmes juridiques.

1.2 Problématiques sous-jacentes

La figure 1.6 représente étape par étape le déroulement d'une téléopération chirurgicale. Le geste effectué par le chirurgien **(1)** sur les manettes de sa console est émis sous forme d'une commande **(2)** sur le réseau dédié reliant la console au bras articulé du robot **(3)**. Cette commande reçue permet d'effectuer le geste demandé **(4), (5)**. Le chirugien visualise son geste sur le champ opératoire à travers le viseur binoculaire en 3D. Tout au long de l'opération, les données de la caméras binoculaire, constituées de deux flux vidéo de 270 Mbits/s chacun, en résolution SD (Standard Definition), sont transmises sur le réseau après leur synchronisation **(6), (7), (8)**. En effet, la caméra binoculaire étant composée de deux caméras distinctes **9**, chacun des deux flux est indépendant. Afin de visualiser correctement en 3D le champ opératoire, la synchronisation des deux vues de la caméra binoculaire est nécessaire **8**.

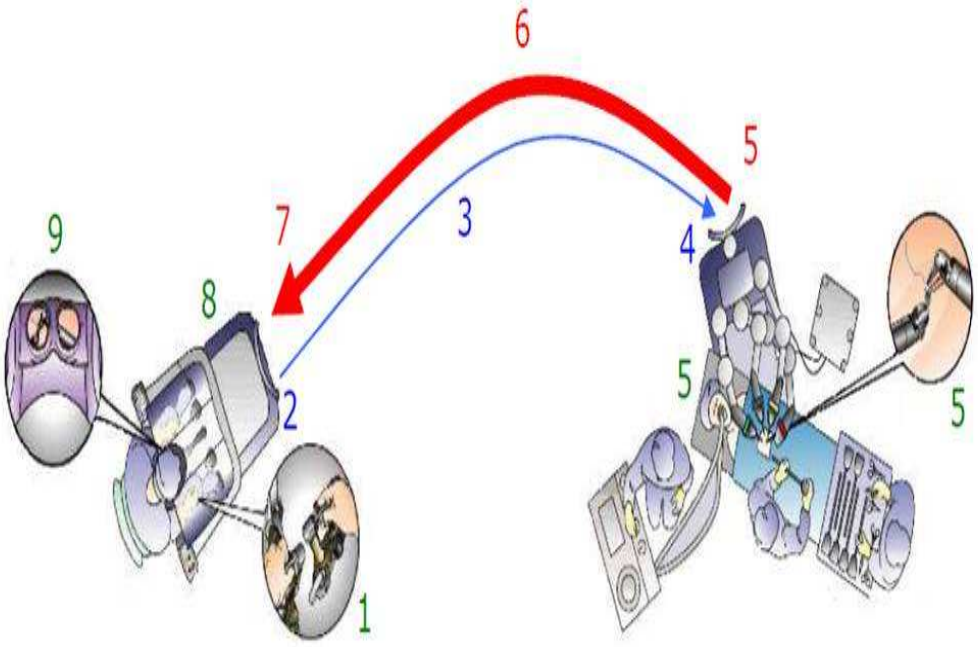


FIGURE 1.6. *Schéma général d'une téléopération*

Dans ce qui suit, nous présentons trois problématiques essentielles : la **qualité de service** liée à la transmission des **vidéos médicales**, la **qualité** de la vidéo et la **synchronisation des flux stéréoscopiques**.

Transmission des vidéos médicales : qualité de service Une téléopération (Figure 1.8) est une application temps réel qui nécessite pour son bon déroulement une qualité de service (QoS) de bout en bout de la chaîne de transmission . Lors d'une téléopération, une séquence video va subir trois étapes : encodage, transmission et décodage. Suite à ces traitements, la séquence vidéo originale passera par quatre états (originale, encodées, transmise et décodée) durant lesquels plusieurs dégradations peuvent être introduites. Il s'agit d'abord des dégradations apportées par l'algorithme de compression avant la transmission dont les dégradations apportées peuvent être considérables (pertes de paquets, bruit sur le canal de transmission). Quand les défauts de transmissions sont grands, le décodeur , en bout de chaîne, va légèrement améliorer la qualité de décodage

avec des méthodes de masquage d'erreur.

Au final, la qualité de la vidéo perçue par l'utilisateur est donc un cumul de l'ensemble des dégradations subies sur toute la chaîne de transmission. Par ailleurs, le temps de latence mesuré entre le geste du chirurgien et le retour d'image du geste effectué, est un paramètre qui doit être pris en compte. En effet, les résultats obtenus lors de l'opération « Lindberg », ont montré qu'une limite du temps de latence acceptable pour effectuer un geste chirurgical à distance est fixée à 200 ms. Le temps de latence global est un cumul des durées de traitement de chaque étape de la chaîne de transmission. Il dépend aussi de la bande passante du réseau et de la complexité des techniques d'encodage et de décodage utilisées. La figure 1.7 (inspirée de [Zhang *et al.*, 2006]) synthétise l'inter-dépendance entre la qualité de la vidéo et le temps de latence.

Un autre problème qui peut avoir une influence sur la vidéo transmise sur un réseau IP est la gigue de transmission. La gigue est la variation du délai de transmission. Le protocole utilisé pour transporter les paquets sur un réseau IP est UDP (User Datagram Protocol). Le protocole UDP fonctionne en mode non connecté : les paquets n'empruntent pas forcément le même chemin, d'où une variation du délai de transit. Une autre cause de la variation de ce délai est le nombre de routeurs traversés et de la charge de chacun d'entre eux. Pour restituer un flux synchrone à l'arrivée, il est possible de prévoir des buffers de compensation de gigue ; mais ce stockage allonge encore le délai de transmission (temps de latence supplémentaire). La gigue doit rester inférieure à 100 ms pour garantir une qualité acceptable. Par ailleurs, le protocole UDP ne garantit pas que les paquets arrivent à destination. Une erreur sur l'en-tête du paquet peut entraîner sa perte ou l'envoi vers une mauvaise destination. D'autre part, lorsque les routeurs IP sont congestionnés, ils libèrent automatiquement de la bande passante en détruisant une proportion des paquets entrants en fonction de seuils prédéfinis. Le taux de perte des paquets dépend de la qualité des lignes empruntées et du dimensionnement du réseau.

Dans ce qui suit, nous évoquerons les problématiques liées à la **qualité** de la vidéo et la **synchronisation des flux stéréoscopiques**.

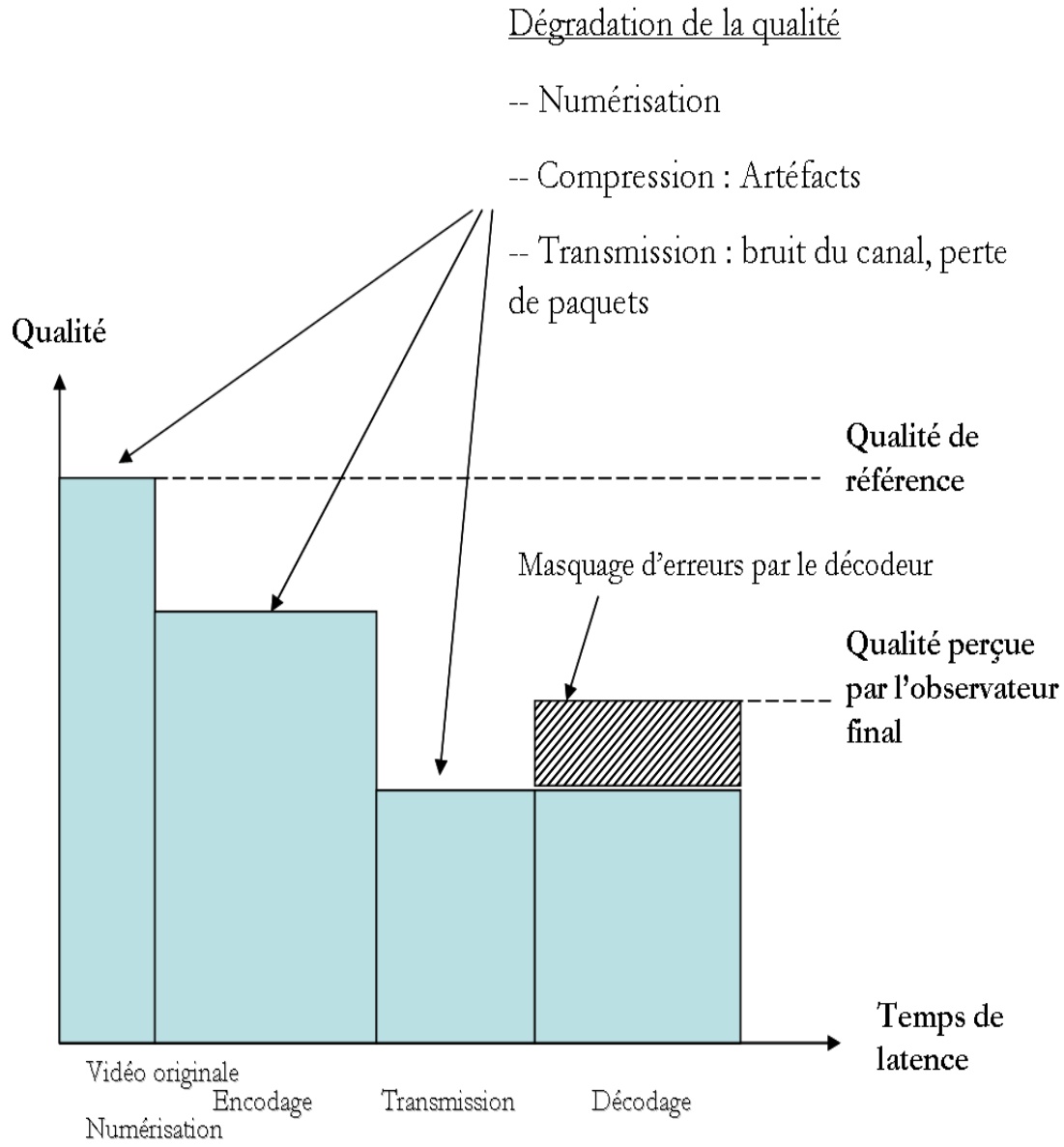


FIGURE 1.7. *Dégradation de la vidéo et temps de latence*

Qualité de la vidéo L'évaluation de la qualité des vidéos chirurgicales compressées, dans un contexte de téléopération, est primordiale pour garantir la sécurité du patient et le confort du chirurgien. Dans une application aussi sensible, il s'avère préférable de se référer au jugement humain pour estimer la qualité des informations perçues visuellement par le chirurgien [ITU-R, 2000]. Le terme de qualité de vidéo ou de critère qualité

n'est pas un terme intuitif, il est donc nécessaire de préciser sa signification. La notion de qualité vidéo est indispensable pour caractériser le besoin selon les applications. Cette notion est donc relative et dépend du besoin en aval. La qualité est un ensemble de perceptions du degré global d'excellence de l'image ou de la vidéo [Engeldrum, 2000]. Par exemple, dans le contexte de l'imagerie médicale, la **qualité** se rapporte à la capacité au diagnostic et des protocoles spécifiques pour effectuer cette évaluation. Dans la plupart des domaines, la qualité de l'image ou vidéo est un constat « esthétique » selon les perceptions de l'observateur. Ce sont des attributs perceptuels (généralement visuels) sur lesquels l'observateur s'appuie pour juger la qualité : celui-ci ne se base pas sur des variables techniques mais sur les attributs intrinsèques de l'image (netteté, couleurs, luminosité, brillance). Dans notre cas, nous utiliserons le terme **qualité vidéo** pour désigner la qualité subjective de la vidéo (qualité perceptuelle) [Arnaud *et al.*, 2009].

Dans ce qui suit, nous évoquerons les problématiques liées à la **synchronisation des flux stéréoscopiques**.

Synchronisation des deux flux stéréoscopiques

Principe des vidéos stéréoscopiques Les vidéos stéréoscopiques, sont un cas particulier des vidéos multi-vues, puisqu'elles ne sont obtenues que par deux caméras constituant une vue gauche et une vue droite. Afin de restituer la perception 3D, il existe des algorithmes spécifiques de reconstruction [Theobalt *et al.*, 2007], [Desurmont *et al.*, 2007]. Cependant, dans le cas de la chirurgie robotique, la restitution se fait par un viseur binoculaire. La vision stéréoscopique est basée sur la disparité binoculaire du système visuel humain. Celui-ci produit deux images légèrement différentes qui sont projetées sur la rétine (yeux) avant d'être fusionnées dans le cortex (cerveau) pour ne représenter qu'une seule image 3D. L'utilisation de deux caméras, dans certaines applications et notamment dans le cas de la chirurgie robotique, permet de reproduire ce fonctionnement. En effet, les deux caméras sont positionnées avec la même distance inter-oculaire humaine (en moyenne cette distance vaut 65 mm). Chaque caméra correspond à un œil de l'observateur. Selon la technique d'affichage, la bonne image est présentée devant chaque œil (par filtrage).

Transmission des flux stéréoscopiques Sur un réseau à commutation de paquets, les données sont découpées en paquets afin d'accélèrer leur transfert. Chaque paquet est composé d'un en-tête contenant des informations sur le contenu du paquet ainsi que sur sa destination, permettant ainsi au commutateur d'aiguiller le paquet sur le réseau vers son point final. Cependant, l'inconvénient majeur réside dans la désynchronisation temporelle au moment de la transmission sur un réseau (à commutation de paquets) perturbant ainsi la visualisation et la perception 3D en temps réel. En effet, le récepteur d'un flux stéréoscopique doit être en mesure de synchroniser la vue de droite et la vue de gauche du flux stéréoscopique. Ceci est très important, car si deux flux ne sont pas synchronisés, les objets mobiles dans une scène peuvent être perçus comme ayant des valeurs de parallaxe fausses, leur position observée serait différente de leur position réelle. La parallaxe est le déplacement entre les deux projections d'un même point physique, d'un point de vue d'observation à l'autre. Les caméras stéréoscopiques sont habituellement ajustées pour que la parallaxe correspondant à un point de 65 mm, c'est-à-dire la distance moyenne entre les pupilles d'un adulte. Les étiquettes « temps » du protocole RTP (Real Time protocol) [Schulzrinne *et al.*, 1996] peuvent être utilisées pour synchroniser les deux flux s'ils proviennent de la même horloge, ce qui n'est possible que si les deux flux proviennent du même serveur émetteur. Par ailleurs, le récepteur peut mettre en correspondance les étiquettes « temps » de RTP avec les étiquettes NTP (Network Time Protocol) [Mills, 1992] qui permet de synchroniser, via le réseau, l'horloge locale du récepteur ou de l'émetteur sur une référence d'heure.

Compression des flux stéréoscopiques La compression des deux flux stéréoscopiques pour leur transmission est incontournable. Pour compresser les flux stéréoscopiques, une approche possible est d'exploiter les redondances inter-vues. Parmi les algorithmes basés sur cette approche, on peut citer MPEG-2 multi-view profile [Puri *et al.*, 1997]. [Oh *et al.*, 2004] propose des modifications de l'algorithme précédent pour augmenter son efficacité en se basant sur les corrélations entre les deux vues stéréoscopiques. Quant à [Darazi *et al.*, 2009], il propose d'encoder conjointement la vue de droite et la vue de gauche en exploitant de façon optimale la corrélation existante entre ces deux vues. Ceci grâce

à une nouvelle transformée qui s'inspire des schémas de lifting. L'étape de prédiction dans cette approche est remplacée par une étape hybride : le calcul des cartes de disparité est suivi d'une correction de la luminance et une prédiction optimale. Une autre approche possible est d'exploiter les propriétés du système visuel humain (SVH). En effet, pour la compression des vidéos monoscopiques, il est commun de sous-échantillonner les canaux de chrominance de la vidéo car le SVH est moins sensible aux variations de la chrominance. D'une façon similaire, le SVH peut percevoir l'information haute fréquence d'une des deux vues dans la vidéo 3D même si la deuxième vue est filtrée par un filtre passe-bas [Stelmach *et al.*, 2000]. Des méthodes combinant les deux approches ont également été proposées [Aksay *et al.*, 2007]

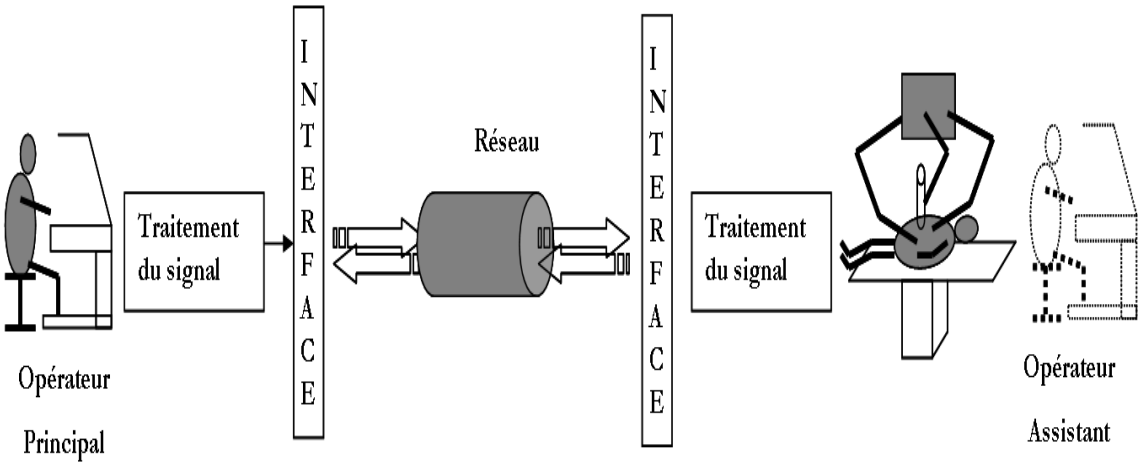


FIGURE 1.8. *Principe de la téléchirurgie*

Dans cette thèse, nous nous focaliserons sur le seuil de compression d'un seul flux stéréoscopique toléré par les chirurgiens à travers un test subjectif de la qualité des vidéos chirurgicales, d'une part. Nous démontrerons la faisabilité de la transmission en temps réel sur un réseau standard de flux compressés à bas débit, d'autre part.

1.3 Objectifs scientifiques

Cette thèse se place résolument dans un contexte de transmission **à distance** de flux vidéo issus de la chirurgie assistée par robot, avec toutes les contraintes associées : le temps de latence, la qualité de service réseau et la qualité perceptuelle de la vidéo.

La première contribution de ces travaux consiste à déterminer un **seuil de compression vidéo** toléré par les chirurgiens, dans un contexte de **télérobotique chirurgicale**. L'hypothèse de base étant que la qualité globale d'une vidéo stéréoscopique dépend de la vue de meilleure qualité [Stelmach *et al.*, 2000]. Nos travaux se sont donc naturellement focalisés sur l'étude de techniques permettant l'évaluation de la qualité des vidéos dans un contexte de robotique chirurgicale. Deux approches méthodologiques sont possibles : l'une à caractère subjectif et l'autre à caractère objectif. Pour ce qui est de la compression des vidéos chirugicales, deux pistes sont possibles : MPEG-2 ou H.264. Si le choix du codeur MPEG-2 a été motivé à la fois par la maturité de cette norme de compression et pour ses bonnes performances en termes de débit-distorsion, la norme H.264, plus récente, a démontré des performances jusqu'à deux fois supérieures à celles de MPEG-2 à débit égal. Dans le chapitre 4, nous décrirons les travaux [Nouri *et al.*, 2010], développés autour du standard MPEG-2, et qui ont permis de mettre en exergue la possibilité de compresser des vidéos médicales en identifiant le débit seuil à partir duquel une opération chirurgicale à distance est envisageable tout au moins concernant la qualité des vidéos compressées MPEG-2 en vue de leur transmission. Nous présenterons également les résultats de l'étude portant sur la qualité des vidéos chirurgicales compressées H.264. A notre connaissance, il n'existe pas dans la littérature de travaux similaires, traitant de la qualité des vidéos dans le domaine chirurgical compressées avec pertes. En effet, le milieu médical utilise la compression avec perte qui permet d'atteindre des taux de compression importants. Différents travaux émergent cependant aujourd'hui montrant qu'il existe une certaine tolérance aux pertes en imagerie médicale, y compris dans le domaine très sensible de la radiologie [Gaudeau et Moureaux, 2009], [Schelkens *et al.*, 2003]. L'ACR (American College of Radiology) recommande d'ailleurs à présent l'utilisation de techniques de compression avec perte sous la responsabilité d'un praticien qualifié.

La seconde contribution de ce travail est de démontrer la **faisabilité de la transmis-**

sion en **temps réel**, sur un **réseau IP** avec une **bande passante partagée** par plusieurs utilisateurs, de flux vidéos compressés provenant d'un bloc opératoire avec une qualité de service de bout en bout. Cette application directe, décrite dans le chapitre 5, s'appuie sur les seuils de compression déterminés dans la première partie de ces travaux.

1.4 Conclusion

La vocation de ce chapitre était de mettre en exergue les problématiques liées à la transmission en temps réel de flux vidéo issus d'un robot de chirurgie. Dans un contexte d'opération à grande distance, de télé-enseignement ou encore de télé-conseil (Telementoring), la transmission à travers les réseaux de communication des volumineux flux vidéo nécessite des capacités de bande passante très importantes, rendant la compression avec perte de ces flux incontournable. Cependant, lorsque le taux de compression augmente, les encodeurs introduisent dans la vidéo compressée des artefacts pouvant affecter sa qualité visuelle. Dans le contexte de l'imagerie médicale et en particulier celui de la télé-chirurgie, cette altération de la qualité peut être rédhibitoire. Nous aborderons la question de la qualité dans le chapitre 3. Les études subjectives de la qualité des vidéos fournissent des données intéressantes pour évaluer la performance des méthodes d'évaluation objective de la qualité. L'objectif est donc d'établir des corrélations entre la perception subjective de la qualité par un panel d'experts et les mesures objectives développées dans la littérature. Par ailleurs, les résultats obtenus suite aux essais subjectifs de la qualité ont permis également de mettre en exergue une corrélation entre les mesures subjectives effectuées et une mesure objective utilisant l'information structurelle de l'image (métrique SSIM). Ceci permet de prédire la qualité telle qu'elle est perçue par les observateurs humains.

Chapitre 2

Compression vidéo : principes généraux et conséquences

Le chapitre précédent a montré la nécessité de la compression des vidéos médicales et les enjeux liés à l'application de télé-chirurgie robotique. Les données visuelles, notamment les vidéos, demandent des ressources très importantes quant à la bande passante des réseaux de transport et à l'espace de stockage nécessaires pour leur traitement, leur compression devient incontournable. Les méthodes de compression de vidéos sont actuellement de plus en plus nombreuses et efficaces. Ce chapitre est consacré aux principes de la compression de vidéos ainsi qu'aux conséquences de cette opération (distorsions inhérentes à la compression). Nous décrivons également d'autres distorsions des vidéos liées à la transmission ou aux traitements des séquences, et en particulier issues la conversion analogique/numérique des séquences.

2.1 Principes fondamentaux de la compression vidéo

Un flux vidéo est composé d'une succession d'images qui défilent à un rythme fixe (25 par seconde dans la norme française ou 30 par seconde dans d'autres normes) pour donner l'illusion du mouvement. Chaque image est décomposée en lignes horizontales, chaque ligne étant une succession de points. La lecture et la restitution d'une image s'effectue donc séquentiellement ligne par ligne comme un texte écrit : de gauche à droite puis de haut en bas. La compression vidéo consiste à réduire la quantité de données de

la séquence en limitant l'impact sur la qualité visuelle de la vidéo et du son. L'intérêt est de diminuer les coûts de stockage et de transmission des fichiers vidéo. Les informations pas ou peu perceptibles par le système visuel humain sont supprimées. La perte d'information est irréversible.

2.1.1 Codage des couleurs

Les standards vidéo analogique comme PAL (Phase Alternating Line), NTSC (National Television System Committee) ou numérique comme la famille MPEG (Moving Picture Experts Group) ou DV (Digital Video), se basent sur le système visuel humain pour représenter et traiter l'information de couleur contenue dans la vidéo. En particulier, l'acuité visuelle pour la chrominance est plus faible. Il en résulte que les paramètres de chrominance peuvent être transmis avec moins de détails que la luminance pour une économie de bande passante. La chrominance désigne la partie de l'image correspondant à l'information de couleur, fournie à partir des 3 couleurs primaires : rouge, vert et bleu par synthèse additive. Le modèle YUV [Wharton et Howorth, 1967] définit un espace colorimétrique à trois composantes. La première représente la luminance (Y) et les deux autres représentent la chrominance (U et V). YUV est utilisé par le standard PAL. La réduction des données est possible par sous-échantillonnage des composantes de chrominance. Les différents formats de sous-échantillonnage sont les suivants :

- 4 :4 :4 ce format n'est pas sous-échantillonné,
- 4 :2 :2 ce format est sous-échantillonné d'un facteur 2 horizontalement. Il est défini par exemple par l'IUT-R BT. 601-5,
- 4 :2 :0 ce format est sous-échantillonné d'un facteur 2 horizontalement et verticalement. Il s'agit de l'approximation la plus proche du comportement du système visuel humain et il est utilisé par exemple dans JPEG et MPEG,
- 4 :1 :1 ce format est sous-échatillonné d'un facteur 4 horizontalement.

2.1.2 Méthodes de compression vidéo

Nous décrivons ici le contenu hiérarchisé d'un flux vidéo type MPEG. En effet, les standards MPEG se veulent génériques, c'est pour cela qu'une syntaxe du flux a été dé-

finie. Le schéma du décodeur MPEG pouvait ainsi être standardisé laissant aux développeurs le soin de mettre en oeuvre les encodeurs.

2.1.2.1 Hiérarchisation du flux vidéo

Une séquence MPEG est structurée de la façon suivante :

- **GOP :** le GOP (Group of picture) est un ensemble d'images ;
- **Image :** l'image est un ensemble de pixels et un espace à trois dimensions constitué d'une composante d'intensité et de deux composantes chromatiques ;
- **Tranche :** chaque image est divisée en tranches, qui sont un ensemble de macroblocs adjacents. Chaque tranche est un élément important dans la gestion des erreurs ; le décodeur peut sauter une tranche et passer à la suivante si une erreur se présente ;
- **Macrobloc (MB) :** un macrobloc est une matrice contenant des blocs. Le nombre de blocs dans un MB de luminance est de 4 blocs pour MPEG-2 et dépend du sous-échantillonnage effectué pour la chrominance ;
- **Bloc :** un bloc est un tableau de $n \times m$ pixels de luminance ou de chrominance. Les étapes de compression spatiale et temporelle vont être appliquées au bloc.

Dans la plupart des méthodes de compression, on exploite quatre types de redondances : **spatiales** (transformée), **subjectives** ou psychovisuelles (quantification), **statistiques** (codage) et **temporelles** (compensation du mouvement).

2.1.2.2 Redondance spatiale

Il existe des redondances au sein de chaque image du flux vidéo. L'étude de ces redondances se fait en deux étapes : après une **décomposition en blocs** de l'image, un **passage dans le domaine fréquentiel** est effectué à travers une transformée mathématique par exemple la Transformée en Cosinus Discrète (DCT).

Décomposition en blocs Le passage dans le domaine fréquentiel de l'image ainsi que la compression temporelles ne sont pas effectués en une fois sur l'image, mais plutôt sur des blocs (de taille 8x8 pour MPEG-2 et de taille variable pour H.264). La décomposition de l'image d'abord en macroblocs, ensuite en blocs est donc effectuée en amont de

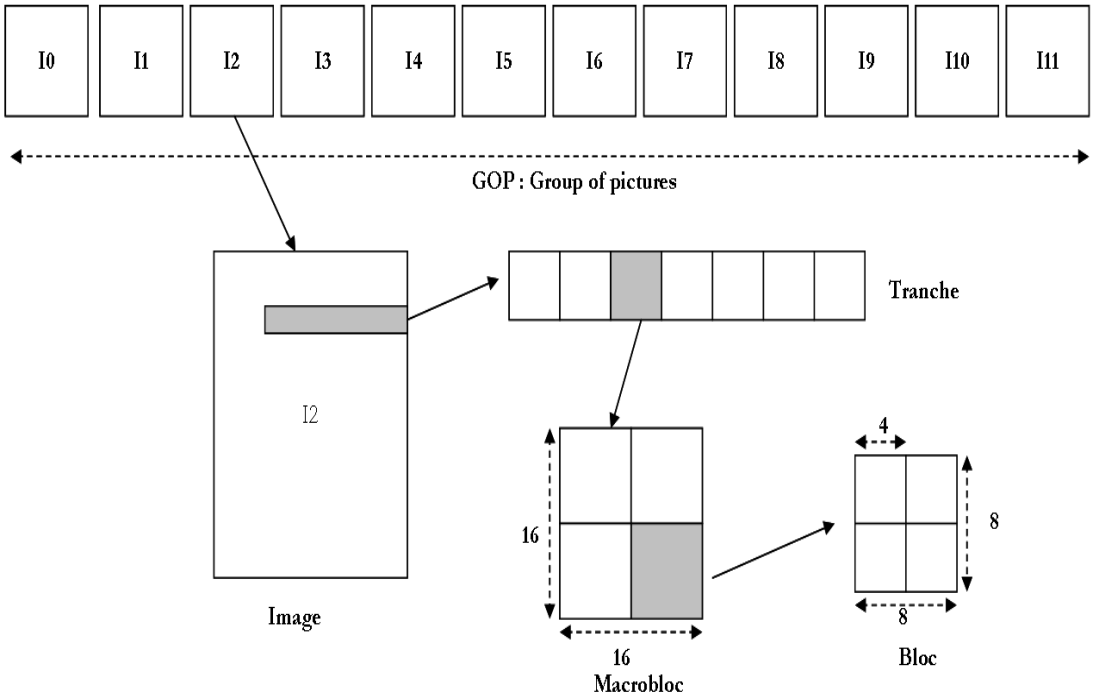


FIGURE 2.1. *Hiérarchie du flux vidéo*

la compression.

Passage dans le domaine fréquentiel La compression spatiale se base sur une propriété du système visuel humain (SVH) : le SVH est moins sensible aux zones de hautes féquences et se satisfait d'une résolution assez faible pour les détails d'une image. Afin de mettre en évidence ces zones de l'image, le passage dans le domaine fréquentiel se fait au moyen d'une transformée mathématique linéaire, le plus souvent une Transformée en Cosinus Discrète (DCT) qui permet de présenter une carte des fréquences et amplitudes de chaque bloc (motifs). Pour un bloc de taille NxN, la DCT s'exprime selon la formule :

$$DCT\left(x,y\right)=\frac{2}{N}C\left(i\right)C\left(j\right)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}p\left(i,j\right)\cos\left[\frac{\left(2x+1\right)i\pi}{2N}\right]\cos\left[\frac{\left(2y+1\right)j\pi}{2N}\right] \qquad (2.1)$$

où C est une constante : C = $\left\{\begin{array}{l}\frac{1}{\sqrt{(2)}}\ \text{si}\ x,y=0\\ 1\ \ \text{si}\ x,y>0\end{array}\right\}$

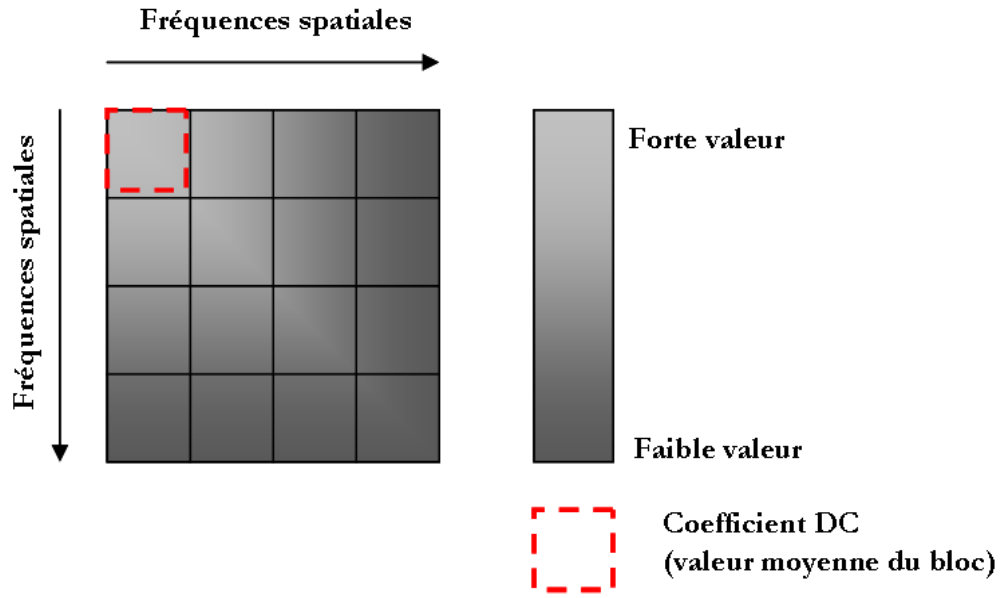


FIGURE 2.2. *Transformation d'un bloc*

Le spectre calculé par la DCT peut être représenté par un bloc fréquentiel de la même taille que le bloc spatial initial, comme l'illustre la figure 2.2.

2.1.2.3 Redondance subjective

La perception des hautes fréquences contenues dans une image étant naturellement atténuées par le SVH, celles-ci seront donc codées (et quantifiées) avec moins de finesse. Un seuillage est effectué pour éliminer les coefficients inférieurs à un seuil fixé par le standard de compression utilisé. Des tables de quantification sont employées pour associer à chaque valeur du coefficient DCT son équivalent quantifié. Le but étant d'éliminer les informations non visibles pour le SVH, les coefficients de DCT les plus significatifs représenteront le bloc. Puisque les informations de basses fréquences sont plus pertinentes que les informations de hautes fréquences, une pondération ad hoc doit être réalisée. Pour cela, on utilise une matrice de quantification contenant des entiers par lesquels seront divisées les valeurs de la matrice DCT. Cette matrice réalise ainsi le filtrage des hautes fréquences. Le choix de la matrice de quantification dépend donc des caractériques du

SVH (redondances subjectives) et détermine le taux de compression. Il est à noter que la perte de données engendrée par la quantification est irréversible.

2.1.2.4 Redondance statistique

Après la quantification des données dans un ensemble fini de valeurs, on peut coder ces données sans pertes en exploitant les redondances statistiques entre les coefficients quantifiés. Le codage entropique est souvent utilisé dans cette étape. Il se base sur l'occurrence des symboles.

Le balayage en zig zag Afin de faciliter le codage des valeurs quantifiées dans un bloc, un balayage en zig-zag est effectué en partant du coin en haut à gauche du bloc (cf. Figure 2.3). Ceci a pour objectif d'organiser les coefficients quantifiés de telle sorte que les plus significatifs soient transmis en premier suivis par les valeurs nulles.

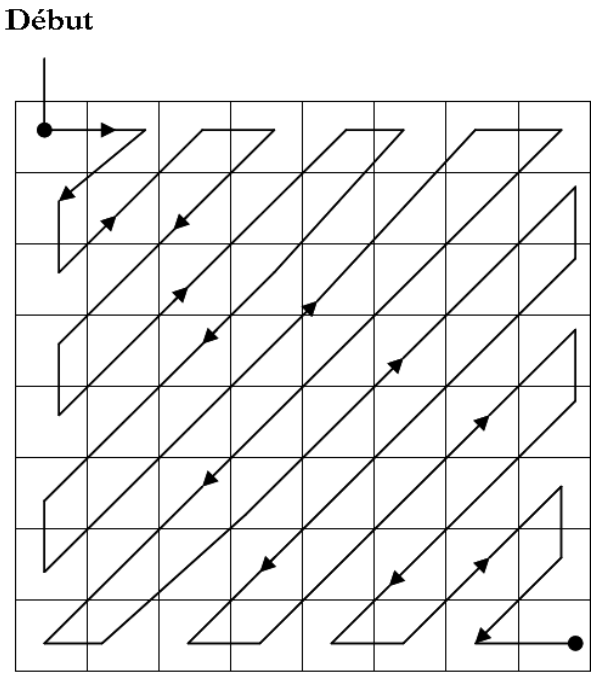


FIGURE 2.3. *Balayage en zig-zag*

Le codage par plage La trame des coefficients quantifiés et organisés après le balayage en zig-zag est codée de la manière suivante : toute suite de coefficients identiques est remplacée par un couple (nombre d'occurrences ; coefficent répété). Il s'agit du codage RLC (Run Length Coding).

Le codage entropique Le codage entropique est une méthode de codage de source sans pertes. Parmi les codes entropiques, on peut citer le codage de Huffman et le codage arithmétique.

Le codage entropique utilise les statistiques de la source pour construire un code dont la longueur dépend de la fréquence d'occurrence des coefficients. On construit un code à longueur variable, qui attribue les mots de codes les plus courts aux coefficients les plus fréquents et inversement.

2.1.2.5 Redondance temporelle

Les méthodes de compression temporelle exploitent les redondances qui existent entre les images successives dans une séquence en ne transmettant que les différences entre deux images consécutives (Figure 2.4). Cette compression s'applique souvent à un groupe d'images (GOP) (*cf.* figure 2.5) de la séquence vidéo. Elle crée à partir des images Intra I du GOP, des images prédites P et bidirectionnelles B grâce aux étapes de compensation de mouvement. Un GOP est constitué d'images I codées en mode intra c'est-à-dire que seules les données spatiales de l'images sont exploitées et codées par compression spatiale. L'image I d'un GOP ne dépend d'aucune autre image et constitue donc un point de référence pour le décodage. Les images P sont quant à elles prédites à partir d'images précédentes I ou P. Il s'agit donc d'un codage du déplacement par rapport à l'image de référence (vecteur mouvement). Elle sert également de référence pour les images B qui sont obtenues par interpolation bidirectionnelle du vecteur mouvement provenant des images passées ou futures (I ou P). Elles jouent un rôle dans la réduction du débit d'un flux vidéo.

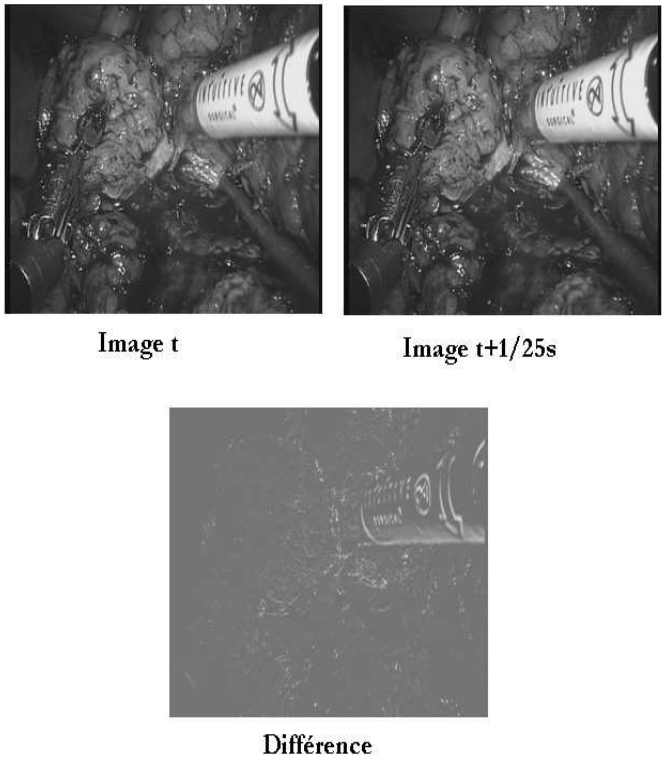


FIGURE 2.4. *Différence entre deux images successives dans une séquence*

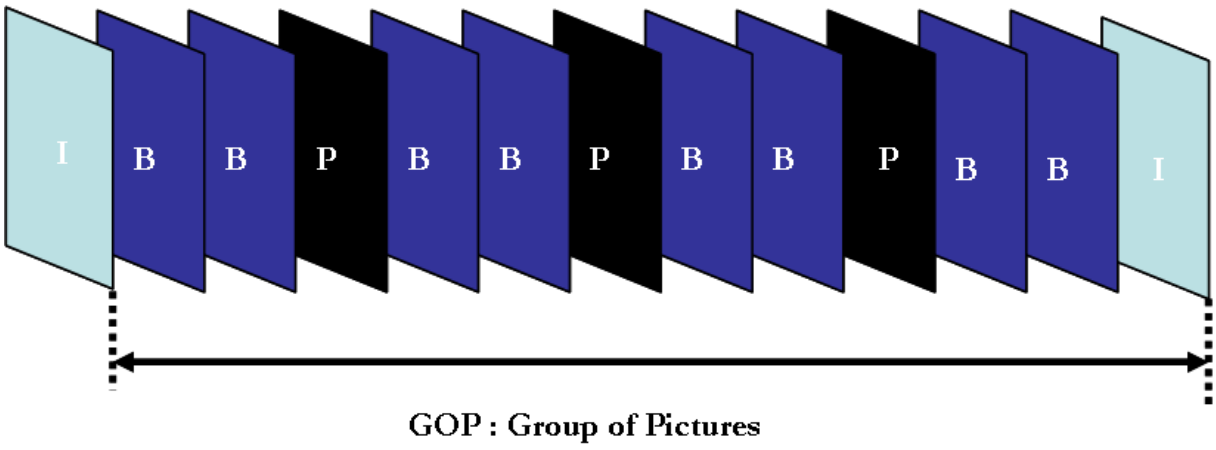


FIGURE 2.5. *GOP de longueur 12*

La compensation du mouvement Les images I permettent de créer les images P et B par une recherche préalable de macroblocs (MB) semblables entre une image et celle qui la précède. Ensuite, la caractérisation des déplacements dans le MB se fait par codage des vecteurs mouvement. Ceci va permettre la construction d'une image prédite qui sera comparée à l'image originale pour générer des données d'erreurs de prédiction. Enfin, seuls les vecteurs et les erreurs de prédiction seront codés et transmis. Une mauvaise estimation du vecteur mouvement peut introduire des pertes d'informations. Cependant, une prédiction de l'erreur obtenue entre un MB estimé et le MB de référence peut minimiser cette erreur. Dans le cas des codeurs vidéos non basés sur une transformée DCT (transformée en ondelettes par exemple) [Cagnazzo *et al.*, 2007], d'autres critères plus complexes d'estimation de mouvement sont préconisés.

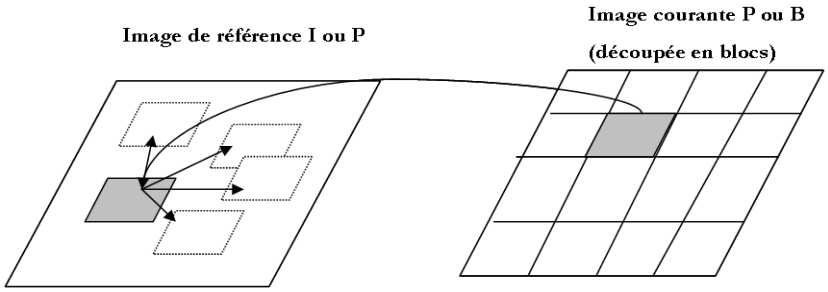


FIGURE 2.6. *Estimation et compensation de mouvement*

2.1.3 Standards de compression

Le groupe MPEG (Moving Picture Experts Group)) est un groupe de travail de l'ISO/IEC (International Organization for Standardization / International Electrotechnical Commission) chargé du développement des standards internationaux de compression, décompression, traitement des images animées, de l'audio et de leurs combinaison. MPEG regroupe quelques standards bien connus pour la compression vidéo. L'activité de ce groupe de travail a permis de produire les normes ci-dessous, nous nous focaliserons ensuite sur les deux standards les plus répandus à savoir MPEG-2 et MPEG-4/AVC ou

H.264.

- MPEG-1, est une norme pour le stockage et l'extraction de la vidéo et de l'audio, qui a été approuvée en 1992. La normalisation n'a concerné que la syntaxe de codage et le schéma de décodage, ce qui en fait une norme générique. Elle définit une technique de codage DCT/DPCM (Discrete Cosine Transform / Differential pulse code modulation) basée blocs avec prédiction et compensation de mouvement ;
- **MPEG-2,** est le standard le plus utilisé à des fins commerciales. Il est par exemple le standard des DVDs mais aussi de la télévision numérique. La technique de codage proposée est également générique, il s'agit d'un raffinement de MPEG-1. Des conditions particulières sont prévues pour les sources entrelacées. La scalabilité est introduite. Les profils et les niveaux permettent de définir les différentes classes de conformité MPEG-2 ;
- MPEG-4, est une norme qui spécifie d'abord des techniques pour gérer le contenu de scènes comprenant un ou plusieurs objets audio-vidéo. Contrairement à MPEG-2 qui visait uniquement des usages liés à la télévision numérique (diffusion DVB et DVD), les usages de MPEG-4 englobent toutes les nouvelles applications Multimédias comme le téléchargement et le streaming sur Internet, le multimédia sur téléphone mobile, la radio numérique, les jeux vidéo, la télévision et les supports haute définition ;
- **MPEG-4 part 10 AVC (Advanced Video Coding),** il s'agit du standard actuellement utilisé dans les applications Multimédia connu aussi sous le nom H.264 (ITU-T Rec. H.264, 2003). Ce dernier est utilisé dans des applications allant de la vidéo sur mobile à la télévision haute-définition. Il est basé sur la même approche de compression que la famille MPEG. Quelques fonctionnalités supplémentaires ont été rajoutées à ce standard. Elles incluent des tailles de bloc plus petites, une prédiction spatiale et temporelle plus adaptative et une insertion dans la boucle d'encodage d'un filtre anti-bloc pour réduire la visibilité des artéfacts de type effet de bloc (cf. section suivante). Toutes ces améliorations permettent à H.264 d'avoir approximativement deux fois plus d'efficacité de codage que les autres standards de compression MPEG ;

- MPEG-7, est une norme qui permet de décrire les contenus Multimédia ;
- MPEG-21, est une norme proposant une architecture pour l'interopérabilité et l'utilisation simple des contenus Multimédia.

Dans ce qui suit, nous décrivons les deux standards que nous avons retenus pour la compression des vidéos chirurgicales : MPEG-2 et H.264.

MPEG-2 : Schéma général MPEG-2 est la norme de seconde génération (1994) du Moving Picture Experts Group qui fait suite à MPEG-1. MPEG-2 définit les aspects compression de l'image et du son [Bosi, 1997] et le transport à travers des réseaux pour la télévision numérique. Les aspects Systèmes (synchronisation, transport, stockage) sont définis dans la norme ISO/CEI 13818-1 (Codage générique des images animées et du son associé - Partie Systèmes) [ISO, 2000b]. Les aspects compression, quant à eux, sont définis dans les normes ISO/CEI 13818-2 et 3 (Codage générique des images animées et du son associé - Parties vidéo, audio) [ISO, 2000a]. Ce format vidéo est utilisé pour les DVD et SVCD avec différentes résolutions d'image. Il est également utilisé dans la diffusion de télévision numérique par satellite, câble, réseau de télécommunications ou hertzien (TNT). MPEG-2 permet de transformer un signal vidéo numérisé en un train binaire destiné à être stocké ou transmis sur un réseau. MPEG 2 s'appuie principalement sur une exploitation des redondances spatiales, temporelles, subjectives et statistiques existant dans une séquence vidéo. Le train binaire obtenu est décrit selon la norme MPEG afin que l'on puisse restituer le signal par n'importe quel décodeur respectant cette même norme. Le schéma d'encodage (Figure 2.7) de MPEG-2 est conforme aux principes généraux de compression vidéo décrits dans le paragraphe précédent : transformation des données, quantification, codage entropique, compensation et estimation du mouvement.

H.264 : Schéma général Les techniques de compression de la famille MPEG et H.264 sont détaillés dans les ouvrages [Watkinson, 2006] et [Richardson, 2008]. Le standard H.264 introduit la notion de « tranche » (slice) pour décrire une région de l'image codée contenant un ensemble de macroblocs. L'encodeur H.264 inclut deux chemins (Figure 2.9 : le chemin direct (de gauche à droite) et le chemin de reconstruction (de droite à gauche).

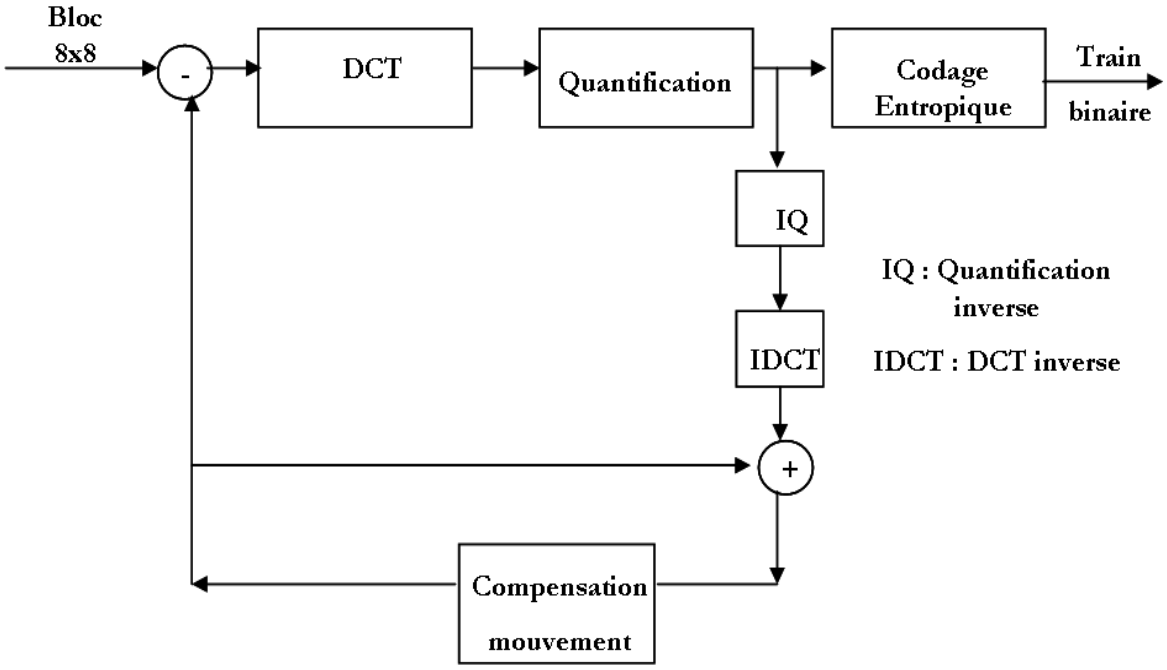


FIGURE 2.7. *Encodeur MPEG-2*

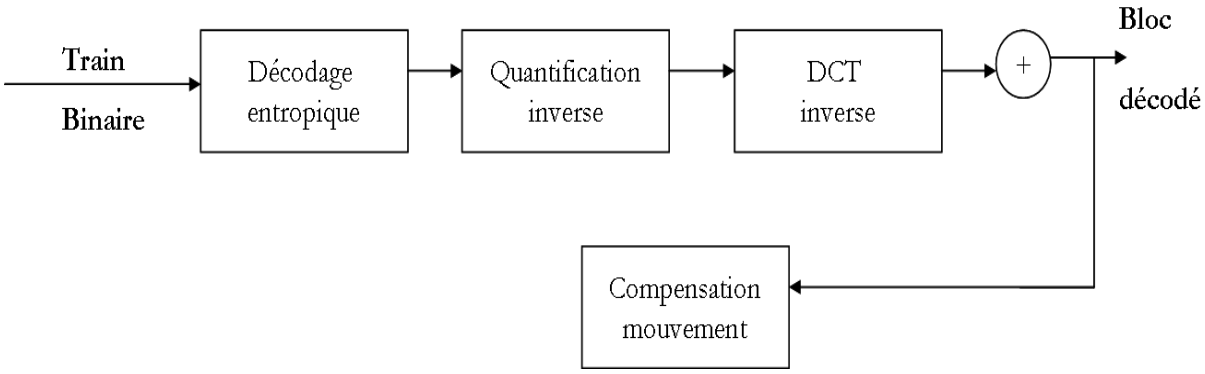


FIGURE 2.8. *Décodeur MPEG-2*

Dans le chemin direct, F_n est un macrobloc encodé en mode intra ou inter et pour chaque bloc dans le macrobloc, une prédiction P est calculée en se basant sur les échantillons d'image reconstruits. Dans le mode intra, P est formé d'échantillons de la tranche courant qui ont été encodés, décodés et reconstruits au préalable (dans la figure 2.9, il s'agit de uF'_n. Dans le mode inter, P est formé de la prédiction de mouvement depuis une des images de référence (F'_{n-1} dans la figure 2.9). Ensuite, P est soustraite du bloc

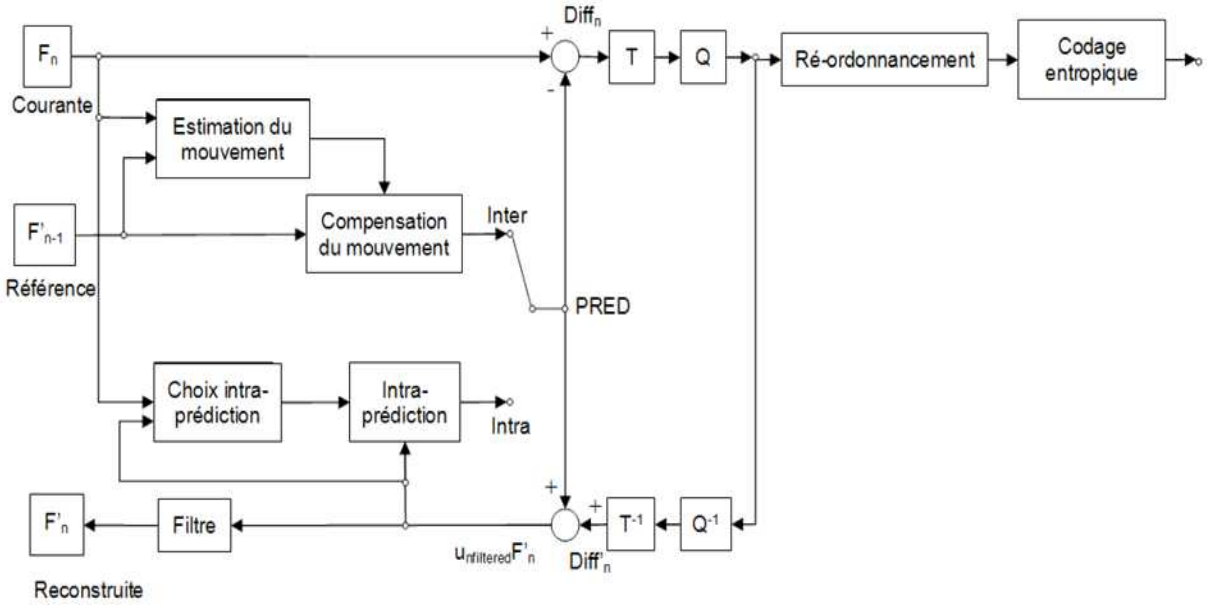


FIGURE 2.9. *Encodeur H.264*

courant pour former un bloc D_n (difference) qui est transformé et quantifié, fournissant ainsi X, ensemble de coefficients quantifiés qui sont ré-ordonnés et codés (codage entropique). Les coefficients codés sont requis pour décoder chaque bloc dans le macrobloc (mode de prédiction, paramètres de quantification, information du vecteur mouvement). La trame binaire ainsi obtenue est ensuite passée à la NAL (Network Abstraction Layer) pour être transmise ou stockée.

Dans le chemin de reconstruction, l'encodeur décode (reconstruit) chaque bloc dans le macrobloc pour fournir une référence pour les prochaines prédictions. Une quantification inverse suivie d'une transformée inverse sont effectuées sur X pour produire un bloc de différence D'_n. Ce bloc est additionné à P pour créer le bloc reconstruit uF'_{n-1} (version décodée du bloc original). Enfin, un filtre est appliqué pour réduire les effets de bloc (distorsions) et une image reconstruite de référence est créee à partir de la série de blocs F'_n.

Le décodeur reçoit la trame binaire de la NAL et la décode (décodage entropique) pour produire un ensemble de coefficients quantifiés X. Une quantification inverse suivie d'une transformée inverse sont effectuées sur X pour produire un bloc D'_n (le même que

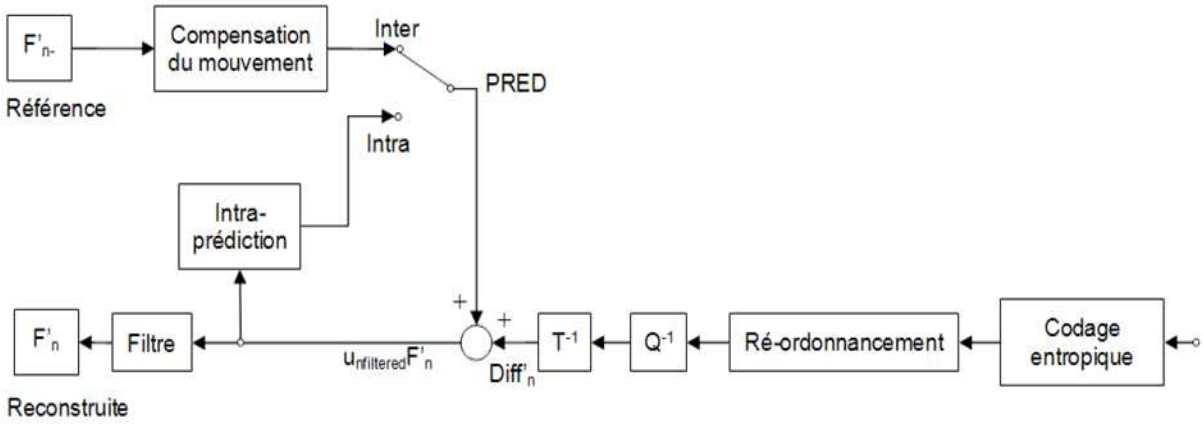


FIGURE 2.10. *Décodeur H.264*

pour l'encodeur). Le décodage de l'en-tête de la trame binaire permet au décodeur de créer des blocs de prédiction P, qui additionnés à D'_n produisent uF'_n. Après filtrage, on obtient les blocs décodés F'_n.

H.264 : Main profile Le profil d'un codec vidéo est l'ensemble des possibilités qu'il peut offrir. H.264 définit trois profils (Baseline, Main et Extended), chacun fournissant un ensemble de fonctionnalités d'encodage et chacun spécifiant les requis pour les encodeurs ou décodeurs afin d'être conformes au profil. Le Baseline Profile permet d'encoder en mode intra et inter (utilisant les tranches I-slice et P-slice) et le codage entropique CAVLC (Context-Adaptive Variable-Length Codes). Le Main Profile, qui est adapté aux applications de broadcast telles que la télévision numérique mais aussi au stockage vidéo.

Nous avons choisi ce profil pour compresser les vidéos chirurgicales. Ce profil convient aux vidéos entrelacées et permet un mode inter utilisant les tranche B-slice ainsi qu'un codage entropique CABAC (Context-Based Arithmetic Coding).

2.2 Artéfacts

Dans le paragraphe précédent, nous avons décrit les principes communs à la plupart des méthodes de compression vidéo qui se basent sur une compensation de mouvement

et une transformée DCT des blocs suivie d'une quantification des coefficients et de leur codage sans pertes. Dans de tels schémas, la quantification des coefficients transformés apporte l'essentiel des dégradations de la séquence compressée. Cependant, d'autres facteurs tels que la prédiction de mouvement ou la taille du buffer de décodage peuvent affecter la qualité visuelle du flux par leur effet sur le processus d'encodage. Les paragraphes suivants décrivent ces principales sources de dégradations dues à la compression et à la transmission.

2.2.1 Artéfacts liés à la compression

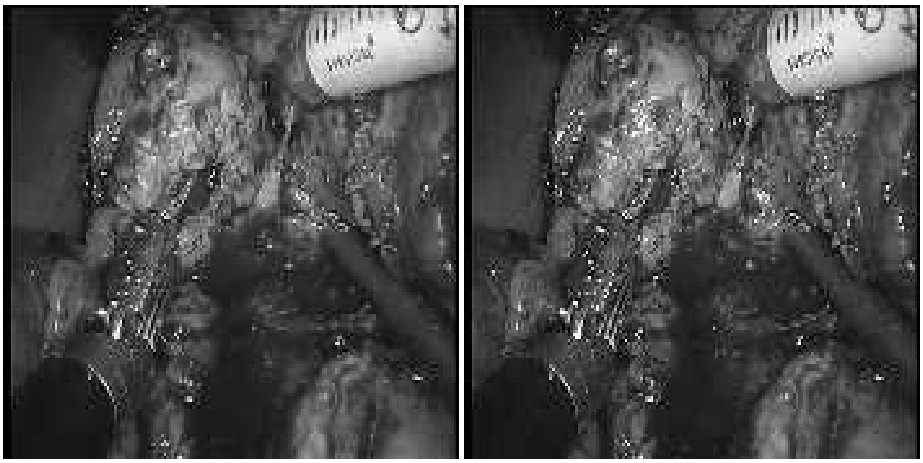
Dans [Yuen et Wu, 1998], les auteurs décrivent l'ensemble des dégradations que l'on peut distinguer dans une séquence compressée. Ces distorsions peuvent être spatiales ou temporelles.Nous ne décrirons ici que les effets ayant le plus d'impact visuel, à savoir **l'effet de bloc**, **l'effet de flou** et **le ringing** dans le domaine spatial, **l'effet de moustique**, **le mouvement saccadé** et **le scintillement** dans le domaine temporel.

2.2.1.1 Distorsions spatiales

Effet de bloc L'effet de bloc est un effet bien connu lié à la décomposition en bloc de la séquence avant quantification. La quantification de chaque bloc de coefficients de la DCT (généralement de taille 8x8) indépendamment génère des discontinuités aux frontières des blocs adjacents. Cet effet de bloc est une distorsion visuelle très importante pour la plupart des séquences compressées (figure 2.11). H.264 utilise un filtre nommé "deblocking filter" pour réduire la visibilité de cet artéfact.

Effet de Flou Le flou est dû à la perte de détails spatiaux et la réduction de l'épaisseur des contours de l'image par la suppression des coefficiens de haute-fréquence (suite à une forte quantification) : figure 2.12.

Ringing Le ringing est associé au phénomène de Gibbs (oscillation de reconstruction d'un signal discontinu avec une somme de signaux continus). Il résulte d'irrégularités dans la reconstruction des hautes fréquences du bloc reconstruit. Après transforma-



(a) (b)

FIGURE 2.11. *Effet de bloc*

FIGURE 2.12. *Effet de flou*

tion inverse, des erreurs apparaissent sous forme d'ondulations, particulièrement visibles le long des contours fortement contrastés.

2.2.1.2 Distorsions temporelles

Effet de moustique L'effet de moustique est une distorsion que l'on peut voir essentiellement dans les surfaces lisses. Il s'agit de fluctuations de la luminance/chrominance autour des contours à haut contraste ou des objets en mouvement. Il est la conséquence

de l'encodage différent de la même zone d'une scène dans les images successives de la séquence.

Effet de mouvement saccadé Cet artéfact apparaît quand la limite de la zone de recherche dans laquelle peut être estimé le vecteur de mouvement est insuffisante. Le mouvement n'étant pas correctement représenté d'une image à l'autre, sa fluidité est cassée.

Scintillement Cet artéfact est dû à la variation dans le temps du facteur de quantification par blocs des zones texturées d'une scène. Il affecte principalement les scènes riches en textures.

2.2.2 Erreurs de transmission

La transmission du flux sur un canal bruité peut être source de dégradations. Les données vidéo numériques compressées sont généralement transmises sur des réseaux à commutation de paquets. Cette technique de commutation est fondée sur le découpage des données afin d'en accélérer le transfert. Chaque paquet est composé d'un en-tête contenant des informations sur le contenu du paquet ainsi que sur sa destination, permettant ainsi au commutateur d'aiguiller le paquet sur le réseau vers son point final. Ce transfert peut se faire selon des protocoles de transport comme ATM ou TCP/IP. Deux types de dégradations peuvent survenir lors du transport des données sur canal bruité :

- les paquets peuvent êtres corrompus ou perdus et par conséquent, ils ne sont pas acheminés dans les temps vers le décodeur,
- le routage des paquets et leur mise en file d'attente dans les routeurs et commutateurs, rend une partie du flux indisponible.

Ces paquets seront perdus définitivement et n'atteindront jamais le décodeur. Les pertes, peuvent affecter les données relatives aux blocs d'une image ou les informations liées au mouvement. A titre d'exemple, un macrobloc MPEG endommagé suite à une perte de paquets réseau, va avoir une influence sur tous les MB précédents et suivants dans la tranche.

2.2.3 Autres distorsions

Outre les artéfacts de compression et les erreurs de transmission, une séquence vidéo peut être altérée par des pré ou post-traitement comme par exemple :

- La conversion analogique/numérique de la séquence,
- le sous-échantillonnage de la chrominance,
- la conversion du nombre d'images en fonction du format d'affichage,
- le désentrelacement c'est-à-dire le processus de création de séquences progressives à partir de séquences entrelacées.

2.3 Conclusion

Dans ce chapitre, nous avons présenté les principes généraux de la compression vidéo. Nous avons également décrit les standards MPEG-2 et MPEG-4 Part10 ou H.264, que nous avons retenus pour la compression des séquences vidéo chirurgicales. D'une part, le choix de la norme MPEG-2 est motivé à la fois par sa maturité et ses bonnes performances débit/distorsion. D'autre part, H.264 offre des performances supérieures à MPEG-2 et des artéfacts différents mais aussi une plus grande complexité qui peut augmenter le temps de latence. Il existe aujourd'hui plusieurs normes de codage et particulièrement celles appartenant à la famille MPEG. La norme de codage vidéo MPEG-2 a été le pilier technique des systèmes de télévision numérique dans le monde. La norme MPEG-2 est structurée en profils et niveaux, en définissant clairement pour chacun d'eux le débit binaire maximum que le décodeur doit pouvoir traiter. Les performances de cette norme sont différentes selon l'algorithme de compensation de mouvement, les valeurs de la matrice de quantification et le dispositif de contrôle du débit. En effet, la norme MPEG-2 définit uniquement la syntaxe du flux binaire et les caractéristiques du décodeur. Quant à la norme MPEG-4, comme pour MPEG-2, l'efficacité du codage est liée à la complexité du matériel de source et à la mise en oeuvre du codeur. Cette norme a été définie pour les applications multimédia à faible débit binaire, puis étendue au domaine de la radiodiffusion. Une évaluation officielle subjective indique que le codage MPEG-4 partie 2, offre un gain d'efficacité de 15 à 20% par rapport à MPEG-2. En 2003, le système AVC

(Advanced Video Coding) est intégré en tant que partie 10 de cette norme et repris sous l'appellation H.264. Par ailleurs, nous avons répertorié les distorsions de la vidéo entre celles inhérentes à la compression, dues à la transmission ou au traitement en amont de la séquence. La question qui se pose, dans le contexte de notre étude, est la suivante : quels taux de compression peut-on raisonnablement appliquer dans le domaine chiurgical tout en maintenant une qualité de la vidéo acceptable pour un bon déroulement d'une opération à distance en temps-réel ?

Chapitre 3

Mesure de la qualité des vidéos compressées

Ce chapitre est consacré à l'évaluation de la qualité des images et des vidéos. Nous faisons une synthèse des approches normalisées d'évaluation subjective de la qualité. Nous décrivons ensuite les métriques objectives avec référence complète et leurs principes, les métriques avec référence réduite et sans référence. Nous terminons, enfin, par une discussion sur les performances de ces approches et nous motivons nos choix pour l'évaluation de la qualité des vidéos dans le contexte chirurgical.

3.1 Introduction

Les recherches dans le domaine de la qualité des images et des vidéos se concentrent sur la conception des métriques de qualité capables d'approcher celles obtenues en moyenne par un panel d'observateurs lors de tests d'évaluation de la qualité perceptuelle. En effet, l'évaluation subjective reste le moyen de référence pour évaluer la qualité des vidéos car l'observateur humain est l'utilisateur final dans la plupart des applications. Le score moyen d'opinion ou Mean Opinion Score (MOS), qui est l'unité de la perception subjective de la qualité obtenue à partir d'un panel d'observateurs, a toujours été considéré comme le score de perception de qualité le plus fiable. Malgré le coût élevé de la mise en place de telles mesures, elles sont utiles car elles fournissent des données in-

contournables pour estimer le jugement humain. Elles permettent également de calibrer les performances des méthodes d'évaluation objective de la qualité. Le principe de ces métriques est de modéliser et de simuler le comportement du système visuel humain (SVH) sur une image altérée en renvoyant un score de qualité. En général, ce type d'approche étudie les fonctionnalités de chaque composant du SVH dans le but de simuler ses caractéristiques. Ces approches dites « ascendantes » ont pour objectif de construire un système numérique dont le fonctionnement s'approche de la perception humaine. Une approche « descendante » est possible en faisant des hypothèses sur l'ensemble des fonctionnalités du SVH (modélisation globale).Il faut noter que la définition de ces deux termes est difficile. En effet, il est impossible, pour une approche ascendante, de simuler chacune des caractéristiques du SVH. Il est également difficile pour une approche descendante de modéliser correctement tout le SVH.

3.2 Mesure subjective de la qualité

3.2.1 Introduction

L'objectif de ce paragraphe est de décrire différents aspects à prendre en compte dans la mise en oeuvre et la conduite de tests subjectifs de la qualité. L'élaboration de ce type de tests est complexe et les sources de biais sont nombreuses. Les recommendations existentes préconisent des conditions générales d'observation, qui vont permettre de réduire les effets parasites de l'environnement de test en partuculier, et d'autre part des techniques de traitement et d'analyse des résultats. Plusieurs méthodologies d'essais subjectifs existent. Cependant, le choix parmi tous ces protocoles n'est pas évident.

3.2.2 Tests subjectifs

Les tests subjectifs, appelés aussi « tests psychovisuels » représentent le moyen le plus précis pour évaluer la qualité d'une vidéo. Pour les tests subjectifs, un nombre de participants est appelé à visionner un ensemble de séquences-tests et à donner un jugement sur leur qualité ou sur le degré de gêne occasionnée par les dégradations. La moyenne des jugements collectés pour chaque séquence-test représente le MOS pour cette séquence.

En général, les expérimentations subjectives sont coûteuses (temps, moyens humains, contraintes de mise en œuvre) et très longues à dérouler. La mise en place d'un protocole d'essais subjectifs, la réalisation et l'analyse des données consomment une grande quantité du temps total d'expérimentation. Afin de réaliser un test subjectif, il est indispensable de réunir un panel d'observateurs, des équipements et un espace physique (laboratoire d'essais). Une méthodologie appropriée doit être ensuite choisie. Avant de sélectionner la méthode à utiliser, il est nécessaire de prendre en compte l'application visée et ses objectifs précis comme l'indique la figure 3.1).

Les documents de l'International Telecommunications Union (ITU) donnent des informations sur les conditions d'observation, les critères de sélection des observateurs, le matériel de test, la procédure d'évaluation et les méthodes d'analyse des données collectées. Ces conditions sont communes et partagées par les méthodologies normalisées. Selon l'ITU, il existe deux classes d'approches pour l'évaluation subjective :

- **Mesure de la qualité :** les jugements des observateurs sont donnés sur une échelle de qualité où le jugement porte sur la bonne ou mauvaise qualité perçue de la vidéo affichée. Ces mesures établissent donc la performance des systèmes sous des conditions optimales,
- **Mesure de la dégradation :** les jugements des observateurs sont donnés sur une échelle de dégradation où le jugement porte sur la perception ou non de la dégradation de la vidéo affichée. Ces mesures établissent donc la performance des systèmes sous des conditions non optimales (relatives à la transmission par exemple).

Selon ces approches, les échelles d'évaluation peuvent être discrètes ou continues, catégorielles ou non-catégorielles, adjectivales ou numériques. Le type de la méthode d'évaluation va aussi dépendre de la forme de présentation du stimulus (séquence) :

- **Méthode à simple stimulus :** dans les approches à simple stimulus, on ne présente à l'observateur que l'image ou la séquence de test,
- **Méthode à double stimulus :** une paire d'images ou de séquences est présentée à l'observateur (référence, séquence de test).

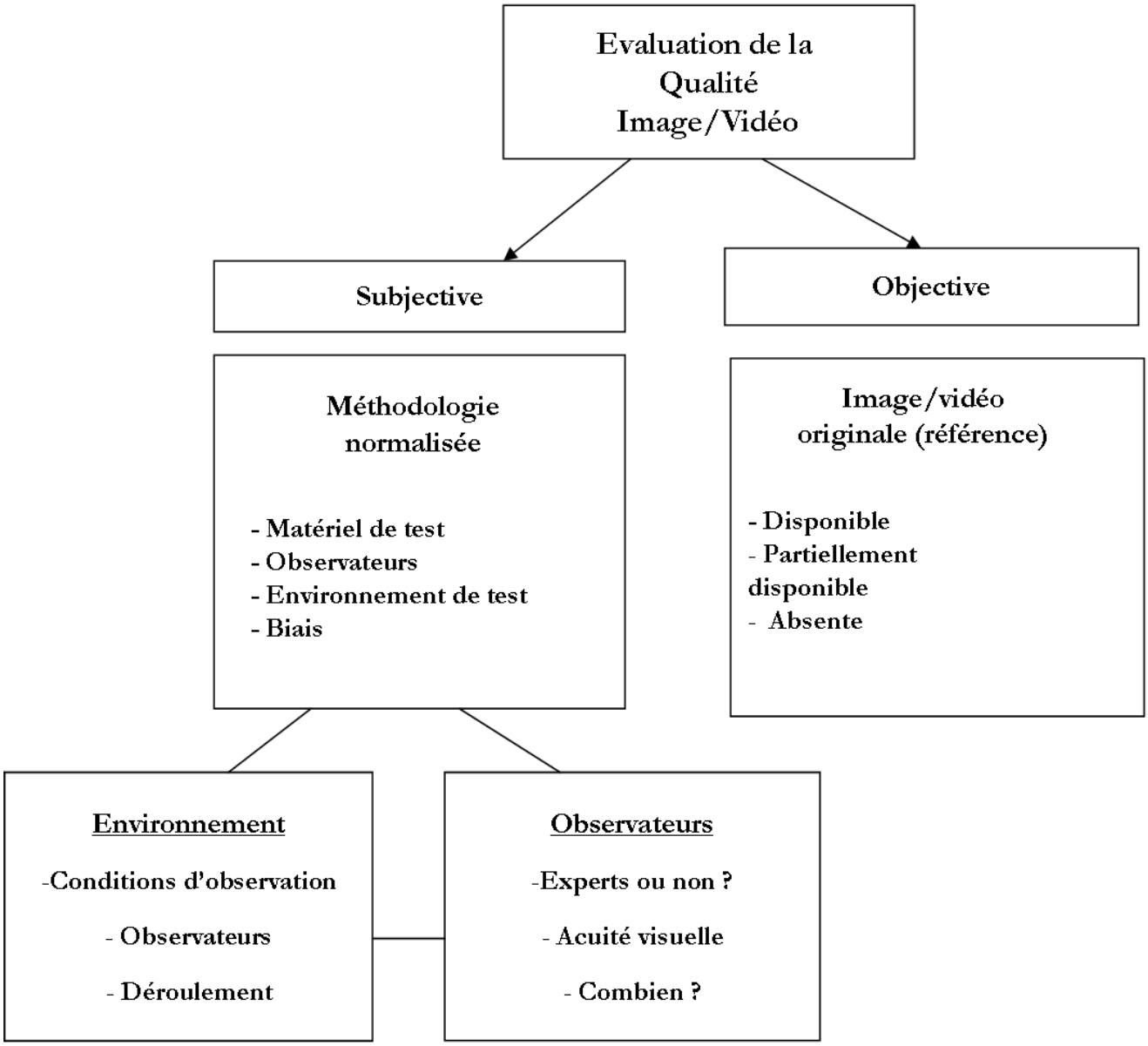


FIGURE 3.1. *Evaluation de la qualité*

3.2.3 Méthodologies normalisées d'évaluation subjective de la qualité

La mesure subjective de la qualité des images est un procédé complexe et contraignant nécessitant des ressources importantes. Cette mesure se base sur un jugement qualitatif d'observateurs humains dont le nombre et les conditions d'observation sont normalisés. Elle pose, cependant, un problème de mise en place car le jugement humain peut être variable, les conditions d'évaluation strictes et le nombre d'observateurs nécessaires pour approcher le jugement humain étant discutable.

Il existe des méthodologies normalisées avec différentes approches les distinguant (chaque méthodologie mesure une grandeur différente). Plusieurs méthodes sont formalisées dans ITU-T-R Rec. BT-500 [ITU-R, 2000] parmi lesquelles : Double Stimulus Continuous Quality Scale (DSCQS), Absolute Category Rating (ACR), Double Stimulus Impairment Scale (DSIS) et Single Stimulus Continuous Quality Evaluation (SSCQE) décrites dans ce qui suit.

3.2.3.1 DSCQS (Double Stimulus Continuous Quality Scale)

Introduction La méthodologie DSCQS est particulièrement utile lorsqu'il n'est pas possible de créer des conditions expérimentales et des stimulus d'essai représentant toute la gamme de qualité. Elle permet de présenter à l'observateur plusieurs paires de séquences comprenant une séquence de référence et une séquence dégradée. Chaque paire est présentée deux fois (cf. figure 3.2). L'ordre de la séquence de référence dans une paire varie de manière pseudo-aléatoire.

A la fin de chaque présentation, les observateurs expriment leur jugement par une note sur une paire d'échelles verticales continues. A ces échelles graduées de façon continue sont associés des qualificatifs du protocole DSCQS : excellent, bon, assez bon, médiocre et mauvais, comme le montre la figure 3.3. Il est cependant erroné d'associer aux notes un seul qualificatif car elles ne sont pas absolues mais représentent la différence de note entre l'image de référence et l'image dégradée.

Exploitation des résultats L'ITU préconise que les données recueillies au cours des essais subjectifs soient traitées selon les mêmes techniques statistiques recommandées.

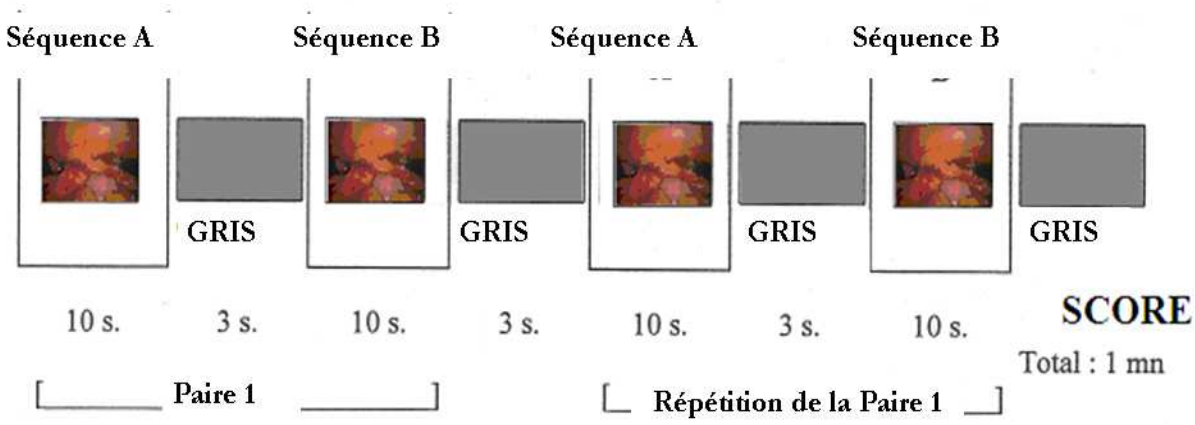


FIGURE 3.2. *Structure d'une séance d'évaluation DSCQS*

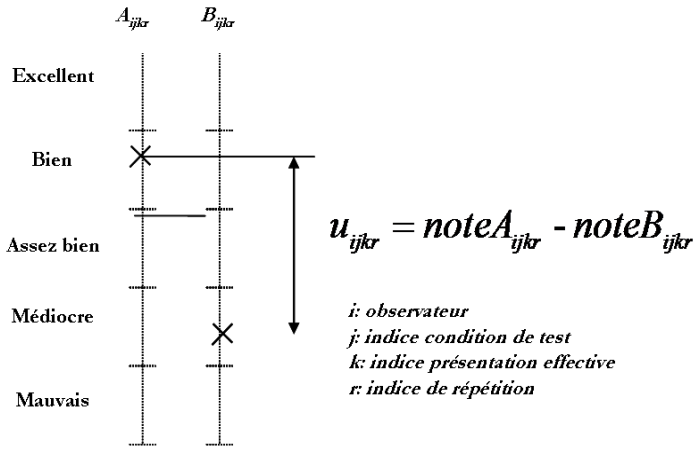


FIGURE 3.3. *Echelle d'évaluation DSCQS*

En effet, l'analyse des résultats issus des tests subjectifs dépend de la méthode utilisée. Ainsi, pour la DSCQS, la cohérence des résultats sera vérifiée en étudiant les notes données par le même observateur à la même séquence pendant la même séance. Si les notes diffèrent de 2 points ou plus (pour une échelle continue allant de 1 à 5), ces notes seront rejetées, l'observateur étant jugé « non fiable ». Après chaque séance, on calcule les valeurs moyennes et les écarts-type associés à chaque niveau de dégradation. La valeur moyenne est donnée par :

$$\overline{\mu}_{jkr} = \frac{1}{N_{obs}} \cdot \sum_{i=1}^{N_{obs}} \mu_{ijkr} \tag{3.1}$$

où N_{obs} représente le nombre d'observateurs et μ_{ijkr} la note de l'observateur i pour la dégradation j de la séquence k et la répétion r.

Afin d'évaluer au mieux la fiabilité des résultats, on associe à chaque moyenne un intervalle de confiance. En général, il est convenu d'utiliser l'intervalle de confiance à 95% donné par :

$$\left[\overline{\mu}_{jkr} - \delta_{jkr}, \overline{\mu}_{jkr} + \delta_{jkr}\right] \tag{3.2}$$

où $\delta_{jkr} = 1,96 \cdot \frac{\sigma_{jkr}}{\sqrt{N_{obs}}}$

L'écart-type de chaque présentation est donné par :

$$\sigma_{jkr} = \sqrt{\sum_{i=1}^{N_{obs}} \frac{\left(\overline{\mu}_{jkr} - \mu_{ijkr}\right)^2}{N_{obs} - 1}} \tag{3.3}$$

Ces valeurs moyennes reposent sur une loi de distribution dont les deux variables aléatoires sont les scènes et les observateurs. Afin de vérifier si cette distribution suit une loi normale, on calcule son cœfficient d'aplatissement (kurtosis), défini comme étant le rapport entre le moment d'ordre 4 et le carré du moment d'ordre 2, soit :

$$\beta_{2jkr} = \frac{m_4}{{m_2}^2} \tag{3.4}$$

avec $m_x = \frac{\sum_{i=1}^{N_{obs}} \left(\mu_{ijkr} - \overline{\mu}_{ijkr}\right)^x}{N_{obs}}$

La norme indique que si β_2 est compris entre 2 et 4, on peut considérer la distribution comme normale. Les résultats de chaque distribution sont alors à comparer à la valeur moyenne dans un intervalle selon l'algorithme donné ci-dessous. Chaque fois que les résultats d'un observateur se situent en dehors de cet intervalle, il faut les enregistrer dans un compteur associé à chaque observateur. Il faut donc deux intervalles pour les valeurs supérieures et inférieures. Cette procédure est récapitulée dans la norme [ITU-R, 2000] et s'exprime comme suit :

Pour chaque observateur i

Pour j,k,r =1 : J,K,R

Si $2 \leq \beta_{2jkr} \leq 4$ alors :

Si $\mu_{ijkr} \geq \overline{\mu}_{jkr}$ + 2.σ_{jkr} alors $P_i = P_i + 1$

Si $\mu_{ijkr} \leq \overline{\mu}_{jkr}$ - 2.σ_{jkr} alors $Q_i = Q_i + 1$

Sinon :

Si $\mu_{ijkr} \geq \overline{\mu}_{jkr}$ + $\sqrt{(20)}.\sigma_{jkr}$ alors $P_i = P_i + 1$

Si $\mu_{ijkr} \leq \overline{\mu}_{jkr}$ - $\sqrt{(20)}.\sigma_{jkr}$ alors $Q_i = Q_i + 1$

Fin Pour

Si $\frac{P_i+Q_i}{J.K.R} > 0,05$ et $\left|\frac{P_i-Q_i}{P_i+Q_i}\right| < 0,03$ Alors

Rejeter l'observateur i

Fin Pour

avec :

J : nombre de conditions de test y compris la référence

K : nombre d'images ou séquences de test effectives

R : nombre de répétitions

L : nombre de présentations de test

3.2.3.2 DSIS (Double Stimulus Impairement Scale)

Cette méthodologie permet de mesurer la gêne ressentie par l'observateur. La référence est présentée aux observateurs avant la séquence test et la paire n'est présentée qu'une seule fois. L'échelle de notation est une échelle de dégradations associant un attribut : « imperceptible », « perceptibles mais non gênantes », « légèrement gênantes », « gênantes » et « très gênantes » à une catégorie correspondant à l'appréciation de la gêne ressntie par l'observateur. Cette méthodologie est adaptée dans le cas où des artéfacts sont visibles sur la vidéo.

3.2.3.3 ACR (Absolute Category Rating)

L'ACR est une méthode à simple stimulus où les observateurs visualisent la séquence-test sans la séquence de référence. Ils attribuent une note globale de qualité en utilisant une échelle d'évaluation discrète à 5 niveaux allant de « Mauvais » à « Excellent ». Le déroulement de cette méthode est plus rapide que DSIS ou DSCQS en raison de l'absence de séquence de référence à chaque présentation.

3.2.3.4 SSCQE : Single Stimulus Continuous Quality Evaluation

SSCQE est une méthodologie d'évaluation de la qualité subjective à simple stimulus et à échelle de notation continue [Alpert et Evain, 1997]. A l'origine, son objectif était de permettre des évaluations subjectives rapides de la qualité des services numériques (notamment la télévision numérique) dans des conditions proches de conditions domestiques mais également de surmonter la plupart des difficultés rencontrées lors du recours aux méthodes à double stimulus. Elle permet d'obtenir, sur une échelle d'évaluation continue (2 notes par seconde), une estimation de la qualité de la vidéo. Dans cette méthode, les observateurs notent la qualité de la vidéo présentée au moyen d'un dispositif coulissant qu'ils déplacent dans un sens, ou dans l'autre, sur une échelle continue, en fonction de leur perception momentanée de la qualité de la vidéo. La durée des vidéos présentée est plus importante qu'avec les autres méthodologies. Le recours à des séquences d'essai plus longues a suscité de nouvelles difficultés, par exemple pour déterminer la longueur des séquences et la forme que devrait prendre la procédure d'acquisition en fonction du comportement de l'observateur. Plusieurs études ont montré les effets de mémoire récente et de tolérance de l'observateur en insérant des artefacts en différents points de séquences de longueur variable et en recueillant une évaluation unique de la qualité à la fin de chaque présentation. La longueur d'une séquence varie entre 30 et 60 minutes. L'intérêt de cette méthode, qui se base sur le phénomène de la mémoire humaine, a été démontré notamment dans le cadre de la télévision numérique. Les résultats obtenus sont plus précis car ils illustrent la réaction immédiate de l'observateur face à un défaut qui peut appraître sur l'image. Cependant, le coût de cette méthode et les ressources qu'elle nécessite ne nous permettent pas de la retenir pour les tests subjectifs de la qualité des vidéos chirurgicales.

3.2.4 Eléments communs aux métriques subjectives de la qualité

La mise en oeuvre d'un test subjectif de la qualité des vidéos doit respecter les préconisations de l'ITU pour assurer la fiabilité et la reproductibilité du test. Certaines conditions doivent être respectées pour le matériel et l'environnement de test ainsi que pour les participants.

3.2.4.1 Sélection du matériel de test : Séquences

Le choix des séquences dépend de deux principaux paramètres : le critère de dégradation (dans notre cas, nous nous intéressons à la plage des débits de compression) et le contenu des images. En effet, il existe des séquences où la dégradation n'est que peu ou pas visible même à bas débits (peu de mouvement notamment). Chaque variable de conception du test aura potentiellement un impact sur la qualité de l'image au moment où elle est visualisée. Il est très connu, que décroître le débit de compression va introduire des artéfacts majeurs visibles sur l'image. Il est donc important de multiplier le type de vidéos et de balayer un large intervalle de débits de compression.

3.2.4.2 Sélection des participants

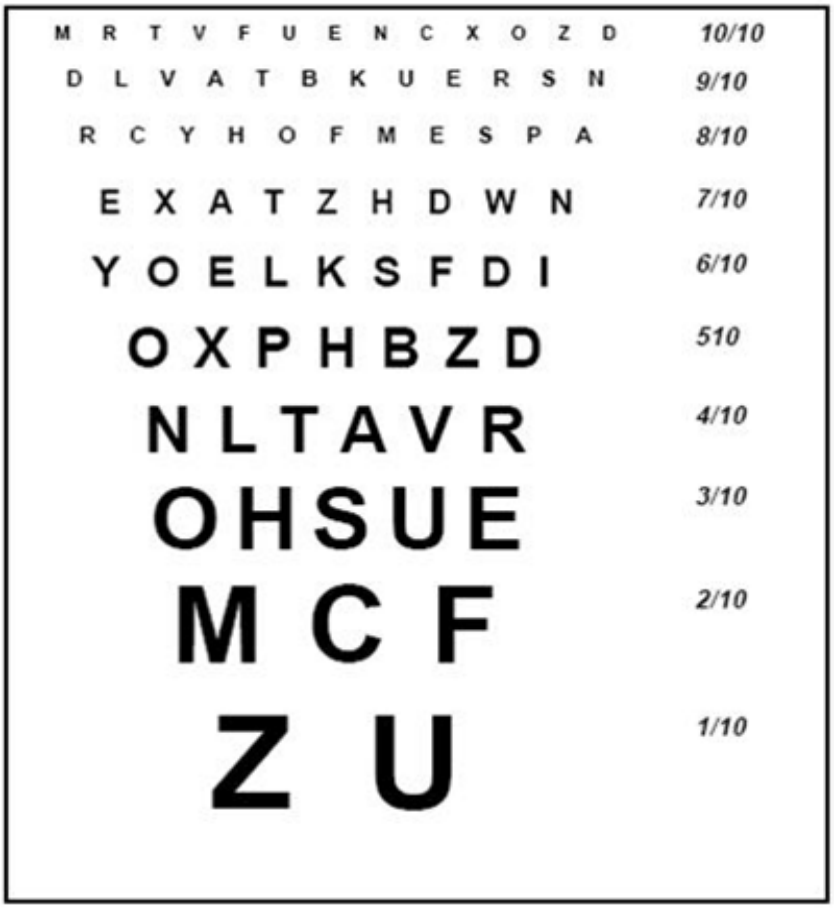


FIGURE 3.4. *Test de l'acuité visuelle (Monoyer)*

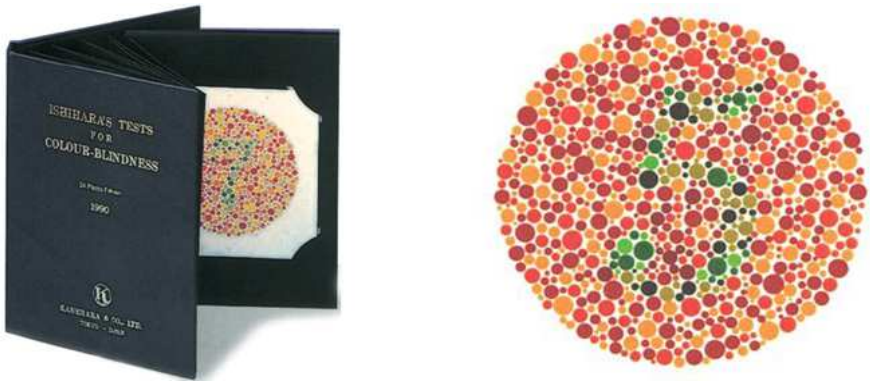


FIGURE 3.5. *Test des couleurs (Ishiara)*

Les particpants peuvent être experts ou non experts de l'image. Dans notre application, nous faisons appel à des experts de l'image médicale (experts chirugiens). Tous les participants doivent être examinés pour leur acuité visuelle à travers le test de Monoyer (voir Figure 3.4 et leurs défauts de perception des couleurs (Test Ishiara, figure 3.5). Pour plus de fiabilité des résultats, un panel entre 15 et 24 participants va donner des résultats exploitables statistiquement. Un observateur habituel peut voir des artéfacts car il n'a pas d'a priori sur la visualisation des séquences vidéo. Le panel doit également être représentatif en âge, genre et expérience. Pour l'évaluation de la qualité de la télévision numérique, la norme définit un nombre minimum de 15 observateurs s'ils sont non experts, ce nombre pouvant être plus faible (minimum 4) lorsque les observateurs sont expérimentés.

3.2.4.3 Laboratoire d'essais/Environnement de test

La durée totale d'observation des séquences ne doit pas dépasser 90 minutes. En fonction de la méthodologie sélectionnée, le nombre de séquences est variable. Pour la méthodologie DSCQS, chaque présentation a une durée de 1 minute. Les sessions doivent, de plus, être divisées en sous-sessions de 30 minutes au maximum. Les trois élémens majeurs qui doivent être considérés pour l'environnement de test sont : la luminosité, le bruit ambiant et la qualité/calibrage du dispositif d'affichage. La normalisation des conditions de test facilite l'appréciation des résultats et minimise l'influence des biais. L'ITU propose plusieurs recommendations dont par exemple la recommendation [ITU-R,

2000] pour la télévision numérique. Cette recommendation contient un certain nombre de règles pour normaliser l'environnement de test. Ces règles initialement définies pour des tests de qualité des images de télévision et le Multimédia [P.910, 1999], sont transposées ici pour l'évaluation subjective de qualité d'images et de vidéos et notamment les vidéos chirugicales. Trois éléments définissent la structure d'un test : l'espace de visualisation, les observateurs et le déroulement d'une séance. Par ailleurs, le positionnement de l'observateur, dans notre cas, doit tenir compte des éléments suivants :

- en accord avec la position du chirurgien, les observateurs sont assis en face de l'écran,
- le rapport taille de l'écran/ distance à l'écran est calculé selon les préconisations de la norme,
- l'angle de vision est également paramètré selon la méthodologie choisie.

Conditions générales d'observation Les éléments les plus importants de l'espace de visualisation à maîtriser sont : la distance d'observation, la luminosité ambiante et les caractéristiques de l'écran. La **distance de visualisation** a une influence directe sur la perception ; de cette distance dépend la répartition des fréquences spatiales de la vidéo projetées sur la rétine. Le contrôle de la luminosité ambiante est important car il n'y a qu'une faible partie du champ visuel qui est excité par la vidéo de test, le reste l'est par l'environnement. De plus, il est souhaitable d'adapter la luminosité ambiante afin de limiter l'éblouissement et la fatigue visuelle des observateurs. Le tableau 3.1 ainsi que les figures 3.6 et 3.7 montrent la relation qui existe entre la taille du moniteur et la distance de visualisation.

Résolution et contraste du moniteur La résolution du moniteur doit être adaptée à la plage de luminance lumineuse recommandée par l'ITU (exprimée en candela par mètre carré 200 cd/m^2). Cette dernière exprime l'intensité lumineuse d'une source lumineuse étendue dans une direction donnée, divisée par l'aire apparente de cette source dans cette même direction. D'autre part, le contraste du moniteur est très fortement influencé par la luminosité ambiante de l'environnement de test.

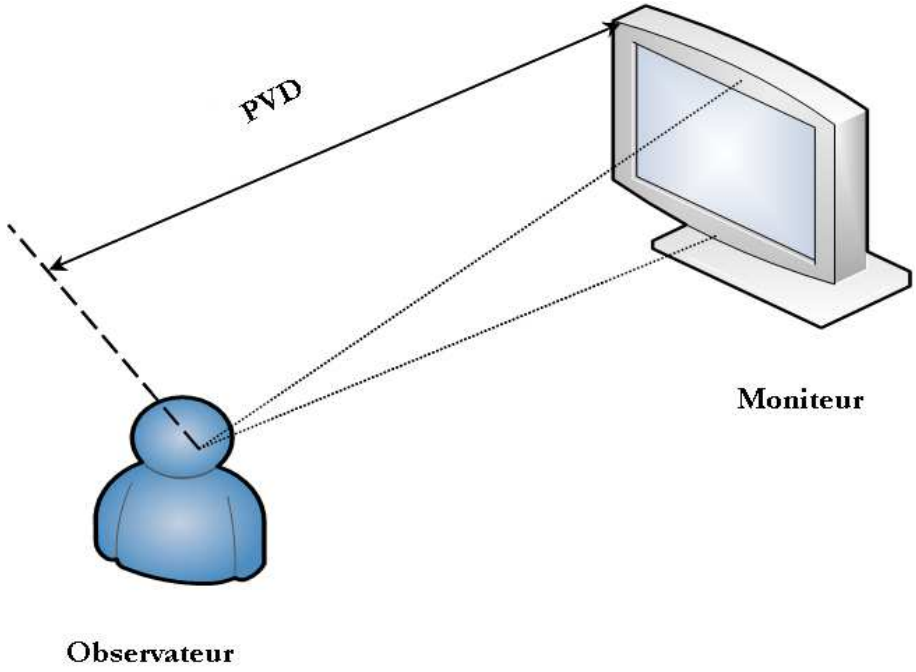


FIGURE 3.6. *PVD : Distance de visualisation*

Diagonale de l'écran		Hauteur de l'écran	PVD
ratio 4/3	*ratio 16/9*	*(m)*	*(H)*
12	15	0,18	9
15	18	0,23	8
20	24	0,3	7
29	36	0,45	6
60	73	0,91	5
>100	>120	> 1,53	3 ou 4

TABLE 3.1. *Relation entre taille de l'écran et la distance de visualisation*

3.2.4.4 Déroulement du test

Les tests basés sur la méthodologie DSCQS se composent de L présentations (L paires de séquences), chaque présentation représente un ensemble J de conditions de test. De

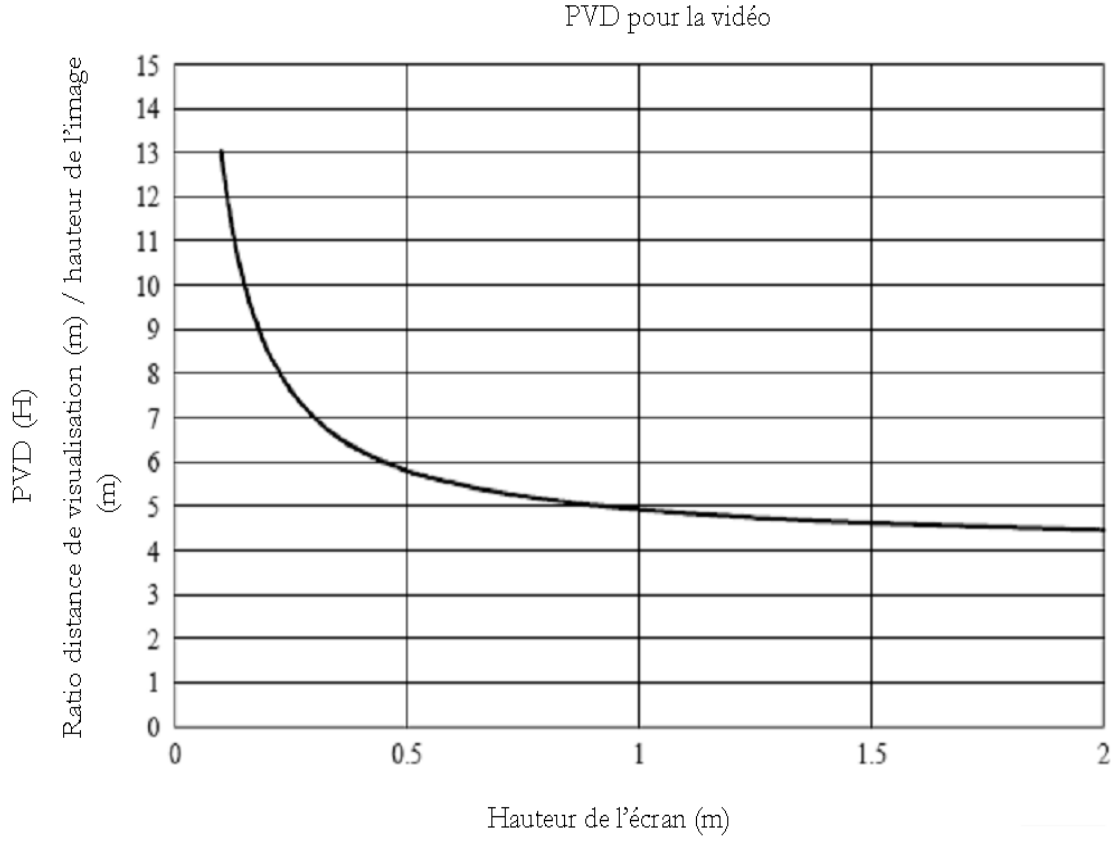


FIGURE 3.7. *PVD en fonction de la taille du moniteur*

plus, chaque paire de séquences peut être répétée R fois afin de pouvoir déceler les incohérences et de telle sorte qu'une séquence n'influe pas sur le jugement de l'observateur pour la séquence suivante. Il faut noter que $K < L$ présentations sont réellement effectives.

3.2.4.5 Biais

La mise en oeuvre et le déroulement d'un test subjectif de la qualité nécessitent une prise en compte des facteurs pouvant influencer les résultats. Ces facteurs sont appelés : les **biais**. Avant le début de chaque test, il est possible de limiter l'impact de ces biais

en les identifiant mais aussi en respectant le protocole préconisé par [ITU-R, 2000]. Cependant, étant donné le nombre important de sources de biais, une correction des notes des observateurs (Z-score) avant d'effectuer les calculs statistiques s'impose dans certains cas (voir chapitre 4). Les sources de biais sont nombreuses. Elles peuvent dépendre du **contexte** du test subjectif, des **facteurs cognitifs** et **psychologiques** de l'observateur. Dans sa thèse, [Ninassi, 2009] classe ces facteurs en trois types d'effets : l'effet contextuel, les styles cognitifs et les facteurs psychologiques.

L'effet contextuel La réponse d'un observateur à un stimulus donné peut dépendre des stimulus précédents. Un effet « dynamique » est observé lorsque les observateurs n'utilisent qu'une portion de l'échelle de notation. Pour compenser cet effet, conditions de dégradation extrêmes appelées *conditions d'ancrage* [Corriveau *et al.*, 1999] sont présentées aux observateurs, en début de séance de test. Ces conditions peuvent être rencontrées pendant le test. Un autres effet contextuel est constaté lorsque la note de la séquence courante est influencée par la dégradation de la séquence qui la précède et plus particulièrement si elle est fortement dégradée. C'est pour cette raison que l'ordre des séquences présentées aux observateurs est choisi de façon aléatoire. Les utilisateurs ont parfois tendance à n'utiliser qu'une portion de l'échelle de notation ou à ne pas juger la séquence globalement, plus particulièrement si les dégradations sont multifactorielles.

Les styles cognitifs Indépendamment de la physiologie du système visuel humain, les observateurs ne perçoivent pas de la même manière les stimulus. Par exemple, un observateur expert aura tendance à localiser les dégradations pour juger une image ou une vidéo tandis que l'observateur non expert jugera globalement la séquence présentée. L'aspect culturel peut avoir une influence sur le style cognitif. Afin de limiter les biais cognitifs, il est impératif d'expliquer aux observateurs en début de la séance le contexte de l'application et l'importance de l'expérience.

Les facteurs psychologiques La disposition psychologique de chaque observateur influence de manière importante les résultats des tests subjectifs. Il existe un phénomène d'initiation où la question qui se pose est de savoir à partir de quelle durée de la séance

d'observation, un observateur peut devenir expert. La motivation des observateurs peut être un facteur d'influence s'ils ne sont pas suffisamment impliqués dans le test : une explication des enjeux du test en début de séance les incitent à rester motivés. Enfin, le degré de concentration, souvent lié à une fatigue visuelle, peut influencer les notes. Afin de remédier à ce phénomène, on limite la durée des séances de test.

3.2.5 Justification du choix de la méthodologie de test

Pour l'évaluation subjective de la qualité des images et des vidéos, l'ITU recommande les méthodes générales d'essai, les échelles et les conditions d'observation. Le choix entre les différentes méthodes de test dépend de l'application considérée. Ainsi, lorsqu'on dispose d'une séquence de référence, la méthode à double stimulus utilisant une échelle de qualité continue DSCQS est la méthode la plus adaptée, en particulier pour mesurer la qualité d'un système par rapport à une référence ou encore comparer des systèmes entre eux. Les recommandations de l'ITU pour l'évaluation subjective de la qualité des vidéos incluent les spécifications sur la manière de dérouler n'importe quel test subjectif. Certaines de ces méthodes sont à double-stimulus où les observateurs évaluent le changement de qualité entre deux vidéos (référence et dégradée). D'autres sont à simple stimulus, où les observateurs ne jugent que la qualité de la séquence dégradée. Chaque méthodologie a ses avantages. Par exemple, la méthodologie DSCQS est moins sensible au contexte c'est-à-dire que les mesures subjectives sont moins influencées par le niveau de dégradation ou son ordre de passage dans la session de test. En effet, afin d'évaluer l'effet du contexte sur les différentes échelles de qualité subjective, [Corriveau *et al.*, 1999] utilise chacune de ces échelles pour évaluer les mêmes séquences vidéo. Les notes obtenues par les différentes échelles se sont révélées fortement corrélées. Le degré auquel chaque méthode était affectée par des effets contextuels a été évalué. Les résultats n'ont révélé aucun effet contextuel pour la méthode DSCQS, des effets contextuels modérés pour la méthode de comparaison proposée par les auteurs et des effets contextuels importants pour la méthode DSIS II. Ils en concluent que la méthode DSCQS est la meilleure méthode à utiliser afin de minimiser les **effets contextuels** pour l'évaluation subjective de la qualité d'images et de vidéos. Dans notre cas, il est plus facile d'écarter certaines

méthodes qui ne répondent pas au besoin de l'application chirurgicale. Ainsi, une méthode à simple stimulus où la note de qualité est globale ne peut pas convenir. Notre choix s'oriente donc vers les méthodes à double-stimulus, où la note de qualité est relative (différence entre la note de la séquence de référence et la séquence dégradée).

3.2.5.1 Conclusion

Malgré leur complexité, les études subjectives de la qualité des vidéos sont la référence pour évaluer la qualité des vidéos. Nous avons vu que l'influence des biais peut être minimisée grâce au respect des recommandations de la norme de l'ITU sur notamment l'environnement et les conditions de test. Le choix parmi tous les protocoles de test n'est pas toujours aisé, mais il est possible d'écarter certaines méthodologies qui ne correspondent pas aux exigences de l'application visée. Dans notre cas, une méthodologie à échelle continue, non coûteuse et à double stimulus est retenue : DSCQS. D'autre part, les études subjectives fournissent des données intéressantes notamment pour évaluer la performance des méthodes d'évaluation objective (ou automatique) de la qualité. Dans le paragraphe suivant, nous décrivons les méthodes objectives de la qualité et leurs principes.

3.3 Mesure objective de la qualité

La littérature propose plusieurs approches d'évaluation objective de la qualité d'images et de vidéos adaptées à des applications très variées comme la compression, la transmission des vidéos ou la qualité des moniteurs. Il existe différents critères sur lesquels on peut se baser pour classer toutes ces métriques : la disponibilité d'un signal vidéo original (supposé non dégradé), la connaissance du processus de distorsion ou le système visuel humain et l'application visée. Le premier critère sur lequel on se base pour classer les métriques objectives de la qualité est la disponibilité ou non d'une image " originale " qui n'a subi aucune distorsion. Cette image peut donc servir de référence dans l'évaluation de l'image altérée. Dans certains cas, la référence n'est que partiellement disponible. Certaines caractéristiques sont extraites de la référence et sont calculées sur la

vidéo dégradée ; leur comparaison permet d'obtenir la note finale de qualité : on parle de métriques avec **référence réduite**. Cependant, dans plusieurs applications, l'accès à la référence n'est pas aisé. Certaines métriques ont été conçues pour évaluer la qualité de manière aveugle ce qui les rend particulièrement complexes. Peu de critères **sans référence** sont disponibles actuellement dans la littérature. La plupart des méthodes d'évaluation objective de la qualité suppose l'existence et la disponibilité de la référence non dégradée. On parle alors de métrique objective de la qualité avec **référence complète**. Enfin, bien que le terme « qualité de l'image ou de la vidéo » soit généralement utilisé pour qualifier ces métriques, il est plus précis dans ce cas d'utiliser le terme « mesure de fidélité ».

3.3.1 Métriques de qualité visuelle avec référence complète

Dans ce type d'approches, on suppose fortement que la perte de qualité est directement liée à un signal d'erreur qui vient se rajouter à un signal initialement " parfait ". On cherche, donc, à quantifier l'erreur entre une séquence originale et une séquence dégradée. La mise en oeuvre la plus simple et presque exclusive de cette catégorie de métriques est l'erreur quadratique moyenne ou Mean Square Error (MSE) et le rapport signal à bruit crête ou Peak Signal to Noise Ratio (PSNR).

3.3.1.1 PSNR

Le PSNR est tout particulièrement utilisé en compression afin d'évaluer les performances des codeurs en mesurant la qualité de reconstruction d'une image par rapport à la version originale. Le PSNR est défini par :

$$PSNR = 10 \cdot \log\left(\frac{d^2}{EQM}\right) \tag{3.5}$$

où d est l'amplitude maximale (crête) du signal, et EQM est l'erreur quadratique moyenne, définie par :

$$EQM = \frac{1}{MN} \cdot \sum_{m=1}^{M} \sum_{n=1}^{N} \|I_o(m,n) - I_d(m,n)\|^2 \tag{3.6}$$

où Io et Id sont respectivement l'image originale et l'image dégradée, M et N les dimensions des images.

Le PSNR est d'usage très courant car il existe plusieurs situations où son utilisation a un sens et il est très adapté aux méthodes d'optimisation. De plus, sa simplicité de calcul et sa rapidité d'exécution sont des arguments qui justifient son utilisation quasi-exclusive par la communauté de traitement du signal. Par ailleurs, il n'existe actuellement que peu de métriques remettant en question son usage. Cependant, ce type de métriques d'évaluation globale de la qualité est très critiqué car elles ne corrèlent pas bien à la perception humaine de la qualité mesurée [Winkler, 1999]. En effet, elles ne modélisent pas le système visuel humain, mais font l'hypothèse que la qualité visuelle décroît quand la distorsion du signal augmente alors que la qualité ne dépend pas uniquement des distorsions mais aussi du contenu de l'image, ou encore de la localisation des distorsions. Par ailleurs, dans le cas des vidéos, les approches, type PSNR, ne prennent pas en compte le contenu de la vidéo car elles sont calculées sur chaque image pixel par pixel, ce qui a un effet parfois désastreux sur la métrique (désynchronisation temporelle, désalignement spatial ou temporel). Enfin, il n'existe pas de relation forte et cohérente entre ces métriques et le score moyen d'opinion obtenu en moyennant les notes données par un panel d'observateurs. Les recherches, depuis quelques décennies, tendent à développer des métriques objectives, essentiellement avec référence complète, qui prennent en compte les caractéristiques du système visuel humain. D'autres approches, telles que les approches structurelles, ont été mises en oeuvre en se basant sur les similarités locales. En effet, ces métriques ont été développées en se basant sur l'existence d'erreurs structurelles qui auront des effets différents dans la perception de la qualité. Ces méthodes sont présentées ci-après.

3.3.1.2 SSIM

On passe, avec cette catégorie d'approches, d'une mesure de l'erreur globale à une mesure de l'erreur structurelle ; la dégradation est considérée comme une perte de l'information structurelle perçue. La première métrique ayant utilisé ce concept est SSIM (Structural Similarity) [Wang *et al.*, 2004a] appliquée d'abord aux images fixes et adaptée ensuite aux vidéos.

SSIM utilise l'index UQI (Universal Image Quality Index) [Wang et Bovik, 2002]. Cet

index définit des mesures de comparaison de luminance $l(x, y)$, de contraste $c(x, y)$ et de structure $s(x, y)$ entre deux signaux de luminance x et y.

$$l\,(x,y) = \frac{2\mu_x \cdot \mu_y}{\mu_x^2 + \mu_y^2}, \quad c\,(x,y) = \frac{2\sigma_x \cdot \sigma_y}{\sigma_x^2 + \sigma_y^2}, \quad s\,(x,y) = \frac{cov_{xy}}{\sigma_x^2 + \sigma_y^2} \tag{3.7}$$

avec μ_x la moyenne de x, μ_y la moyenne de y, σ_x^2 la variance de x, σ_y^2 la variance de y et cov_xy la covariance entre x et y. L'index UQI entre x et y correspond à :

$$UQI(x,y) = l(x,y) \times c(x,y) \times s(x,y) = \frac{4\mu_x\mu_y cov_x y}{\left(\mu_x^2 + \mu_y^2\right)\left(\sigma_x^2 + \sigma_y^2\right)} \tag{3.8}$$

Le passage à SSIM (Wang *et al.* 2002) résulte de la prise en compte des cas où $\mu_x^2 + \mu_y^2$ ou $\sigma_x^2 + \sigma_y^2$ peuvent être proches de zéro. La formule est alors transformée de la manière suivante :

$$SSIM(x,y) = l(x,y) \times c(x,y) \times s(x,y) = \frac{(2\mu_x\mu_y + c_1)\,(cov_x y + c_2)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)} \tag{3.9}$$

avec $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$, L la dynamique des valeurs des pixels, $k_1 = 0,01$ et $k_2 = 0,03$ par défaut.

Cette formule n'est appliquée qu'à la luminance des images. Les grandeurs sont calculées sur des fenêtres de taille 8×8. La fenêtre courante se déplace pixel par pixel sur l'ensemble de l'image. Les auteurs proposent de ne considèrer qu'un sous-ensemble de ces fenêtres pour la simplicité des calculs. Une pondération est ensuite réalisée pour pallier aux effets de blocs qui peuvent être générés par la carte des mesures SSIM. La métrique MSSIM entre deux images X et Y est la moyenne des mesures SSIM sur les N_f fenêtres de luminance :

$$MSSIM\,(X,Y) = \frac{1}{N_f} \cdot \sum_{i=1}^{N_f} SSIM\,(x_i, y_i) \tag{3.10}$$

La métrique SSIM est caractérisée par une simplicité de mise en œuvre et a démontré des performances supérieures au PSNR. En effet, les auteurs, en partant d'une base de données (base LIVE) de 29 images (Sheikh et al., 2002), évaluent les performances de la métrique MSSIM sur des images dégradées avec JPEG et JPEG2000. SSIM a servi de base

de calcul pour plusieurs critères de qualité tels que la méthode ESSIM (Edge-based SSIM) de [Chen *et al.*, 2006a] ou le GSSIM [Chen *et al.*, 2006b]. [Wang *et al.*, 2004b] proposent une extension de SSIM à la vidéo (VSSIM). Dans cette thèse, nous avons retenu la métrique **SSIM** pour objectiver les résultats obtenus lors du test subjectif de la qualité des vidéos chirurgicales suivant la méthodologie DSCQS.

3.3.2 Métriques de qualité visuelle sans référence : approches basées sur la mesure des dégradations

Les dégradations contenues dans une vidéo se définissent comme les causes de la perte de qualité ou ont directement une influence sur la qualité visuelle. Dans notre application, la compression vidéo est le principal processus de dégradation. Ces dégradations sont principalement issues de l'étape de quantification réalisée par l'encodeur. Dans [Yuen et Wu, 1998], les auteurs décrivent une liste d'artéfacts de compression comme le flou ou l'effet de bloc. Nous présentons dans ce paragraphe un catalogue de dégradations dues à la compression. Puis, nous décrivons un exemple de métrique de qualité basée sur cette approche de mesure des dégradations.

3.3.2.1 Catalogues des dégradations

Les algorithmes de compression utilisés dans les différents standards de compression vidéos sont basés sur des principes proches. La plupart utilisent la transformée en blocs DCT suivie d'une étape de quantification des coefficients ainsi que la compensation de mouvement. Outre la quantification qui apporte des dégradations irréversibles à la vidéo, d'autres étapes comme la prédiction de mouvement altèrent la qualité visuelle du flux.

On peut distinguer un certain nombre d'artéfacts dans une séquence vidéo compressée parmi lesquels [Yuen et Wu, 1998] :

– l'effet de bloc (*blockiness*) résulte du traitement par bloc des codeurs de la famille MPEG. Il est dû à la quantification de blocs de coefficients DCT indépendamment (généralement, ces blocs sont de taille 8x8). Il se caractérise par l'apparition de discontinuités horizontales ou verticales sur les frontières des blocs. H.264 utilise un filtre (deblocking filter) pour réduire la visibilité de cet artéfact ;

- Le flou (*blurr*) se manifeste par une perte de détails spatiaux de l'image originale et une atténuation des contours. Il est provoqué par la suppression des coefficients de hautes fréquences ;
- L'effet d'ondulation (*ringing*) résulte d'irrégularités dans la reconstruction des hautes fréquences du bloc reconstruit. Après transformation inverse, des erreurs apparaissent sous forme d'ondulations, particulièrement visible le long des contours fortement contrastés.

Ces effets sont ceux qui ont le plus d'impact visuel et les métriques de qualité basées sur la mesure des dégradations se limitent généralement à ces trois artéfacts ([Farias *et al.*, 2004], [Crété-Roffet, 2007]).

3.3.2.2 Exemple

A titre d'exemple, nous détaillons dans ce paragraphe les travaux de Wang [Wang et Bovik, 2002] sur la mesure des artéfacts des images fixes (à notre connaissance, il n'existe pas de travaux équivalents adaptés à la vidéo). Les auteurs ont mis en œuvre un algorithme de mesure de la qualité dans le domaine spatial à partir d'une base de données d'images compressées avec JPEG. Il faut noter qu'un moyen efficace pour détecter l'effet de bloc et l'effet de flou est de transformer le signal dans le domaine fréquentiel. Pour une image, $x\left(m,n\right)\ pour\ m\ \in [1,M]\ et\ n\ \in [1,N]$, les auteurs calculent un signal différentiel sur chaque ligne horizontale de la luminance :

$$d_h\left(m,n\right) = x\left(m,n+1\right) - x\left(m,n\right) avec\ n \in [1,N-1] \tag{3.11}$$

Un signal 1-D horizontal $f_m\left(n\right) = |d_h\left(m,n\right)|$ est formé pour chaque valeur de m. Le calcul du spectre de puissance de $f_m\left(n\right)\ pour\ m = 1,\cdots,M$ et de leur moyenne, permet d'obtenir une estimation du spectre $P_h\left(l\right)$ où l'effet de bloc peut facilement être identifié par les pics aux fréquences $(1/8, 2/8, 3/8\ et\ 4/8)$ et l'effet de flou est caractérisé par le changement d'énergie des hautes-fréquences aux basses-fréquences.

Cependant, cette technique a le défaut d'utiliser une Transformée de Fourier rapide (FFT), mais qui va être calculée plusieurs fois pour chaque image, ce qui la rend coûteuse en temps de calcul. C'est pour cela que les auteurs proposent de calculer des ca-

ractéristiques de l'image horizontalement et verticalement. L'effet de flou est difficile à estimer sans l'image de référence mais il introduit une réduction de la dynamique du signal. Les auteurs estiment la dynamique du signal image en incluant deux facteurs. Le premier est la différence moyenne absolue entre les images ; le second est le taux de "zero-crossing". Cette métrique donne une bonne corrélation avec les scores MOS (Mean Opinion Score) obtenus avec une expérimentation subjective conduite avec la même base de données d'images. La méthode proposée peut également servir à mettre en oeuvre une métrique sans référence pour les vidéos compressées avec la famille H.26x/MPEG.

3.3.3 Métriques de qualité visuelle avec référence réduite

The Video Quality Experts Group (VQEG) est le groupe d'experts chargé de comparer et d'évaluer les performances des métriques objectives. Il a publié, en 2000 (FRTV Phase I, 2000) et en 2003 (FRTV, Phase II), deux études comparatives sur les critères d'évaluation de la télévision avec référence complète. Ensuite, leurs travaux ont porté sur les métriques avec référence réduite et sans référence. Plusieurs laboratoires ont particpé à ces études. VQEG s'intéresse également aux applications Multimédia et à la télévision haute définition [VQEG, 2008].

Les métriques de qualité avec référence réduite utilisent une description de l'image ou de la vidéo originale pour produire la note de qualité. Dans (Wolf et Pinson, 2005), les auteurs proposent une métrique de qualité avec référence réduite inspirée du modèle VQM (Video Quality Model) [Wolf et Pinson, 2002], standardisé par l'ITU, développé par NTIA (National Telecommunications and Information Administration). VQM est un critère qui a obtenu les meilleurs résultats lors de la campagne du VQEG en 2003 (FRTV phase II) et il existe un logiciel (*VQM Software*) [NTIA, 2011] qui permet d'évaluer la qualité des vidéos avec ce critère. Celui-ci procède en plusieurs étapes. Une étape de calibrage est réalisée pour comparer les séquences à évaluer, en estimant :

- alignement et correction spatiale entre les deux séquences,
- estimation de la région d'intérêt pour l'extraction des caractéristiques et estimation du contraste, de la luminosité,
- alignement et correction temporelle entre les deux séquences.

L'extraction de caractéristiques locales de la séquence de référence et de la séquence à évaluer permet de les comparer. En fonction de ces caractéristiques les distorsions de la séquence dégradées sont mesurées par rapport à la séquence originale. VQM offre un ensemble de modèles dépendant de l'application visée.

3.3.4 Performances

La qualité, telle qu'elle est perçue par des observateurs humains est exprimée en termes de MOS, et constitue le point de référence pour n'importe quelle métrique de qualité visuelle. Il existe un certain nombre de paramètres qui peuvent être utilisés pour caractériser les performances d'une métrique de qualité dans la prédiction des scores subjectifs. Il s'agit de la **précision**, la **monotonie** et la **cohérence de prédiction**.

- Précision : c'est la capacité de la métrique de prédire les scores subjectifs avec une erreur moyenne minimale et peut être déterminée au moyen du coefficient de corrélation linéaire de Pearson LCC défini par :

$$LCC = \frac{\sum_{i=1}^{N} \left(MOS_i - \overline{MOS}\right)\left(Q_i - \overline{Q}\right)}{\sqrt{\left(MOS_i - \overline{MOS}\right)}\sqrt{\left(Q_i - \overline{Q}\right)}} \tag{3.12}$$

  pour un ensemble de N paires (MOS_i, Q_i) où MOS_i est le score moyen d'opinion de la séquence i et Q_i est le score de qualité de la même séquence obtenu par la métrique objective ;

- Monotonie : mesure si la variation (croissance/décroissance) d'une variable est associée à l'autre variable indépendamment de l'ampleur de cette variation. Idéalement, la différence entre les scores de la métrique et les scores subjectifs correspondants doit toujours avoir le même signe. Le degré de monotonie peut être quantifié par le coefficient de corrélation par rangs de Spearman $SROCC$ défini par :

$$SROCC = 1 - 6\sum_{i=1}^{N} \frac{D^2}{N\left(N^2 - 1\right)} \tag{3.13}$$

  où N est l'ensemble des paires (MOS_i, Q_i), avec MOS_i le score moyen d'opinion et Q_i la note de qualité objective pour la séquence i et D la différence de rang sta-

tistique de ces variables. D est une approximation du coefficient de corrélation calculée à partir des données originales. ;

– Cohérence : elle peut être évaluée en mesurant le nombre d'observations aberrantes (outliers). Une observation aberrante est définie par un point de données pour lequel l'erreur de prédiction dépasse un certain seuil (par exemple deux fois l'écart type de la différence de score subjectif en ce point). Le ratio d'observations aberrantes est défini simplement par le rapport entre le nombre d'observations aberrantes et le nombre total de données. On a évidemment plus de cohérence lorsque le ratio d'observations aberrantes est petit.

3.3.5 Conclusion

Dans ce chapitre, nous avons présenté certains aspects de la mesure de la qualité des images et des vidéos. D'abord, les méthodes et techniques d'avaluation subjective de la qualité standardisées par l'ITU. Ensuite, nous avons classé les méthodes objectives selon la disponibilité ou non d'une référence. Parmi ces méthodes, nous avons décrit un ensemble de métriques avec référence complète, avec référence réduite et sans référence. Il est à noter que, malgré la disponibilité d'un catalogue de métriques objectives, il reste encore beaucoup à investiguer dans le domaine de la qualité objective des vidéos et particulièrement pour les métriques avec référence réduite et sans référence mais aussi dans celui de la qualité des vidéos transmises à travers des réseaux type IP (streaming par exemple). Un autre champ de recherche doit intégrer la qualité des vidéos stéréoscopiques et multivues, plus généralement les applications multimédia (HD). Dans le chapitre suivant, nous mettons en exergue la possibilité de compresser des vidéos chirurgicales avec les deux standards MPEG-2 et H.264 et nous déterminons les limites de tolérance à la compression à travers l'étude subjective de la qualité de ces vidéos. Pour cette étude, nous avons retenu la norme DSCQS pour le volet subjectif et les métriques PSNR et SSIM pour le volet objectif.

Chapitre 4

Sensibilité des chirurgiens à la compression vidéo : Résultats expérimentaux

Ce chapitre présente les résultats expérimentaux de deux études subjectives de la qualité des vidéos chirurgicales que nous avons mises en œuvre. Ces études ont deux finalités : évaluer la sensibilité des chirurgiens à la compression MPEG-2 et H.264, d'une part et évaluer les performances de deux métriques objectives de la qualité, d'autre part. Nous montrons qu'il existe un seuil de tolérance à la compression avec pertes de type MPEG-2 autour de 3 Mbits/s pour les vidéos utilisées dans cette étude, ce qui équivaut à un taux de compression d'environ 90 :1 du flux vidéo initialement à 270 Mbits/s ! Ce taux de compression est plus important dans le cas de H.264 et vient confirmer les performances de ce standard énoncés dans la littérature.

4.1 Seuils et seuils différentiels

Lors du déroulement d'un test subjectif, on soumet aux observateurs des séquences-test pour lesquelles ils sont invités à évaluer soit la qualité de perception soit le degré de gêne occasionnée par un système de dégradation. Cette dégradation peut concerner un paramètre particulier de l'image ou de la vidéo (luminance, contraste, etc.) ou un traitement particulier comme la compression dans notre cas. Etant donnés une information

visuelle soumise à un processus de dégradation progressif et une échelle quantifiant la perception de cette dégradation, deux principales questions se posent :

- A partir de quel seuil sur l'échelle, un observateur commence-t-il à percevoir la dégradation ?
- Quelle différence mesure-t-on sur l'échelle lors de la comparaison d'une information dégradée à une référence ?

Ces deux questions sont d'égale importance dans le processus de jugement de la qualité des images et des vidéos mais l'accent sera mis sur l'une ou l'autre en fonction de l'application visée. Dans notre cas, nous cherchons à déterminer **un seuil au delà duquel un chirurgien ne perçoit plus les dégradations** dues à la compression des vidéos chirurgicales. Nous nous intéresserons, donc, en priorité à la deuxième question qui renseigne sur la notion de **seuil différentiel**.

4.1.1 Définitions

Seuil absolu Il s'agit de la valeur pour laquelle une stimulation est détectée et reconnue par un observateur.

Seuil différentiel Il s'agit de la limite à partir de laquelle un observateur n'est plus capable de dire si une stimulation est présente ou non. Par extension, c'est aussi le seuil au-dessus duquel il ne parvient pas à distinguer deux stimulations. On utilise aussi l'expression anglaise *just noticeable difference (JND)*.

4.1.2 Méthodes de détermination des seuils

La psychophysique classique voit la notion de **seuil** comme étant la quantité de stimulus physiques nécessaires pour la détection d'une JND ; les seuils sont une spécification physique du stimulus. Selon [Engeldrum, 2000], le seuil s'exprime en fonction de la propre perception de l'observateur, c'est-à-dire qu'il n'est pas nécessaire de connaître les propriétés physiques de la dégradation pour déterminer un seuil différentiel. Un des concepts sur lequel on se base pour déterminer le seuil JND est la **dispersion discriminante**. En effet, l'observateur est invité à répondre à une question : « Comment la

séquence A est-elle différente de B ? ». Pour de multiples raisons, les réponses des observateurs varient même si le stimulus demeure constant. Afin de déterminer un seuil, il est utile de déterminer un modèle permettant d'ajuster les données et de déduire ses paramètres utiles. Il existe trois familles de méthodes pour trouver le seuil :

- **La méthode des ajustements** : l'observateur fait varier lui-même la stimulation afin de la placer au niveau qu'il juge être la limite de son seuil de détection ;
- **La méthode des limites** : on présente à l'observateur une série de stimulus d'intensité décroissante ou croissante et relève le niveau à partir duquel il ne parvient plus à détecter le stimulus ;
- **La méthode des stimulus constants** : à la différence de la méthode des limites, le niveau de la stimulation varie de façon non prédictible par l'observateur. Ce dernier doit dire à chaque fois si oui ou non un stimulus était présent.

Nous avons retenu la méthode des stimulus constants pour déterminer un seuil de compression des vidéos chirurgicales tolérable par les praticiens. A partir des résultats obtenus par l'application de la méthodes des stimulus constants, il est possible de tracer une courbe décrivant la variation de la réaction des observateurs au stimulus en fonction de la variation de la dégradation. Dans notre étude, cette courbe représente l'évolution du MOS en fonction de la variation du débit après compression (dégradation).

Modélisation des données expérimentales La figure 4.1, représente une courbe de régression typique du nuage de points représentant la variation de la réaction des observateurs (en termes de MOS (Mean Opinion Score) dans notre cas) en fonction de la dégradation (le débit de compression dans notre cas). En effet, on soumet aux observateurs, pour chaque niveau de dégradation, une séquence de référence et une séquence dégradée. Ces présentations se font dans un ordre aléatoire, et les observateurs sont invités à noter les deux séquences (en les comparant l'une par rapport à l'autre). Pour l'analyse des données, on définit un modèle les approchant (estimation par la méthode des moindres carrés). On appelle ce modèle « la courbe psychométrique ». A notre connaissance, il existe dans la littérature, peu de travaux sur la modélisation des données expérimentales (courbe psychométrique). En général, les modèles sont basés sur une fonction

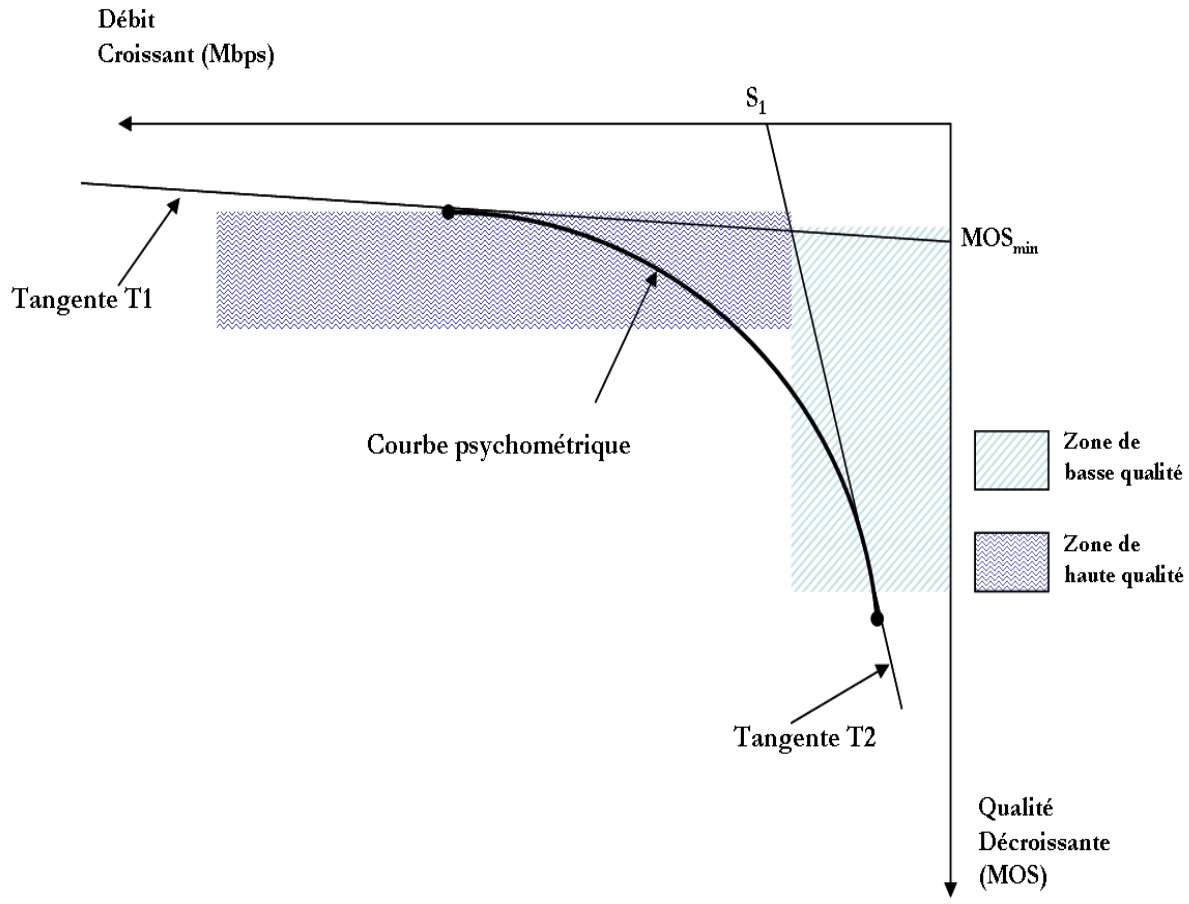


FIGURE 4.1. *Détermination du seuil*

logistique :

$$y = \frac{1}{1 + \exp^{-(\alpha_s + \beta_s . x_{j_s})}} \tag{4.1}$$

où y est la perception du stimulus, α_s et β_s sont des réels et x_{j_s} est le stimulus.

L'avantage d'un tel modèle est la facilité de passage à l'échelle logarithme. La loi de Weber-Fechner décrite dans [Bonnet, 1986] met en évidence la relation entre la grandeur physique d'un stimulus et la perception de ce stimulus (loi affine, exponentielle). Stevens [Stevens, 1957] a proposé en 1957 un modèle qui généralise cette loi : la sensation est liée à la stimulation par une loi puissance. Ainsi, la sensation perçue répond à la formule suivante : $S = k.I^a$ où S est la sensation perçue, k est une constante, I représente l'intensité de stimulation et a est appelé l'exposant de Stevens.

Détermination du seuil Un seuil JND est le changement de stimulus requis pour produire une différence dans la perception des dégradations. Ce seuil est lié à la nature de la courbe psychométrique mais aussi aux réponses souvent hétérogènes des observateurs. Il est à noter que la compression des vidéos chirurgicales, dans notre cas, constitue un défi majeur dans un contexte aussi sensible que le contexte médical, celui de l'impact des pertes sur la qualité des données et leur exploitation et la détermination d'un seuil de compression est d'autant plus complexe dans ce contexte.

Nous proposons ici de fixer ce seuil, à partir de la courbe psychométrique. Deux approches sont possibles. La première approche consiste à définir sur l'échelle de qualité une valeur de seuil ad hoc noté MOS_{min1} : celui-ci correspond à un niveau de qualité jugé excellent. Dans le cas particulier qui nous concerne, on préconise de fixer ce seuil de qualité fixé à 20% de perception de la perte de qualité par rapport à une référence. Cette méthode est utilisée dans l'étude 1 décrite ci-après (voir paragraphe 4.2). Une deuxième approche consiste à définir deux tangentes à la courbe psychométrique (voir figure 4.1), au plus bas et au plus haut niveau de dégradation (ici le débit après compression). Le plan formé par l'intersection de ces deux droites, est ainsi délimité en deux zones : une zone de basse qualité **(1)** et une zone de haute qualité **(2)**. Si on se place dans le contexte de l'évaluation de la qualité des vidéos compressées, la tangente horizontale (T1) permet de simuler le comportement asymptotique du modèle lorsque la débit tend vers l'infini (donc lorsque le taux de compression diminue). L'intersection de cette tangente avec l'axe de la qualité définit une valeur seuil de qualité notée MOS_{min2}. La tangente verticale (T2) définit une valeur S_1 approchant le comportement asympototique du modèle lorsque le débit tend vers 0. Le seuil de débit recherché est déterminé par la projection, sur la courbe psychométrique, du point correspondant à MOS_{min2}. Selon la gamme de mesures qu'on décide de fixer, les points de meilleure qualité et de plus mauvaise qualité n'auront pas la même position sur la courbe, ce qui décale les deux tangentes en ces points. Par conséquent, le seuil n'est plus le même. On note donc l'importance des conditions initiales notamment la plage de dégradations soumise aux observateurs, dans la validité du seuil.

Nous présentons, dans la suite, deux études portant sur la qualité des vidéos chirurgicales. Une première étude, préliminaire, a été menée au démarrage du projet RALTT et

a permis de collecter une base de notes de la qualité. Cette base de données n'a pas pu être exploitée statistiquement dans le passé. Dans le paragraphe 4.2, nous présentons les conditions expérimentales de cette étude ainsi que les résultats de l'exploitation statistique de la base de notes que nous avons menée dans cette thèse. Ensuite, nous décrivons dans le paragraphe 4.3, la campagne de tests que nous avons menée dans le cadre de cette thèse et les résultats qui en découlent. Nous présentons dans le paragraphe 4.6 les résultats de mesures objectives de la qualité à travers deux métriques issues de la littérature.

4.2 Etude 1 : Evaluation subjective de la qualité de vidéos compressées MPEG-2

4.2.1 Environnement de test

Pour juger de la qualité des vidéos chirurgicales, nous avons jugé nécessaire que les chirurgiens soient dans des conditions de fonctionnement habituelles. Par conséquent, le test s'est déroulé dans un bloc opératoire et les vidéos sont visualisées dans la console de visualisation du chirurgien comme c'est le cas en routine. Cependant, l'environnement du bloc opératoire n'est pas aussi maîtrisé qu'un laboratoire d'essais subjectifs tel qu'il est préconisé par l'ITU. Pour cette première étude, nous estimons néanmoins être en conformité avec l'idée directrice de la norme. De ce fait, on permet aux chirurgiens de noter des séquences vidéo sans les influencer : ils visionnent les séquences dans un environnement auquel ils sont habitués.

4.2.2 Sélection des participants

Les participants peuvent être experts ou non experts de l'image. Dans notre application, nous faisons appel à des experts de l'image médicale (experts chirugiens). Tous les participants doivent être examinés pour leur acuité visuelle (Test de Monoyer) et leurs défauts de perception des couleurs (Test Ishiara). Pour plus de fiabilité des résultats, un panel entre 15 et 24 participants va donner des résultats exploitables statistiquement. Un observateur habituel peut voir des artéfacts car il n'a pas d'a priori sur la visualisation

des séquences vidéo. Le panel doit également être représentatif en âge, genre et expérience. Pour l'évaluation de la qualité de la télévision numérique, la norme définit un nombre minimum de 15 observateurs s'ils sont non experts, ce nombre pouvant être plus faible (minimum 4) lorsque les observateurs sont expérimentés. Ici, 7 observateurs ont participé au test.

4.2.3 Matériel de test et déroulement

La session d'essais subjectifs est limitée à 38 minutes pendant lesquelles les 7 observateurs ont noté 38 présentations provenant de 4 scènes typiques de chirurgie selon l'enchaînement décrit dans le chapitre 3. Au début de la séance, on procède à 5 présentations fictives pour stabiliser les jugements des observateurs et dont on ne tiendra pas compte dans le dépouillement des résultats. Dans cette étude, $L = 38$, $J = 25$, $K = 28$ et $R = 3$. Nous avons également 4 scènes différentes compressées à un débit variant de 1,2 Mbits/s et 8 Mbits/s (cf. table 4.1). Ces 4 scènes sont représentatives de la diversité des vidéos chirurgicales :

- scène 1 : représente un mouvement d'outil laparoscopique,
- scène 2 : représente un cas de coagulation du sang,
- scène 3 : représente une zone graisseuse,
- scène 4 : représente l'application d'une compresse.

Ces séquences ont la particularité de contenir des mouvements de faible amplitude (application d'une compresse, mouvement d'outil laparaoscopique), une dominance de la composante de couleur rouge (coagulation du sang) et des textures aléatoires et souvent peu homgènes (tissu graisseux). Toutes ces spécificités constituent un défi à la fois pour le codeur (et la nature des artéfacts dûs à la compression) et pour la perception humaine des dégradations.

Par ailleurs, les séquences sont compressés par une carte de compression MPEG-2 (hardware) à différents débits puis transmises à la console du robot pour être visualisées par les sept observateurs. La vidéo initiale est constituée de 720 pixels sur 576 lignes codés sur 10 bits, ce qui représente un débit de 270 Mbits/s.

TABLE 4.1. *Débits associés à chaque séquence*

Scène	Débit (Mbits/s)							
1	1,02	3	4,02	5,04	6	8,04		
2	1,5	2,56	4,5	5,52	6,54	7,5		
3	1,74	2,76	3,78	4,74	5,76	6,78	7,74	
4	1,2	2,04	2,28	3,24	4,25	5,28	6,24	7,2

4.2.4 Résultats expérimentaux

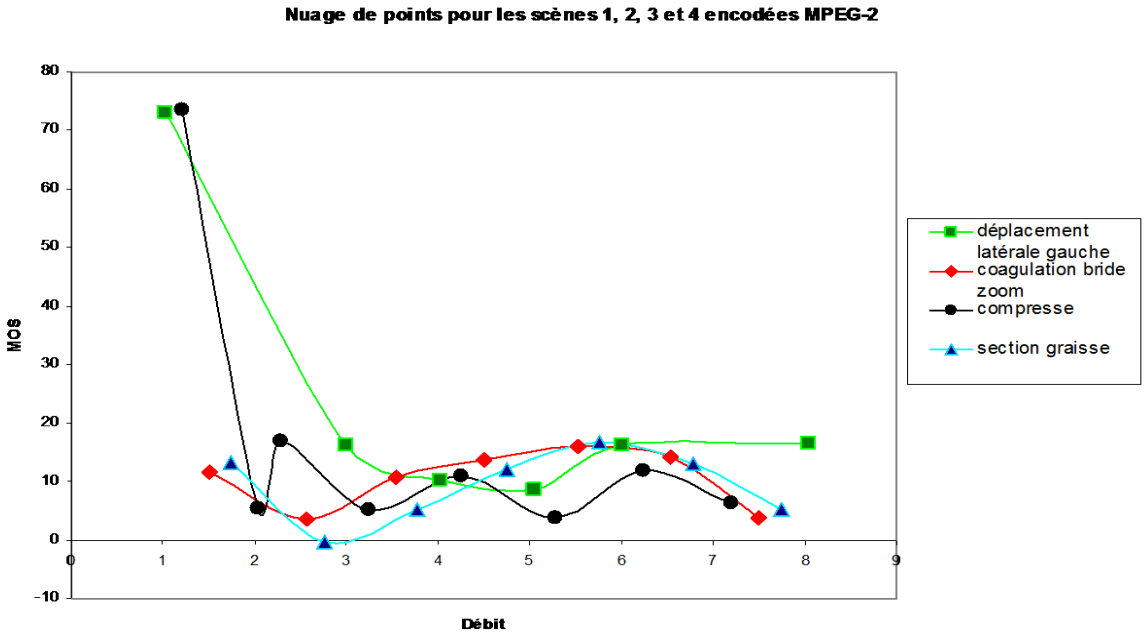


FIGURE 4.2. *Etude 1 : MOS en fonction du débit pour les 4 scènes encodées MPEG-2*

La figure 4.2 représente les nuages de points de la moyenne des notes (MOS) en fonction du débit associé à chaque séquence. On rappelle que si le MOS est faible, la qualité de la vidéo a peu diminué alors que si le MOS tend vers 100, on est face à d'importantes pertes de qualité. La figure 4.2 montre que les scènes 2 et 3 ne sont pas sensibles à la variation de la qualité des vidéos. En effet, pour ces scènes, les chirurgiens observateurs n'ont pas mis en évidence une perte de qualité significative malgré les dégradations introduites

par la compression y compris pour des débits faibles (1,5 Mbits/s). On remarque également que la note moyenne pour ces deux scènes oscille entre 0 et 20 (sur une échelle de 100). Au contraire, les courbes correspondant aux séquences 1 et 4 permettent d'identifier une perte de qualité significative avec la réduction du débit après compression.

La modélisation des données expérimentales par la méthode des moindres carrés pour estimer la dispersion des valeurs du nuage de points obtenu (score moyen en fonction du débit) permet de déterminer une courbe de régression (courbe psychométrique) ainsi que le coefficient de détermination pour chacune des scènes 1 et 4. D'une valeur comprise entre 0 et 1, le coefficient de détermination est un indicateur qui permet de mesurer l'adéquation entre le modèle et les valeurs observées. Il s'agit du carré du cœfficient de corrélation de Pearson donné par :

$$R = \frac{\sum_{i=1}^{N} \left(X_i - \overline{X}\right)\left(Yi - \overline{Y}\right)}{\sqrt{\left(X_i - \overline{X}\right)}\sqrt{\left(Y_i - \overline{Y}\right)}} \tag{4.2}$$

où X sont les données observées et Y les données correspondant au modèle. Nous avons retenu, dans le cas de cette étude, un modèle *puissance* (cf. paragraphe 4.1.2).

Les figures 4.3 et 4.4 montrent le nuage de points correspondant à la variation des MOS en fonction du débit pour les scènes 1 et 4 ainsi que la courbe psychométrique suivant un modèle puissance de la forme : $MOS = k.D^a$ où k est une constante, a est un exposant et D est la valeur du débit.

Les résultats permettent d'obtenir un coefficient de détermination R1=0,79 pour la scène 1 et R4=0,95 pour la scène 4. Ces valeurs montrent une homogénéité des observations et nous permettent de déterminer avec précision la valeur du débit seuil au delà duquel les observateurs ne distinguent pas de perte de qualité de la vidéo compressée. Ce seuil correspond au point à partir duquel la détection de perte de qualité n'est plus perçue par les observateurs dans un intervalle de confiance à 95%. Ainsi, pour la scène 1, on obtient un débit seuil de 3,2 Mbits/s pour un intervalle de confiance à 95%. Tandis que pour la scène 4, on obtient une valeur seuil du débit égale à 2,9 Mbits/s pour le même intervalle de confiance. On obtient ces valeurs de seuil, en fixant une valeur minimale du MOS notée MOS_{min} (ici autour de 20%) et on considère qu'au dessus de MOS_{min},

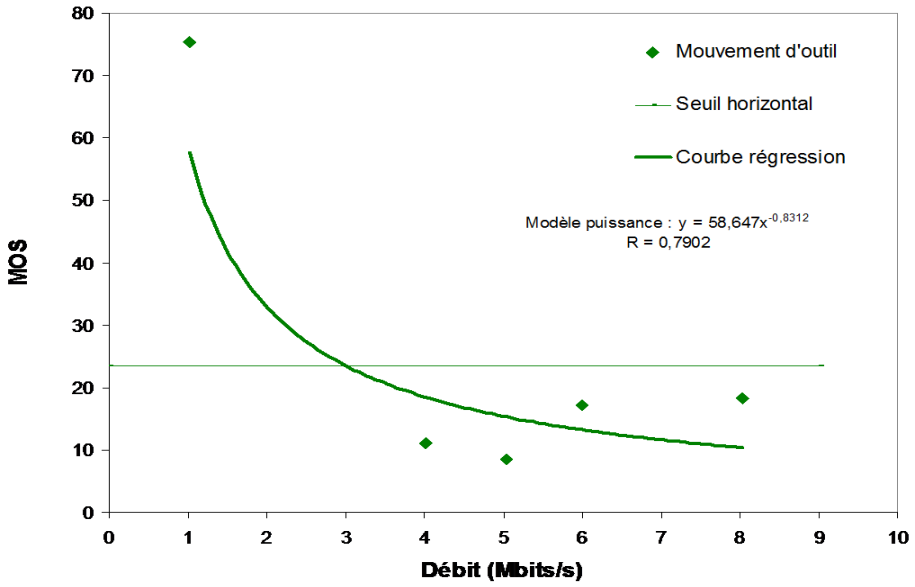


FIGURE 4.3. *Etude 1 : MOS en fonction du débit (Scène Mouvement d'outil laparoscopique, encodée MPEG-2)*

les dégradations sont perçues par les observateurs. Cette approche présente un inconvénient majeur lié au réglage de la valeur du MOS minimal car MOS_{min} dépend à la fois du jugement des observateurs, du contenu de la séquence et du modèle de la courbe psychométrique choisi.

4.2.5 Discussion

Dans le contexte de cette première étude, on peut conclure qu'à partir d'un débit de 3,2 Mbits/s, aucun praticien n'observe une baisse de qualité sur les images compressées et donc aucune gêne. Etant donné que chacune des vidéos nécessitait initialement 270 Mbits/s pour être transmise, un taux de compression autour de 90 :1 peut être adopté en pratique dans un contexte de chirurgie robotisée. Ces résultats sont valables pour des séquences chirurgicales de résolution SD (Standard Definition) encodées MPEG-2. Les choix retenus pour cette étude ne sont pas tous optimaux en particulier ceux concernant l'environnement de test, la criticité des séquences de test choisies et la méthode de déter-

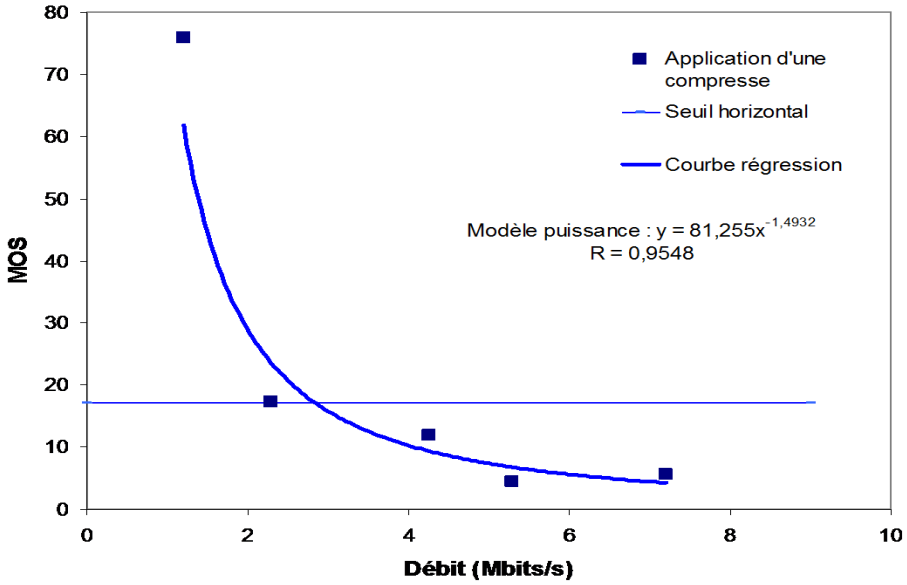


FIGURE 4.4. *Etude 1 : MOS en fonction du débit (Scène Application d'une compresse médicale, encodée MPEG-2)*

mination du seuil. Cependant, les résultats obtenus sont prometteurs, dans un contexte médical et plus précisément dans un contexte de chirurgie à distance. Ils permettent d'ouvrir des perspectives pour une compression des vidéos chirurgicales avec pertes et d'envisager l'utilisation d'autres standards de compression vidéo comme H.264. Ces premiers résultats ont permis de valider l'approche méthodologique utilisée et d'élaborer un plan d'étude approfondie pour la seconde campagne d'expérimentations (étude 2).

4.3 Etude 2 : Evaluation subjective et mesure objective de la qualité de vidéos compressées

Dans le paragraphe suivant, nous allons décrire les conditions expérimentales de notre étude subjective de la qualité des vidéos chirurgicales compressées avec le standard H.264.

4.4 Description des conditions expérimentales

4.4.1 Introduction

La première étude de la qualité des vidéos compressées MPEG-2 a mis en évidence la possibilité de compresser des vidéos médicales en identifiant le débit seuil à partir duquel une opération chirurgicale à distance est envisageable tout au moins concernant la qualité des vidéos compressées en vue de leur transmission. Cette étude a permis d'ouvrir des perspectives de tests subjectifs de la qualité de vidéos compressées avec le standard MPEG-4 AVC/ H.264. Dans ce qui suit, nous décrivons l'environnement, les observateurs, le matériel et les conditions de test pour l'évaluation subjective de la qualité des vidéos chirurgicales compressées H.264.

4.4.2 Environnement de test

Dans cette étude nous avons souhaité maîtriser l'environnement de test. En effet, contrairement au bloc opératoire (cas de l'étude 1), où les observateurs, la luminosité, la distance à l'écran et le réglage du dispositif d'affichage sont en conformité avec les paramètres définis par la norme (cf. chapitre 3). De plus, le déroulement du test dans le bloc opératoire a présenté un inconvénient pour recueillir les notes car les observateurs devaient quitter la console de visualisation du robot de chirurgie pour noter les séquences présentées au risque d'introduire des biais à leur jugement. Ici, au contraire, la disposition de la pièce dans laquelle les tests se sont déroulés (comme le montre la figure 4.5) permet aux chirurgiens de noter les séquences plus confortablement.

4.4.3 Sélection des participants

Un panel de 16 observateurs représentatifs de la population des chirugiens (spécialistes en chirurgie générale et digestive, en urologie et en O.R.L) a été sélectionné. Ce panel a été équilibré en :

- Genre : Homme (H), Femme (F),
- Age : entre 20 et 30 ans (1), entre 30 et 40 ans (2), entre 40 et 50 ans (3) et supérieur à 50 ans (4),

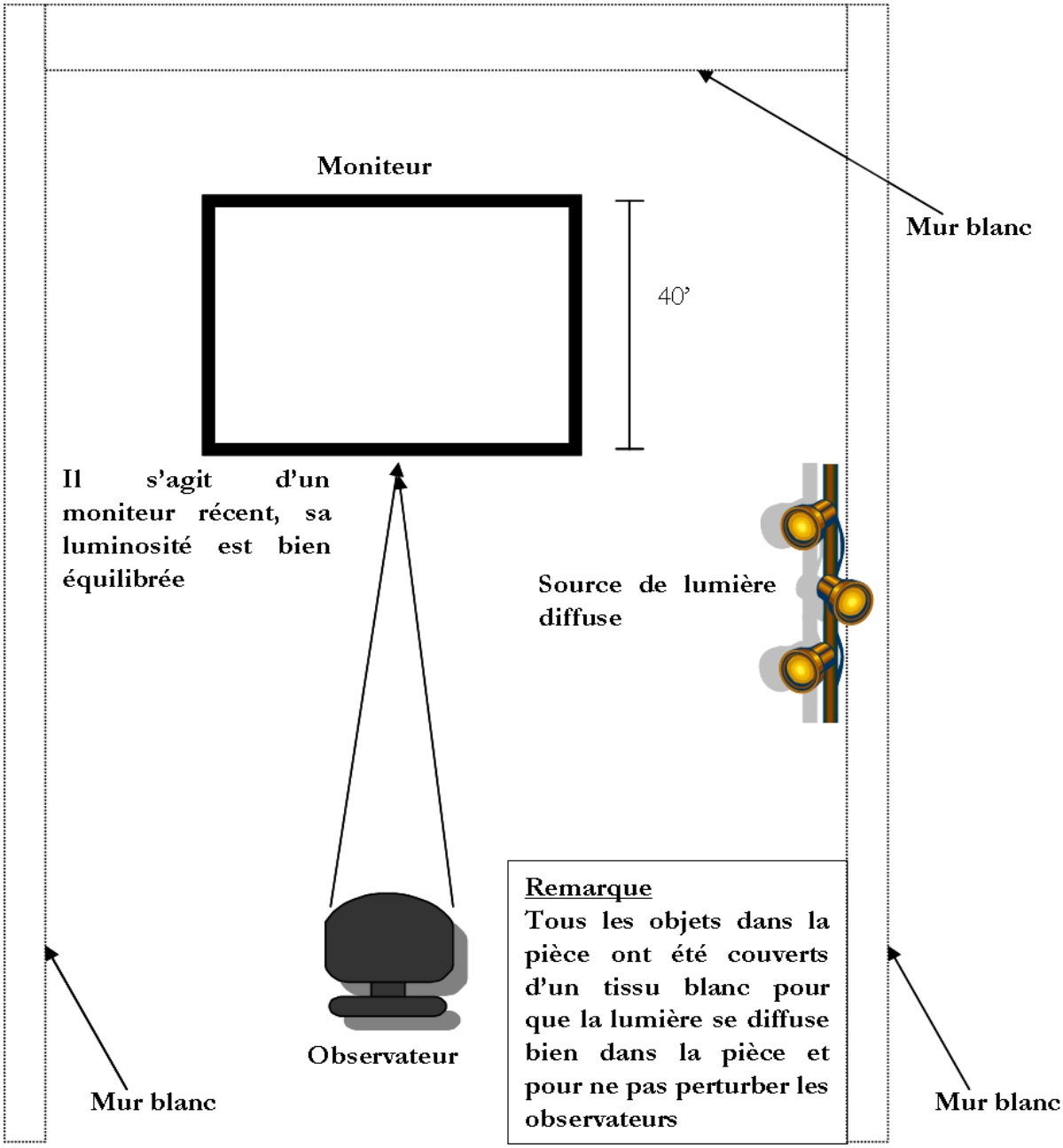


FIGURE 4.5. *Environnement de test*

– Expérience : débutant (D), expérimenté (E), senior(S).

Le tableau 4.2 résume la situation de chaque observateur du panel. Il est à noter que le panel choisi est représentatif de l'organnigramme type d'un service de chirurgie au CHU de Nancy. Il paraît évident que la pyramide des âges des observateurs, dans ce

contexte, varie entre 20 et 65 ans. En conséquence, le panel sélectionné, représentatif de la population de chirurgiens, n'inclut pas d'observateurs mineurs (< 18 ans) ou de retraités.

Par ailleurs, les chirurgiens seniors sont aussi considérés comme expérimentés mais disposent d'une plus grande expérience. Ces derniers ont généralement acquis une longue expérience en chirurgie ouverte avant d'utiliser les robots de chirurgie.

TABLE 4.2. *Observateurs*

Observateur	Genre	Age	Expérience
1	F	3	S
2	H	1	D
3	F	2	E
4	H	1	D
5	F	1	E
6	F	2	E
7	F	2	E
8	H	4	S
9	F	2	E
10	H	1	D
11	H	1	D
12	H	3	S
13	F	2	E
14	F	2	E
15	H	2	E
16	H	1	D

4.4.4 Matériel de test et déroulement

Pendant deux sessions de 35 minutes, les 16 observateurs ont noté 70 présentations provenant de 4 scènes typiques de chirurgie selon le même enchaînement que celui décrit dans le chapitre 3. Avant le début de chaque session, un document comportant une expli-

cation des conditions de test et de son contexte ainsi que l'importance de l'expérience est présenté aux observateurs (voir Annexe A). On procède ensuite à 3 présentations fictives dont on ne tiendra pas compte dans le dépouillement des résultats, afin de stabiliser les jugements des observateurs.

Dans cette étude, nous avons $L = 70$ présentations (c'est-à-dire les paires de séquences référence/séquence de test), $J = 67$ conditions de test (y compris la référence), $K = 19$ conditions de tests effectives (ici, il s'agit des niveaux de dégradation) et $R = 4$ répétitions (chaque condition de test a été répétée pour les 4 séquences). Nous avons également 4 scènes différentes compressées à un débit variant linéairement et présentées aléatoirement. Les débits de compression H.264 que nous avons choisis se situent dans une plage allant de 0,3 Mbits/s à 3,6 Mbits/s avec un pas linéaire de 20%. Les performances de H.264 énoncées dans la littérature, nous incitent à compresser plus fortement les séquences vidéos de cette étude par rapport aux taux de compression de l'étude 1.

Nous avons sélectionné quatre scènes en collaboration avec des experts chirurgiens. Ces scènes sont représentatives de la diversité des vidéos chirurgicales et elles mettent en évidence les gestes effectués en chirugie (à travers la dissection, la section et la suture) :

- scène 1 (Dissection) : décollement d'un espace graisseux (Figure 4.6),
- scène 2 (Contrôle de vaisseaux) : contrôle de vaisseaux (Figure 4.7) ,
- scène 3 (Section) : ablation d'une tumeur (Figure 4.8),
- scène 4 (Suture) : contrôle du saignement et coagulation (Figure 4.9) .

Nos critères de choix des séquences sont également liés à la contrainte imposée par l'application de chirurgie à distance, ce qui nous amène à choisir :

- des gestes typiques de chirurgie caractérisés par une faible amplitude de mouvement, notamment en raison de la faible superficie du champ opératoire où le geste est effectué (une fenêtre d'observation de quelques centimètres) ;
- une variation du débit linéaire entre 0,35 Mbits/s et 3,6 Mbits/s : l'étude précédente a démontré qu'au delà de 3 Mbits/s pour MPEG-2, aucune gêne n'est perceptible pour les chirurgiens.

Les séquences sont compressés avec deux logiciels de compression H.264 (software) aux différents débits. Ces logiciels sont : la bibliothèque **x264** issue du projet VideoLAN

[VideoLAN, 2010] et le logiciel de référence H.264 **JM Software** [hhi, 2010]. Par ailleurs, les séquences ont été également compressées par un logiciel MPEG-2 (FFMPEG) à des débits variant de 1,74 à 3,61 Mbits/s. Ces logiciels constituent des références dans la communauté des chercheurs en imagerie.

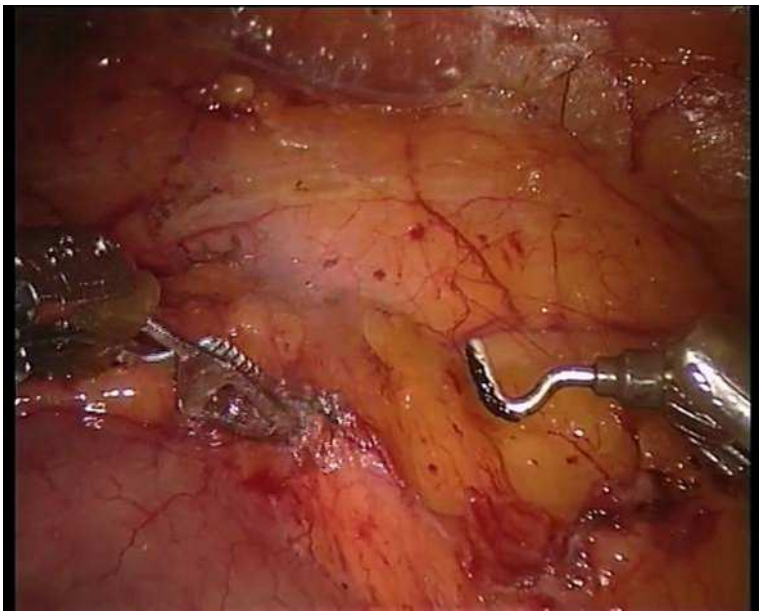


FIGURE 4.6. *Dissection*

4.4.5 Spécificités de l'étude

Nos critères de choix pour les débits de compression ainsi que le choix des codeurs peuvent être une source de biais. De plus, les séquences compressions ont été numérisées en amont : le passage du monde analogique au monde numérique peut introduire des artéfects dans les séquences. Cependant, les observateurs sont habitués à ce type de traitement de la vidéo chirurgicale. D'autres biais possibles proviennent de l'environnement de test, des dégradations choisies ou du contenu des vidéos chirurgicales.

D'autre part, la provenance des chirurgiens, tous excerçant dans le même établissement hospitalier (CHU de Nancy), peut constituer un biais supplémentaire. En effet, nous supposons que le panel sélectionné est représentatif de tous les chirurgiens. Or, une étude de [Norenzayan *et al.*, 2002] a démontré l'existence de différences dans l'attention portée

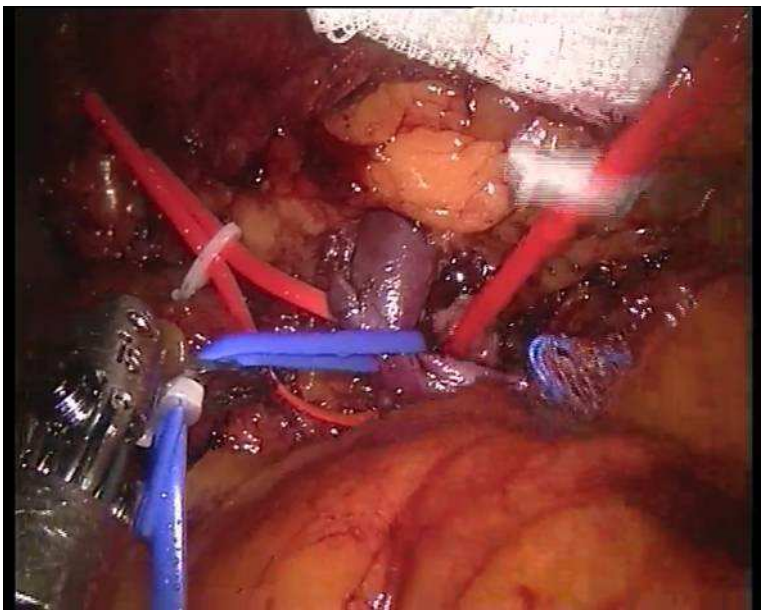


FIGURE 4.7. *Contrôle de vaisseaux*

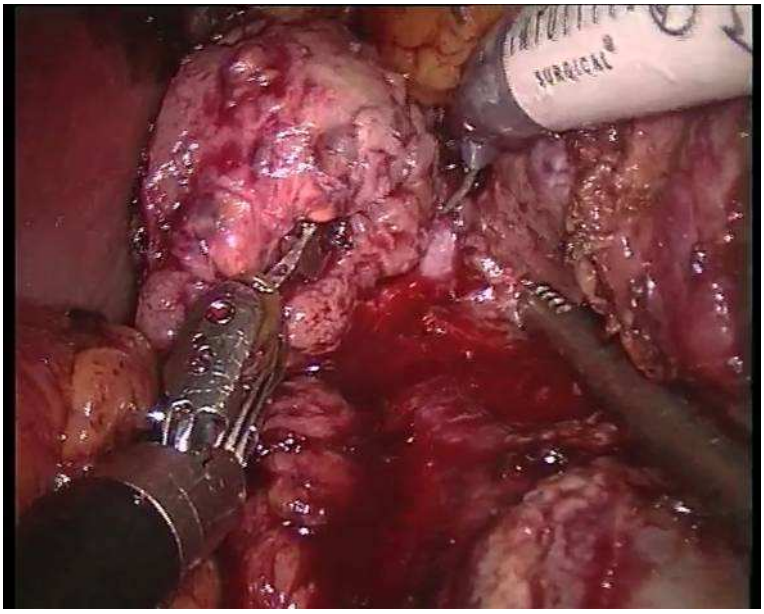


FIGURE 4.8. *Section*

à des photos (de type grand public) par des étudiants de nationalités différentes, en suivant leurs mouvements oculaires lors de l'observation . Alors que les uns concentraient leur attention sur le fond de l'images, les autres regardaient les objets. Dans le cas de la

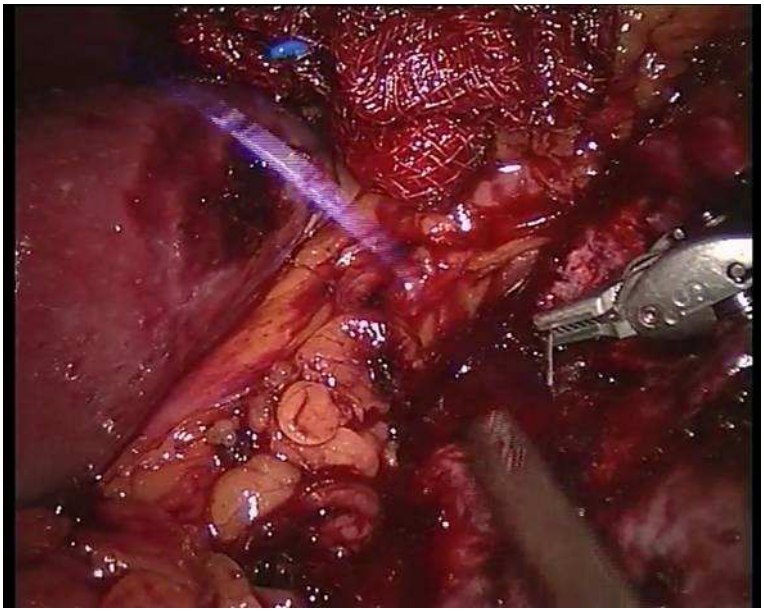


FIGURE 4.9. *Suture*

chirurgie, il est possible que le comportement des chirurgiens face à une même séquence soit légèrement différent en raison des cultures d'apprentissage différentes. Par ailleurs, la nature même du test subjectif, où l'on soumet des vidéos chirurgicales à des chirugiens est une éventuelle source de biais car la séquence doit être notée sans tenir compte de son contenu. C'est pour cette raison, qu'en début de chaque session, les observateurs sont informés qu'ils ne doivent pas juger le geste chirugical présenté dans la séquence mais la qualité globale de la vidéo présentée. Le déroulement de l'étude sur la qualité des vidéos chirugicales est une transposition de la norme initialement prévue pour tester la qualité des vidéos grand public (Télévision Numérique, Multimédia, vidéo sur IP).

4.4.6 Traitement des données expérimentales

La base des notes obtenue après le déroulement du test est exploitée de manière identique à celle décrite dans le chapitre 3. D'abord, la cohérence des résultats est vérifiée en étudiant les notes données par le même observateur à la même séquence pendant la même séance. Si les notes diffèrent de 2 points ou plus (pour une échelle allant de 1 à 5), ces notes seront rejetées. Ensuite, le test de normalisté (ou test du β_2 tel qu'il est décrit

dans la norme [ITU-R, 2000]) est réalisé. Il permet de tester si la distribution des notes suit une loi normale et si les notes des observateurs sont cohérentes.

Par ailleurs, deux autres outils peuvent être utilisés en complément. Le premier, le Z-score, permet de minimiser les variations entre les notes individuelles dûes à la non utilisation des échelles entières par les observateurs. Le second, le test de Student, précise, d'un point de vue statistique, si les valeurs moyennes associées à un niveau de dégradation j sont discernables ou non entre elles.

Enfin, l'exploitation statistique de la base de notes permettra d'obtenir deux types majeurs de résultats : déterminer le seuil de compression H.264 toléré par les chirurgiens et évaluer les performances de métriques objectives de la qualité sur les vidéos chirurgicales. Ces résultats sont détaillés dans le paragraphe 4.5.

4.5 Résultats expérimentaux

Suite au test subjectif d'évaluation de la qualité des vidéos chirurgicales suivant la méthodologie DSCQS, la première étape consiste à déterminer la note de qualité de chaque séquence présentée. Le MOS est estimé par la moyenne des jugements des observateurs, qui ont noté indépendamment les séquences pour chacune des dégradations (on rappelle qu'il s'agit des débits de compression). Pour être fiable, on associe à chaque note moyenne un intervalle de confiance à 95% : intervalle dans lequel se trouvent 95% des réponses et en considérant que les moyennes suivent une loi gaussienne [ITU-R, 2000]. A cause de la multiplicité des facteurs ayant une influence sur le test (cf. chapitre 3), les notes de qualité des observateurs peuvent être biaisées : les observateurs peuvent ne pas utiliser la même dynamique sur l'échelle de notation ou noter des séquences plus ou moins sévérement. Il peut arriver aussi qu'un observateur fournisse des résultats non cohérents en donnant, par exemple, des notes différentes pour des images ayant subi les mêmes niveaux de dégradation. La procédure préconisée par l'ITU est appliquée afin de rejeter ces notes incohérentes et des observateurs incohérents. Enfin, une transformation des notes, par une transformée Z-score, peut s'avérer utile si les observateurs n'utilisent pas la même dynamique pour les notes (utilisation de seulement une partie de l'échelle

de notation par exemple). Cette méthode sert à corriger les notes de chaque observateur pour les ramener à une même dynamique comparable entre observateurs. Un Z-score donne une information sur la différence normalisée d'une note par rapport à la moyenne d'un observateur. Le Z-score est calculé de la manière suivante :

$$Z_{ijk} = \frac{X_{ijk-\overline{X}_i}}{\sigma_i} \tag{4.3}$$

avec :

$$\overline{X}_i = \frac{1}{N_{deg}} \cdot \frac{1}{N_{im}} \sum_{j=1}^{N_{deg}} \sum_{k=1}^{N_{im}} X_{ijk} \tag{4.4}$$

$$\sigma_i^2 = \frac{1}{N_{deg}-1} \cdot \frac{1}{N_{im}-1} \sum_{j=1}^{N_{deg}} \sum_{k=1}^{N_{im}} \left(X_{ijk} - \overline{X}_i\right)^2 \tag{4.5}$$

où :

- N_{deg} est le nombre de dégradations par vidéo originale,
- N_{im} est le nombre de vidéos originales,
- X_{ijk} est la note donnée par l'observateur i, à une vidéo (j,k) dégradée ou non.

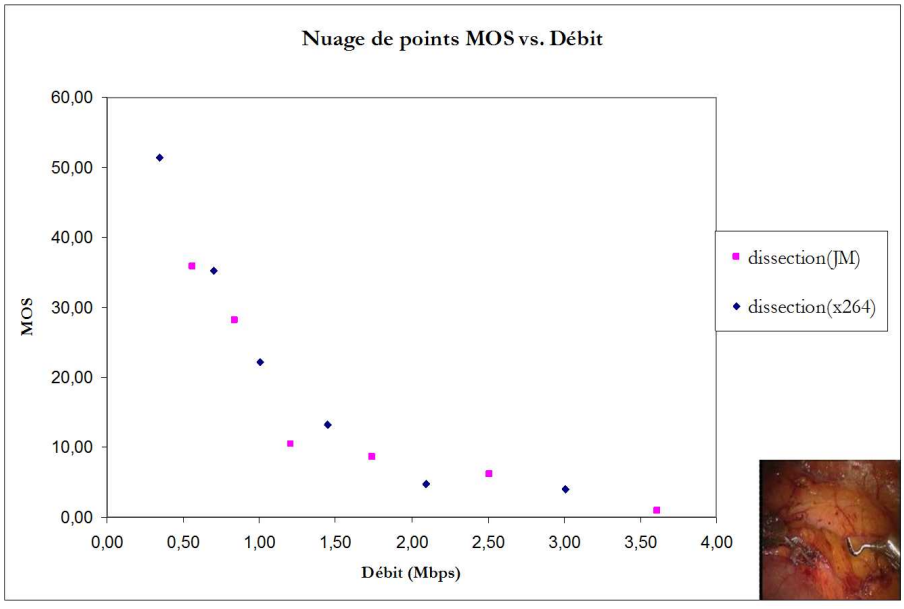


FIGURE 4.10. *MOS en fonction du débit (Dissection)*

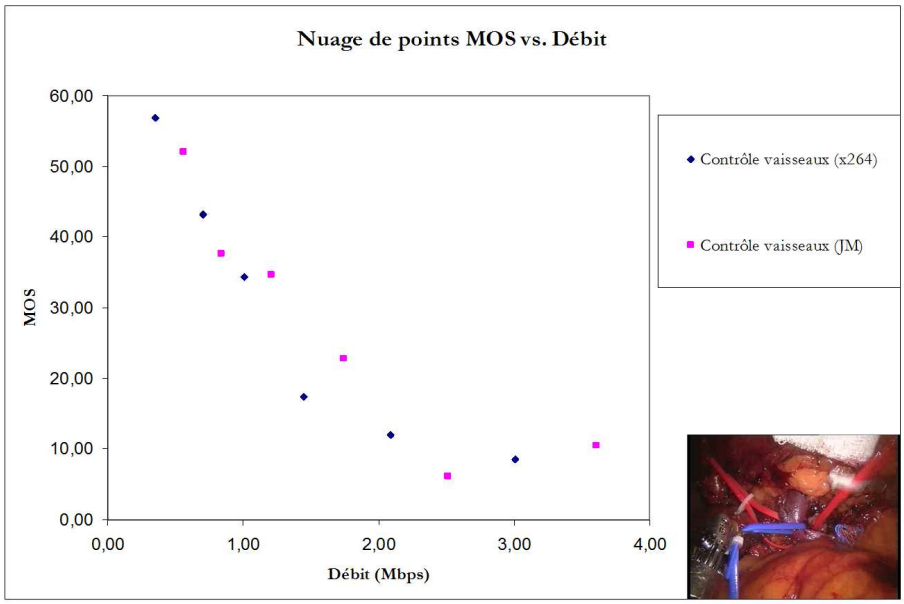


FIGURE 4.11. *MOS en fonction du débit (Contrôle de vaisseaux)*

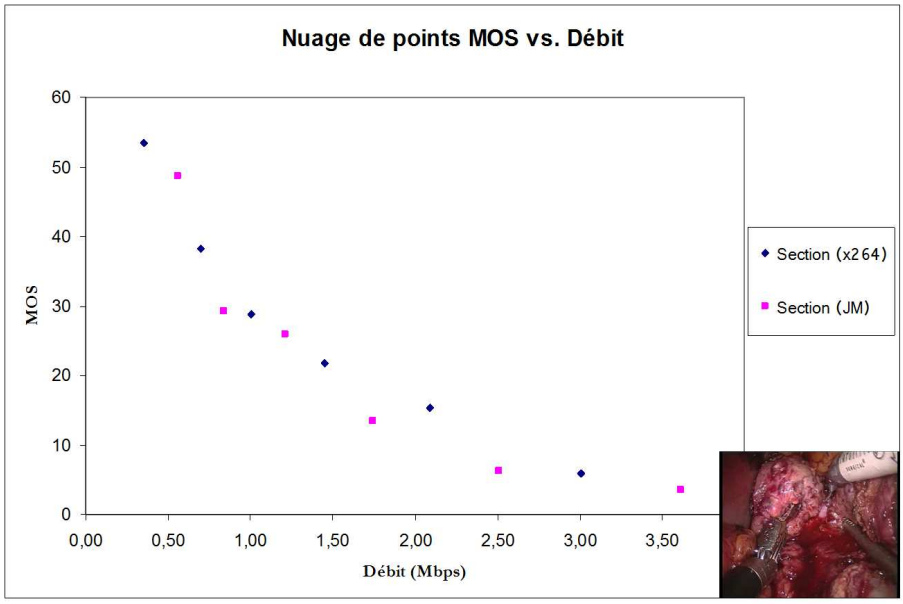


FIGURE 4.12. *MOS en fonction du débit (Section)*

Les figures 4.10, 4.11, 4.12 et 4.13 représentent le nuage de points illustrant la dispersion de la moyenne des notes en fonction du débit associé à chaque séquence. On rappelle que si cette note est faible, la qualité de la vidéo a peu diminué alors que si la note tend

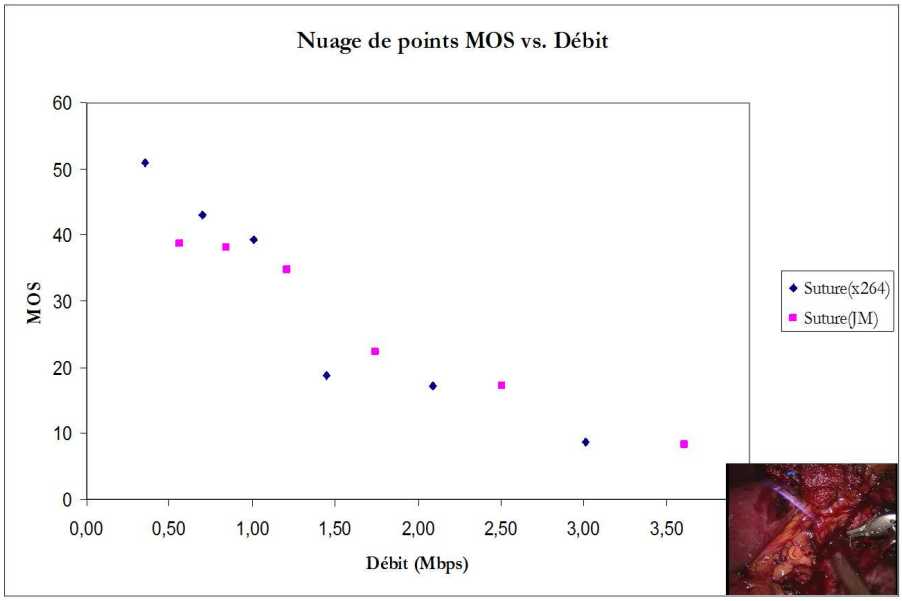


FIGURE 4.13. *MOS en fonction du débit (Suture)*

vers 100, on est face à d'importantes pertes de qualité.

Dans cette étude, on identifie sur chacune des figures 4.10, 4.11, 4.12 et 4.13 une perte de qualité significative avec la réduction du débit après compression. La modélisation des données expérimentales par la méthode des moindres carrés pour estimer la dispersion des valeurs du nuage de points obtenu (score moyen en fonction du débit) permet de déterminer une courbe de régression ainsi que le coefficient de détermination pour chacune des scènes. Dans le cas de notre étude, le modèle de Stevens est le plus adapté. En effet, la modélisation des données expérimentales démontre un comportement asymptotique du modèle quand $x \to 0$ et $x \to \infty$ avec une décroissance des notes (car on évalue la perte de qualité par rapport à une référence). Les tableaux 4.3 et 4.4 mettent en évidence les cœfficients de régression R pour les quatres scènes utilisées dans le test, compressées avec les logiciels *x264* et *JM reference software* et notées par le panel de 16 observateurs.

Ces valeurs montrent une homogénéité des observations et nous permettent de déterminer avec précision la valeur du débit seuil au delà duquel les observateurs ne distinguent pas de perte de qualité de la vidéo compressée. Ce seuil correspond au point à partir duquel la détection de perte de qualité n'est plus perçue par les observateurs dans

Scène	R
Dissection	0,9673
Contrôle vaisseaux	0,9702
Section	0,9444
Suture	0,9817

TABLE 4.3. *Cœfficients de régression R, compression H.264 (x264)*

Scène	R
Dissection	0,9380
Contrôle vaisseaux	0,8944
Section	0,9817
Suture	0,9289

TABLE 4.4. *Cœfficients de régression R, compression H.264 (JM)*

un intervalle de confiance à 95%.

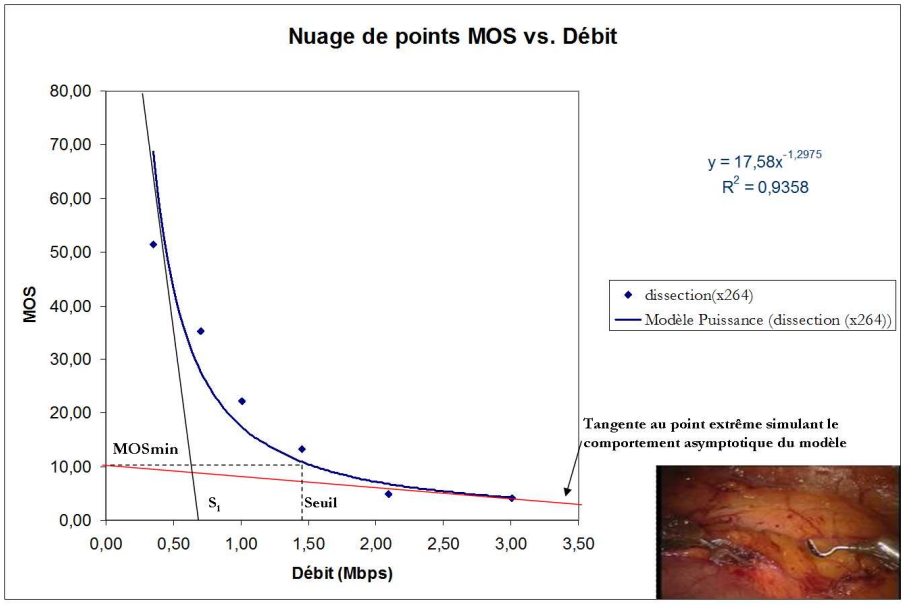


FIGURE 4.14. *Seuil de compression pour la séquence dissection (Modèle¨Puissance)*

La figure 4.14 montre que le seuil S de perception de la perte de qualité des vidéos chirugicales est situé au débit de compression de 1,5 Mbits/s (selon la deuxième approche décrite dans le paragraphe 4.1.2).

4.6 Evaluation des performances des métriques objectives

Afin d'évaluer la pertinence des métriques objectives, il est nécessaire de confronter les jugements humains avec les notes objectives (prédites). Nous avons choisi de mesurer la qualité des séquences évaluées subjectivement dans l'étude 2, par deux méthodes objectives décrites dans le chapitre 3 : PSNR et SSIM. Pour évaluer les performances de prédiction des méthodes objectives, on dispose de deux bases de notes. Les MOS, qui représentent le score moyen donné par les observateurs humains d'une part, et les notes objectives, notées VQR (Video Quality Rating), d'autre part.

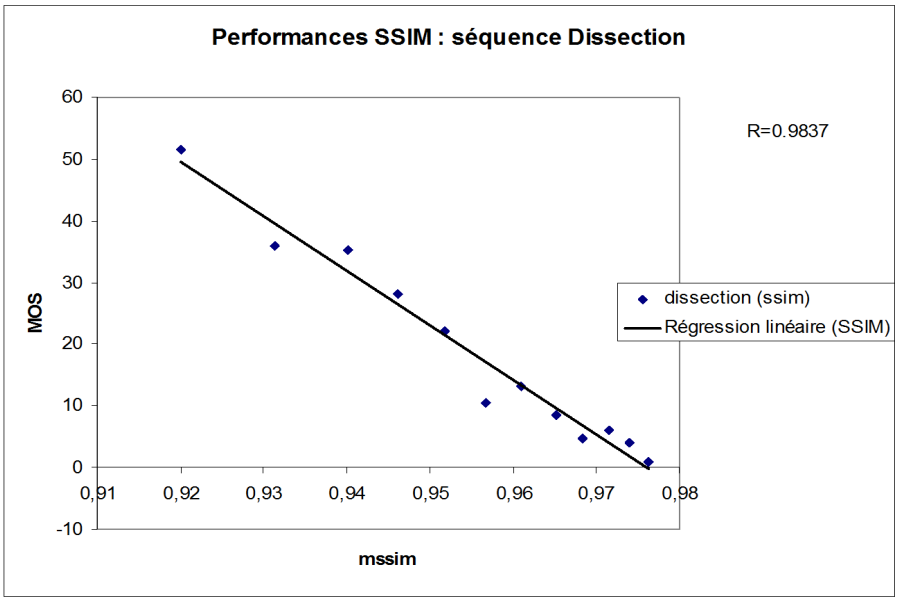


FIGURE 4.15. *Corrélation SSIM vs. MOS (Dissection)*

Les métriques de qualité objectives renvoient un paramètre VQR : la corrélation entre le MOS et le VQR doit être prédictible et répétable. Il s'agit de déterminer un cœfficient de corrélation linéaire entre le paramètre VQR et le MOS (Figure 4.15 pour la séquence *Dissection*). Pour la métrique SSIM, le cœfficient de corrélation est proche de 0,9 pour les

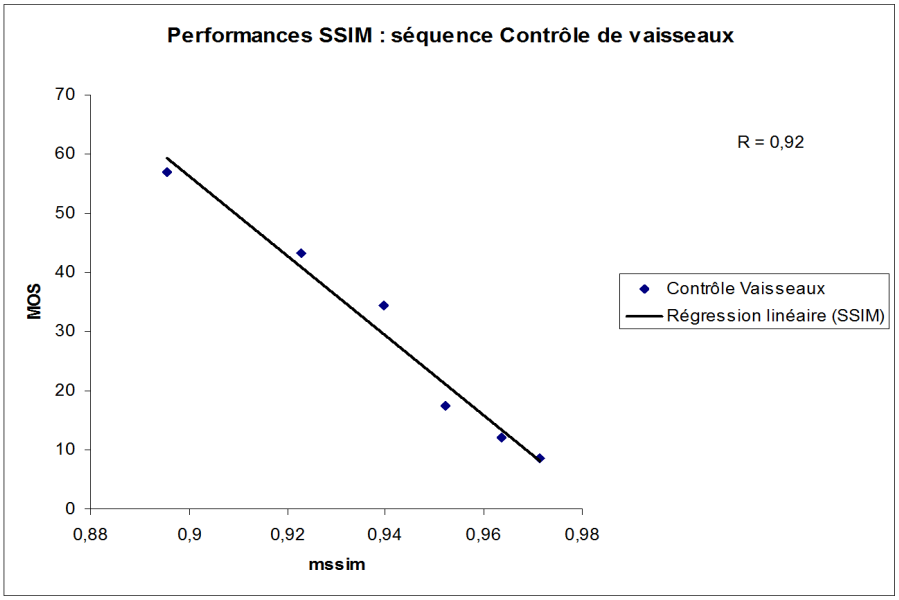


FIGURE 4.16. *Corrélation SSIM vs. MOS (Contrôle de vaisseaux)*

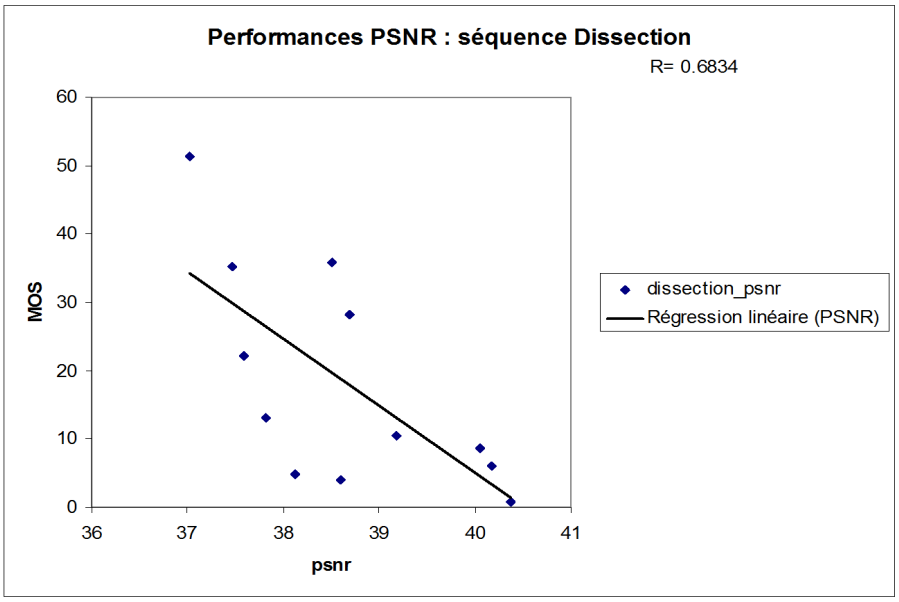


FIGURE 4.17. *Corrélation PSNR vs. MOS (Dissection)*

séquences *Dissection* et *Contrôle de vaisseaux* comme le montrent les figures 4.15 et 4.16. La figure 4.17 montre la régression linéaire entre les notes obtenues par les observateurs et les résultats de la métrique PSNR. Le cœfficient de corrélation pour l'adéquation du

modèle aux données est autour de 0,6. Ceci confirme que la corrélation entre le PSNR et les jugements humains

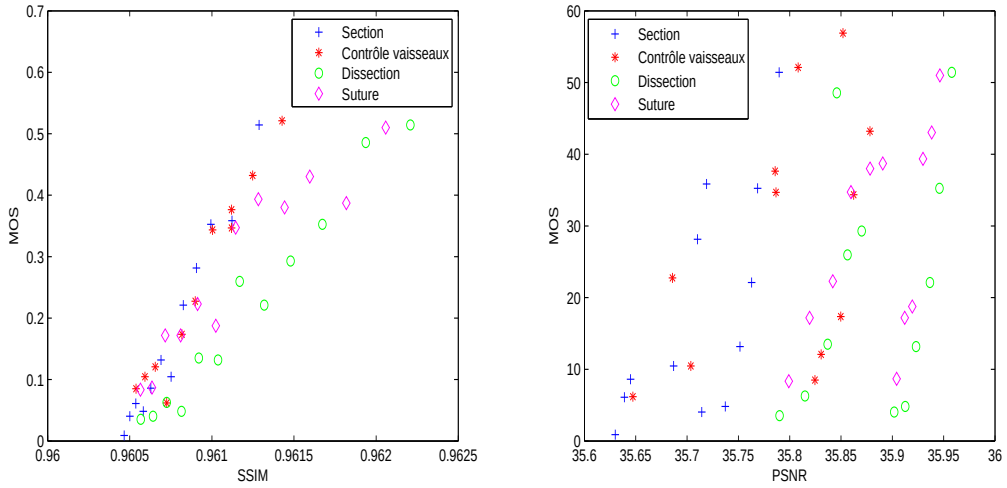
Par ailleurs, les performances des méthodes objectives de la qualité sont évaluées à l'égard de trois aspects de leur capacité à approcher le score moyen d'opinion (MOS) obtenu en moyennant les notes des observateurs lors d'un test subjectif de la qualité. Ces aspects, énoncés dans le chapitre 3 sont :

- Précision de l'estimation ;
- Monotonie de l'estimation ;
- Consistence de l'estimation.

Le VQEG [VQEG, 2008] recommande plusieurs indicateurs pour évaluer les performances d'une métrique de qualité et exprimer les trois critères : précision, monotonie et consistence des notes par rappport au jugement humain. Le coefficient de corrélation linéaire (*Pearson linear correlation coefficient*), noté LCC exprime la dépendance linéaire entre les mesures objectives et les notes subjectives. C'est un indicateur de précision. Le coefficient de corrélation de rang (*Spearman rank order correlation coefficient*), noté SROCC, est une mesure de la monotonie. Il caractérise le degré avec lequel les mesures objectives et les notes subjectives évoluent dans le même sens. Un coefficient de corrélation de rang proche de 1 signifie que la métrique de qualité classe les vidéos, selon leur qualité, selon le même ordre que les observateurs. (SROCC= -1 indiquerait un classement dans l'ordre inverse, ce qui est le cas lorsque le MOS représente la perte de qualité alors que la métrique objective mesure la fidélité par rapport à une référence). L'indicateur de cohérence OR (*Outlier Ratio*) permet de mesurer l'aptitude de la métrique objective à prédire une note de qualité qui soit proche du MOS. Il s'agit de la proportion de notes aberrantes (en dehors de l'intervalle de confiance à 95%). La valeur de l'indicateur OR doit être faible. Le VQEG recommande également d'utiliser une transformation non linéaire qui permet de passer des mesures objectives de qualité VQR à des notes prédites de qualité (MOSp) pour les comparer avec les MOS (ceci permet de mieux tenir en compte de la façon dont le jugement humain est construit). La transformation non linéaire s'exprime sous forme de fonction logistique (psychométrique) :

$$MOSp = \frac{b_1}{1 + \exp^{-b_2(VQR - b_3)}} \tag{4.6}$$

où b_1, b_2 et b_3 sont les trois paramères de la fonction logistique. Cette transformation permet de transposer à la même dynamique toutes les notes objectives quelle que soit la métrique. De plus, elle permet d'effectuer une correction de la dynamique des métriques.



(a) (b)

FIGURE 4.18. *Dispersion du nuage de points des couples (MOS, MOSp) : MOSp est la prédiction du SSIM ou du PSNR*

La figure 4.18 illustre les couples (MOS, MOS_p) et la dispersion

Tous ces paramètres ont été évalués pour les deux métriques : PSNR et SSIM. Dans le tableau 4.6, nous donnons les valeurs de ces paramètres de performance pour la métrique SSIM et dans le tableau 4.5 les performance du PSNR.

Scène	LCC	SROCC	OR
Dissection	0,68353	0,67133	0
Contrôle vaisseaux	0,4952	0,45455	0
Section	0,37495	0,45455	0
Suture	0,55871	0,65149	0

TABLE 4.5. *Performances de la métrique PSNR*

Le PSNR ne conduit pas à une bonne prédiction de la qualité. En effet, le cœfficient de

Scène	LCC	SROCC	OR
Dissection	0,98366	0,98601	0
Contrôle vaisseaux	0,9763	0,95622	0
Section	0,98493	0,97902	0
Suture	0,94156	0,95622	0

TABLE 4.6. *Performances de la métrique SSIM*

Métriques	LCC	SROCC	OR
SSIM	0,87538	0,9054	0
PSNR	0,41023	0,40212	0

TABLE 4.7. *Performances des métriques sur toute la base de test*

corrélation de Pearson (LCC) du PSNR avec les notes subjectives est seulement de 0,41. Ceci montre l'incapacité du PSNR à prédire signficativement la qualité des séquences de vidéos chirurgicales. Ses performances sont significativement inférieures à celles du SSIM, (cf. tableau 4.7). Les performances de ces métriques varient aussi en fonction du contenu des séquences comme le montrent les tableaux 4.6 et 4.5. Le PSNR a de meilleures performances pour la séquence *Dissection* qui contient peu de mouvements que pour la séquence *Suture* où le mouvement est plus important.

4.7 Discussion

Ce chapitre était dédié à l'étude de la qualité des vidéos chirurgicales. Nous nous sommes basés sur un protocole normalisé utilisé dans le domaine grand public. A notre connaissance, ce type d'étude n'a pas été mené auparavant dans le monde médical, très longtemps réticent à la compression de ses données. Le jugement de qualité obtenu nous a permis de détecter un seuil de tolérance à la compression mais nous a également renseigné sur la particularité des vidéos médicales (du point de vue de leur contenu). En effet, malgré l'hétérogénéité du jugement des observateurs, nous avons démontré qu'il

est possible de modéliser la relation entre les dégradations introduites aux vidéos (ici à travers la compression) et la perception des chirurgiens.

Par ailleurs, certains détails de ces vidéos qui peuvent paraître inutiles dans l'appréciation de la qualité globale des vidéos pour un non expert médical, le sont pour le chirurgien. Une première idée qui pourrait être mise en œvre, concerne les zones de la vidéo qui ont été effectivement regardées pour construire le jugement de la qualité. Il serait intéressant de pouvoir déterminer les zones d'attention visuelle et d'en tenir compte dans la construction d'une métrique objective de la qualité plus adaptée au contexte chirurgical. Il existe différents mécanismes d'attention visuelle, décrits dans la littérature, qui se basent essentiellment sur les mouvements oculaires. L'attention visuelle a été étudiée dans le contexte de l'évaluation subjective de la qualité par [Vuori *et al.*, 2004]. Par ailleurs, les premiers résultats obtenus pour la mesure objective de la qualité à travers la métrique SSIM, montrent qu'il est possible de mesurer objectivement la perception humaine de la qualité. Il y a une bonne corrélation entre SSIM et le jugement humain, ce qui en fait un meilleur indicateur de la qualité que le PSNR (ici, la corrélation entre le PSNR et le jugement humain est faible). Cependant, les métriques objectives, y compris SSIM, ont leurs limites. En effet, la mesure de la qualité est relativement complexe car la perception humaine et en particulier la perception du chirurgien est difficile à mesurer par un algorithme. Par exemple, dans le cas de la métrique SSIM, celle-ci prend en compte l'information structurelle de l'image pour déterminer un paramètre de qualité mais elle ne tient pas compte des zones d'intérêt de l'image ou de la vidéo effectivement regardées (points saillants). Davantage de recherches doivent être menées pour mieux comprendre les mécanismes de l'attention visuelle et leur intérêt dans l'évalution de la qualité des vidéos. Il semble important que les régions d'intérêt des vidéos chirurgicales et leurs niveaux de dégradations soient traités conjointement.

4.8 Conclusion

Ce chapitre a été dédié à l'étude de la sensibilité des chirurgiens à la compression vidéo. L'étude subjective de la qualité a permis de mettre en exergue la possibilité de

compresser les vidéos chirurgicales avec les standards de compression « grand public » MPEG-2 et H.264. Dans cette étude nous utilisons des séquences issues de flux vidéo au format PAL provenant d'une caméra tri-CCD et numérisés dans la chaîne de transmission du robot. Les signaux numériques obtenus sont échantillonnés selon la norme 4 :2 :2. La vidéo initiale est constituée de 720 pixels sur 576 lignes codés sur 10 bits, ce qui représente un débit de 270 Mbits/s.

Nous avons identifié le débit (après compression) seuil à partir duquel une opération chirurgicale à distance est envisageable tout au moins concernant la qualité des vidéos compressées en vue de leur transmission. Ce seuil se situe autour de 3 Mbits/s pour MPEG-2 et autour de 2 Mbits/s pour MPEG-4 AVC/ H.264. Ces résultats confirment ce qu'annonce la littérature : H.264 apporte entre 20% et 50% de gain par rapport à MPEG-2.

D'autre part, nous avons tenté d'approcher ces jugements humains à travers deux métriques issues de la littérature : SSIM et PSNR. Pour la première métrique, il existe une très bonne corrélation linéaire entre le VQR et le MOS. Pour le PSNR, et comme l'indique la littérature, cette corrélation est moins bonne.

Chapitre 5

Transmission de vidéos chirurgicales en temps réel sur un réseau IP : Cas concret

Dans ce chapitre, nous décrivons une réalisation concrète de transmission des flux vidéos issus d'un robot de chirurgie. Il s'agit d'une transmission en temps réel de vidéos issues d'opérations chirurgicales dans un objectif pédagogique. Nous définissons les spécifications fonctionnelles et techniques de cette transmission.

5.1 Contexte

Le projet RALTT comporte plusieurs étapes visant à établir la faisabilité des interventions chirurgicales à distance (plusieurs centaines de kilomètres) entre deux robots de chirurgie. Ces étapes sont :

- Le choix des méthodes de compression de vidéos (standards grand public, méthode adaptée à l'application ;
- L'étude de la limite tolérable de compression des vidéos chirurgicales ;
- Le choix et la validation de la configuration réseau adaptée pour ce type de transmission ;
- L'étude de la limite tolérable du retard de transmission ;

– La validation finale du projet (compression, transmission, restitution, adaptation des commandes du robot).

Afin d'opérer à de grandes distances, les informations numérisées (vidéo, audio, commandes, contrôle) issues du robot de chirurgie doivent transiter par des réseaux de télécommunication. La compression des flux vidéo en particulier devient de ce fait incontournable et les temps de latence induits doivent rester raisonnables pour qu'un chirurgien puisse conserver un geste sûr. Cependant, la compression doit permettre une restitution d'images d'une qualité irréprochable pour que l'acte chirurgical puisse être réalisé dans des conditions de sécurité optimales. Par ailleurs, le temps de latence entre le mouvement réalisé par le chirurgien sur la console et le retour de l'image est un facteur critique dans ce projet puisque ce délai ne doit pas constituer une gêne pour le chirurgien qui doit pouvoir être assuré de conserver un geste sûr même à distance. La figure 5.1 montre les étapes nécessaires pour valider une téléopération. L'étape **1** concerne l'acquisition de deux flux stéréoscopiques en vue de leur transmission sur un réseau de télécommunication (étape **2**) afin d'être restitués sur un dispositif d'affichage au niveau du site distant (étape **3**).

Les contraintes liées à ce schéma sont celles énoncées dans le chapitre 1. Il s'agit ici d'identifier les contraintes liées à la transmission par voie IP de données (pour l'essentiel vidéo) d'un seul flux vidéo issu du robot de chirurgie, de déterminer les paramètres de réglage de l'algorithme de compression adapté aux vidéos chirurgicales et de limiter le retard de transmission. Pratiquement, il s'agit de transmettre en temps réel des flux vidéo tout en répondant à des contraintes temporelles (temps de latence, bande passante, retard) et des contraintes de qualité visuelle essentiellement.

Le paragraphe suivant décrit les spécifications fonctionnelles et techniques de la transmission en temps réel des flux vidéo issus du robot de chirurgie (bloc opératoire du CHU de Nancy) vers la salle de conférence de l'Ecole de Chirurgie de Nancy (Faculté de Médecine).

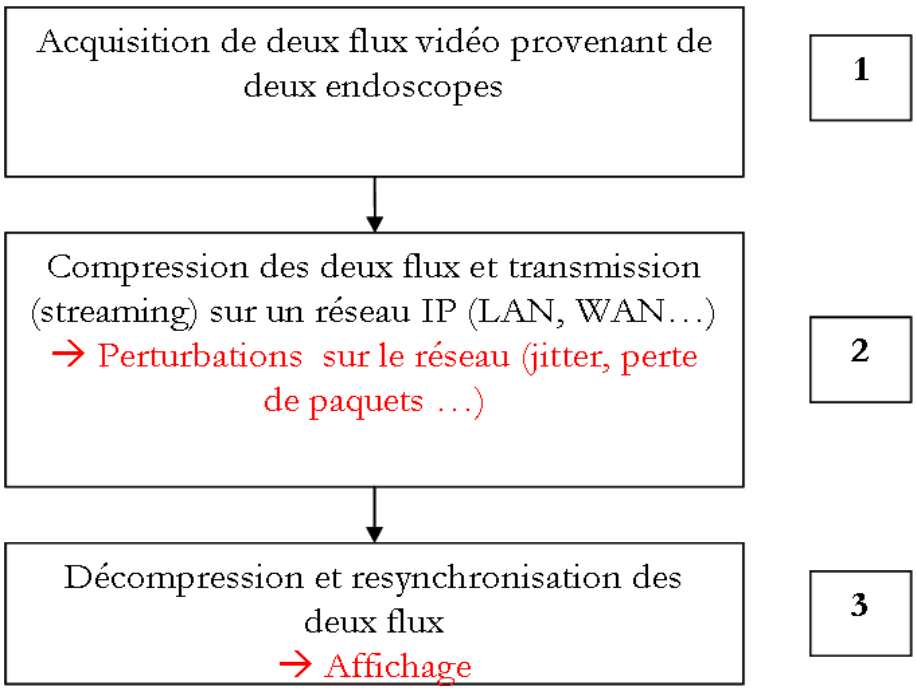


FIGURE 5.1. *Schéma général de transmission des vidéos chirurgicales*

5.2 Cahier des charges de la transmission

Dans le but de former des étudiants à la chirurgie robotique à l'école de chirurgie de Nancy, notre objectif est de mettre en place une plateforme de transmission de vidéos issues de l'endoscope filmant le champ opératoire. C'est une transmission en temps réel, où le chirurgien a la possibilité d'expliquer le geste chirurgical qu'il effectue et d'intéragir directement avec les étudiants.

5.2.1 Compression et transmission des flux vidéo

La plateforme de transmission doit répondre aux contraintes liées à la transmission des données, à leur compression et décompression. Le réseau dont nous disposons a une capacité de 100 Mbits/s dont uniquement 10 Mbits/s sont alloués pour la transmission. Nous rappelons que les vidéos acquises au bloc opératoire de Nancy proviennent d'une caméra tri-CCD et sont constituées de 720 pixels sur 576 lignes codés sur 10 bits, ce qui représente un débit total de 270 Mbits/s. Dans le domaine de l'informatique, la bande

passante indique, par abus de langage, un débit d'informations. Le terme exact est le **débit binaire**. Nous utiliserons le terme « bande passante réseau » pour désigner le débit binaire d'informations dans un réseau.

L'objectif principal ici est de transmettre ces flux en temps réel en limitant le temps de latence et en maintenant une qualité visuelle tolérable de l'image.

La figure 5.4 montre un schéma bloc de la plateforme mise en place. Sur le côté gauche de la figure, on voit la console du robot DaVinci et les connectiques associées. Quant au côté droit, il illustre la restitution dans la salle de conférence ainsi que les connectiques associées.

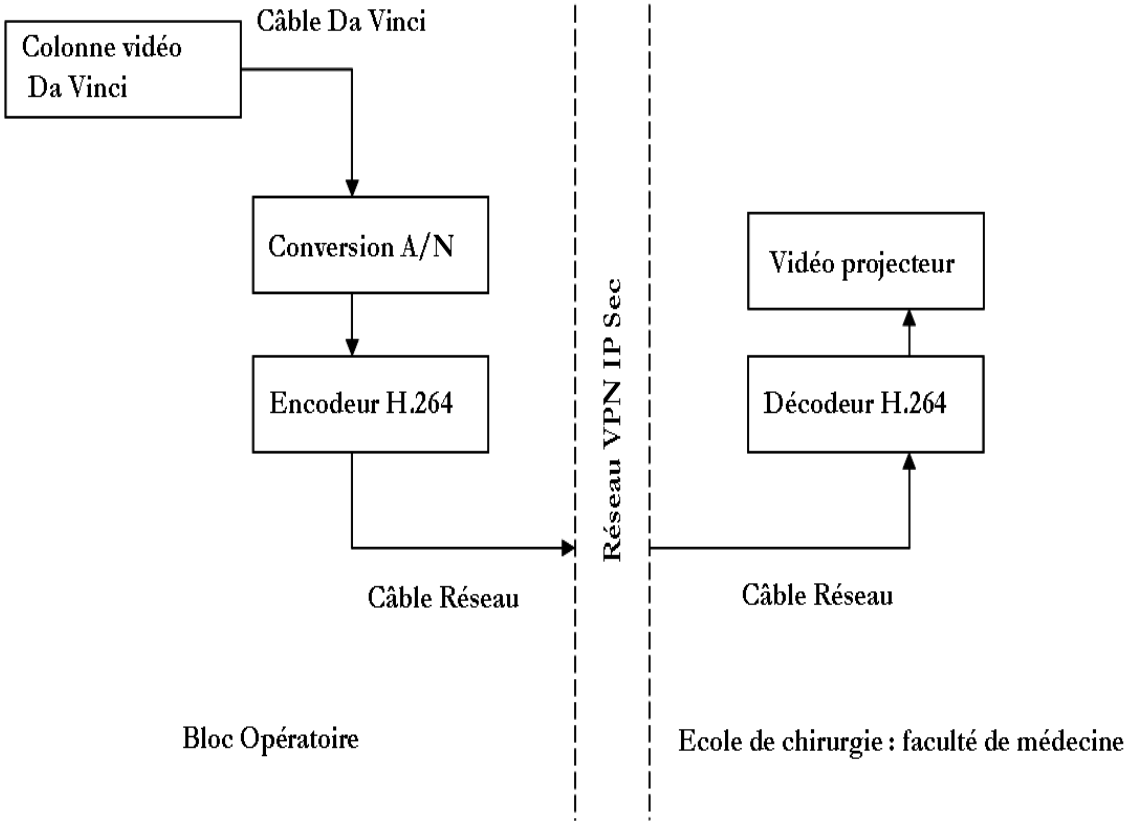


FIGURE 5.2. *Schéma de transmission CHU vers Ecole de chirurgie*

Les formats vidéos pour la transmission en flux continu (streaming) des données ciblent la contrainte de la bande passante réseau limitée et doivent permettre d'acheminer facilement les données. De plus, dans certaines configurations, ils doivent être facilement

lisibles par un lecteur vidéo léger (tel que VLC media Player ou RealPlayer). Historiquement, les standards DIVX (basé MPEG), MPEG-1 et MPEG-4 Part 2 ont longtemps été considérés comme des références. MPEG-4 part 2 est connu par sa capacité à incorporer des objets et des données. Généralement, pour le streaming, les débits supérieurs à 1150 kbits/s et les résolutions supérieures à 352x288 ne sont pas utilisées (cf. tableau des résolutions vidéos, tableau 5.1). Ces standards ont été intégrés dans des encodeurs en temps réel bas coût (caméras de surveillance), dans les cartes de compression et les caméras IP. Cependant, le temps de latence n'était pas un paramètre considérés dans leur conception. Ces formats ne sont pas adaptés à notre application notamment en raison des résolutions

Résolution	NTSC	PAL
QSIF	174x120	-
QCIF	-	174x144
SIF (~QVGA)	352x240	-
CIF	-	352x288
SD	720x480	720x576
HD 720p	1280x720	1280x720
HD 1080i	1920x1080	1920x1080

TABLE 5.1. *Résolutions vidéo*

très basses prises en compte, les faibles débits binaires et le temps de latence aléatoire engendré. Par ailleurs, dans le domaine de la télédiffusion (broadcast), les standards de streaming ne sont pas appropriés pour cette application car ils ne supportent pas toutes les résolutions (au minimum la résolution SD est requise dans le domaine du broadcast). C'est pour cela que MPEG-2 est utilisé à des fins de télédiffusion, ainsi que son extension HD. MPEG-2 est un format de compression vidéo prolifique (également utilisé pour les DVD, la télévision numérique), qui délivre de bonnes performances débits/distorsion. Cependant, historiquement, il a toujours requis une large bande passante réseau pour les applications de réseau étendu WAN (Wide area network). En dehors du domaine de télédiffusion, MPEG-2 est utilisé par des applications WAN telles que le télé-enseignement

ou les applications militaires. Dans le cas de notre application médicale (transmission de vidéos médicales sur un réseau), nous tenons compte de deux contraintes : la qualité visuelle après compression et le temps de latence (induit par l'algorithme de compression mais également par le réseau). L'étude de la qualité des vidéos (vue au chapitre 4) a porté sur les deux standards MPEG-2 et H.264. Le choix du standard H.264 est motivé par plusieurs raisons :

- MPEG-4 AVC (H.264) est un standard adapté aussi bien aux applications de télédiffusion que de transmission en flux continu. Il est donc compatible avec les applications Multimédia et par conséquent avec notre application médicale ;
- Ce standard présente une efficacité de codage qui nécessite moins de bande passante réseau et de capacités de stockage. Ceci est très important dans le cas de transmission sur réseau étendu ;
- C'est un standard qui ne dépend pas de la résolution des vidéos et donc adapté aussi bien à une résolution SD que HD.

Nous avons démontré l'efficacité de ce standard dans le cadre de notre application, à travers l'évaluation subjective de la qualité de la vidéo compressée. Nous avons également déterminé un **seuil de compression** maximal toléré par le panel de chirugiens autour de **2 Mbits/s**.

5.2.2 Spécifications techniques

La plateforme de compression et de transmission doit être intégrée à un réseau reliant le CHU de Nancy à l'Ecole de Chirurgie. Ce réseau d'une capacité totale de 100 Mbits/s est utilisé quotidiennement par diverses applications telles que les applications métier, la téléphonie IP ou les courriers électroniques. Le challenge à relever ici est de transmettre les flux vidéos issus du robots de chirurgie sur ce réseau tout en garantissant une qualité de service pour les autres applications.

Description du réseau La partie réseau de la plateforme de transmission (cf. Figure 5.4), est configurée de la manière suivante. Le lien entre le CHU de Nancy et la Faculté de Médecine se fait sur un tunnel VPN (Virtual Private Network) [Muthukfish-

nall et Malis, 2000] crypté IPSec (Internet Protocol Security) [Kent et Atkinson, 1998] avec un débit maximal de 100 Mbits/s (tous les composants). En effet, les données à transmettre sont encapsulées et de façon chiffrée. On parle alors de VPN (Figure 5.3) pour désigner le réseau ainsi artificiellement créé. Ce réseau est dit virtuel car il relie deux réseaux « physiques » (réseaux locaux) par une liaison non fiable (Internet), et privé car seuls les ordinateurs des réseaux locaux de part et d'autre du VPN peuvent accéder aux données en clair (ici pour des raisons évidentes de sécurité et d'intégrité des données médicales). Toute l'infrastructure, en dehors d'Internet, est répliquée.

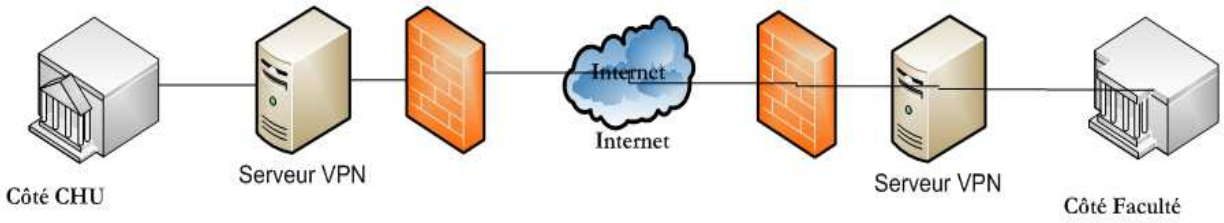


FIGURE 5.3. *Principe d'un VPN entre deux sites*

La bande passante allouée à la transmission vidéo est de 10Mbits/s pour garantir une qualité de service des autres applications transitant sur le même réseau. De nombreux routeurs sont traversés. Sur la figure 5.4, côté CHU, il y a deux groupes de 4 carrés qui sont des routeurs ainsi que le FireWall et les équipements qui construisent le tunnel VPN IPSec. La connexion entre les deux sites est une liaison point à point (Point-to-point Protocol). Lors de la transmission sur ce réseau, plusieurs protocoles sont utilisés : Ethernet, IP, TCP pour le management, UDP pour le flux vidéo ainsi que tous les protocoles pour le tunnel IPSec tels que Sha1 (Secure Hash Algorithm).

Le protocole de transmission du flux vidéo est RTP (Real-Time Transfer Protocol) au dessus de UDP (User Datagram Protocol) [Schulzrinne *et al.*, 2003]. Le but de RTP est de fournir un moyen uniforme de transmettre sur IP des données soumises à des contraintes de temps réel (vidéo, audio, etc.). Le rôle principal de RTP consiste à mettre en oeuvre des numéros de séquence de paquets IP pour reconstituer les informations de voix ou vidéo même si le réseau sous-jacent change l'ordre des paquets.

Compresssion/décompression Le système MAKO HD est un produit de la société HAIVISION[2]. En effet, il s'agit d'un encodeur/décodeur vidéo qui permet de réaliser la compression, la transmission et la décompression des vidéos en assurant un temps de latence assez court variant de 70 à 120 millisecondes et une excellente qualité d'image compressée. Ce système a été intégré à la plateforme du projet RALTT. La norme de compression utilisée est MPEG-4 AVC/H.264 (Main Profile). La vidéo initiale au format PAL de résolution SD est numérisée et est compressée en temps réel à un débit de **3 Mbits/s** (soit un taux de 90 :1). Ce débit est au dessus de la limite de compression tolérée par les chirurgiens est autour de 2 Mbits/s (chapitre 4) car les conditions de la transmission sont différentes des conditions de test : l'encodage H.264 est réalisé grâce au matériel MAKO HD et la restitution se fait sur un moniteur de résolution plus grande justifient cette marge de 1 Mbits/s.

L'intégration du standard H.264, utilisée dans notre application, garantit un temps de latence minimal (autour de 70 ms) grâce à l'encodage progressif des trames au fur et à mesure qu'elles arrivent au niveau de l'encodeur. Cette technique permet d'optimiser le temps d'encodage et de garantir la fluidité de la transmission.

Restitution Les images sont transmises sur le réseau entre le bloc opératoire et la salle de conférence de l'école de chirurgie. Les vidéos transmises sont restituées sur un vidéo-projecteur en résolution XGA (Extended Graphics Array) de résolution 1024x768 pixels. Il s'agit de la norme utilisée par les vidéoprojecteurs (connectique SUB-D Haute Densité 15 broches).

Contraintes Les contraintes sont situées principalement au niveau de la bande passante réseau qui est partagée avec d'autres applications. La transmission vidéo n'ayant aucune priorité sur les autres applications transitant par le même réseau, la question est de savoir si ceci a une incidence directe sur la qualité de service de bout en bout. Par ailleurs, les données médicales transmises en temps réel doivent transiter par un réseau sécurisé pour garantir leur intégrité.

2. http ://www.haivision.com/

5.3 Vidéo-transmission

La mise en place de la plateforme de compression et transmission de vidéos entre deux sites distants (CHU de Nancy et école de chirurgie) a permis de réaliser plusieurs transmissions en direct notamment lors de sessions de formation à la chirurgie robotique (Diplôme inter-universaitaire de chirurgie robotique) en 2010 et 2011. Nous avons relevé un retour positif de la part des participants aux vidéo-transmissions en particulier en ce qui concerne la qualité visuelle des vidéos compressées transmises. Une clé USB contenant un extrait vidéo d'une transmission en temps réel, est fournie avec ce manuscrit. Alors que, lors de congrès de chirurgie, des moyens importants sont mis en œvre pour assurer le même type d'opération (transmission via satellite), nous avons relevé le défi de transmettre des flux compressés sur un réseau partagé au quotidien par plusieurs applications, tout en assurant une qualité optimale.

5.4 Conclusion

Nous avons démontré, dans ce chapitre, la **faisabilité d'une transmission, sur un réseau IP, en temps-réel de flux vidéos** provenant d'une caméra endoscopique (champ opératoire) vers un site distant pour des objectifs pédagogiques (Figure 5.5). Les enjeux d'une telle transmission sont nombreux. D'abord, d'un point de vue pédagogique, il a été prouvé que la visualisation des opérations chirurgicales en en temps réel par les étudiants, permettait d'améliorer leurs compétences en raison de l'interaction avec le chirurgien et la possibilité d'anticiper les actions à effectuer.

Ensuite, cette première phase de validation, permet d'envisager à plus long terme des opérations chirurgicales à distance. La mise en œuvre d'une technique de synchronisation de flux stéréoscopiques adaptée au contexte (chirurgie, réseau, sécurité) sera la prochaine étape à valider.

Enfin, il est important de trouver un meilleur compromis entre des paramètres réseau optimaux (mesure de la bande passante, retards, taux de paquets perdus) et une compression efficace. Le problème de la transmission doit ainsi être traité parallèlement à la compression, d'autant plus qu'il existe de nombreuses perspectives de recherche et

de choix à faire concernant l'infrastructure réseau. Elles sont liées, en particulier, au trafic et à la congestion du réseau mais aussi aux problèmes de sécurité. Faut-il dupliquer toute l'infrastructure réseau de bout en bout pour prévenir les pannes ? Quels sont les paramètres de sécurité à prendre en compte ?

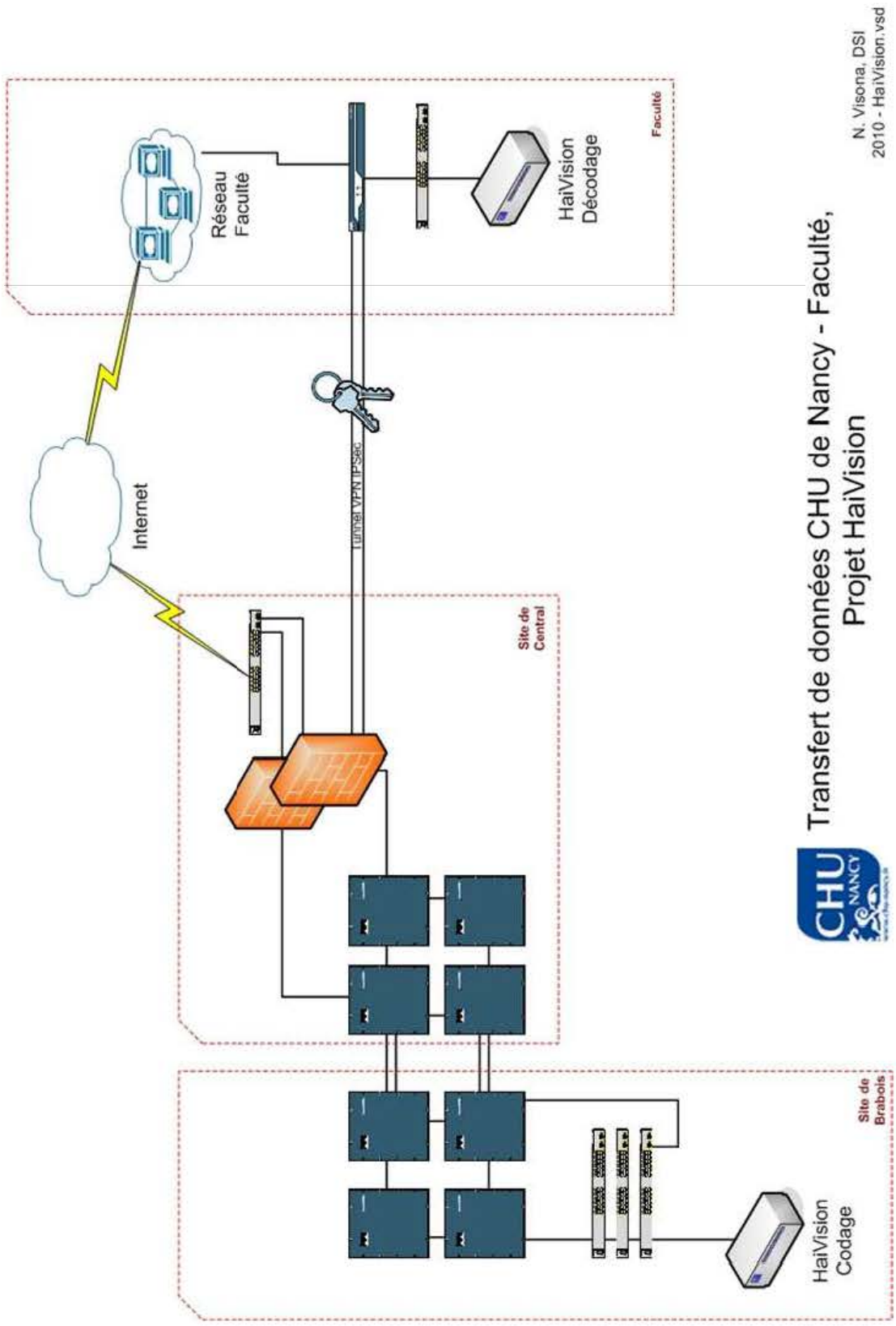


FIGURE 5.4. *Schéma de transmission*

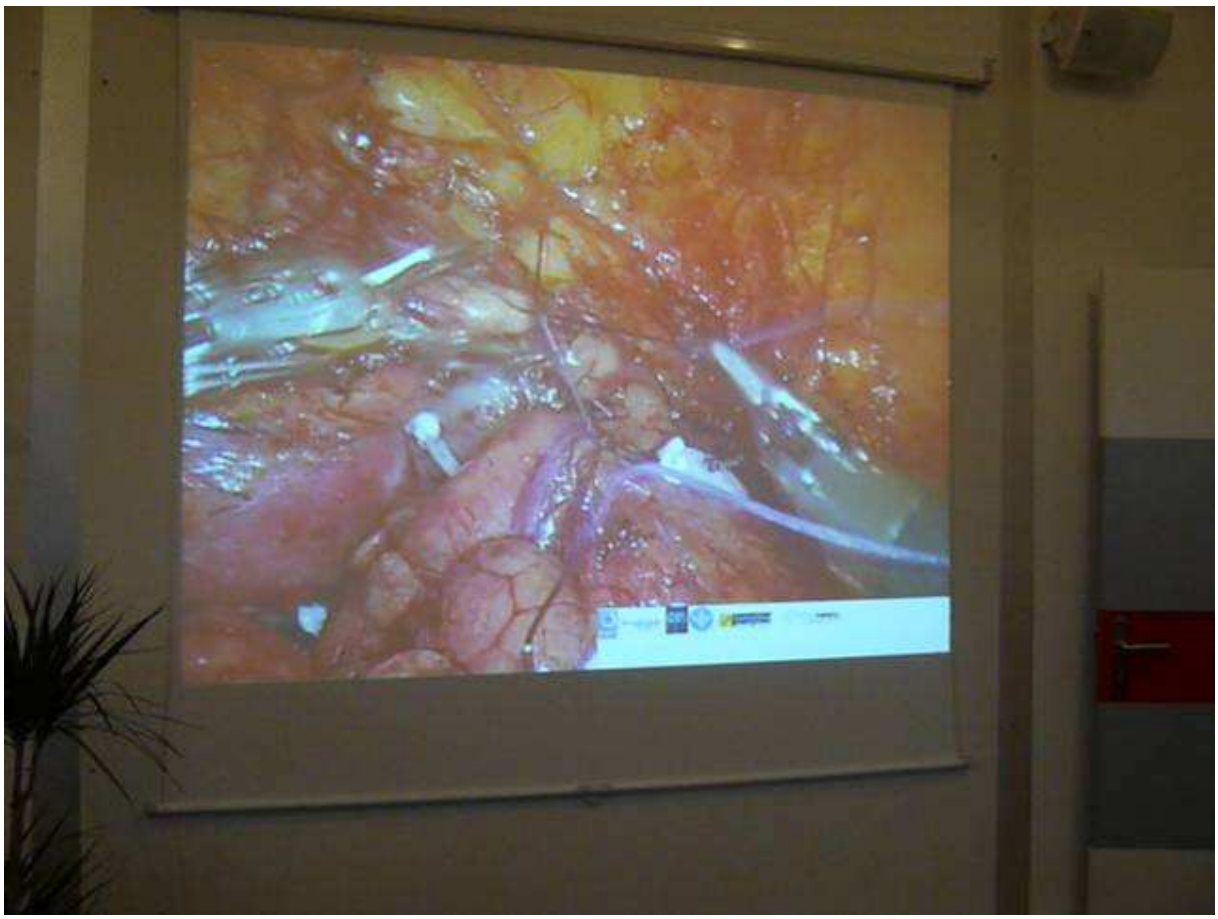


FIGURE 5.5. *Vidéo Transmission : côté Ecole de Chirurgie*

Conclusion générale et perspectives

Les contributions de ce travail touchent à la fois l'estimation de la qualité des vidéos issues d'un robot de chirurgie et la faisabilité de la tranmission de ces flux sur des réseaux IP en temps réel. Ces contributions tentent de répondre aux besoins du monde médical pour la chirurgie robotisée à distance et ouvrent de nombreuses perspectives dans ce contexte.

La première contribution de ce travail se situe sur le plan méthodologique et concerne la **mise en évidence de la possibilité de compresser des vidéos chirurgicales** afin de les **transmettre en temps réel sur des réseaux IP** sans perte sensible de qualité. La détermination d'un **seuil de compression** toléré par les chirurgiens, après des tests subjectifs de la qualité suivant une méthodologie d'essais normalisée, a permis d'une part de fixer un débit de compression pour une transmission en temps réel des vidéos chirurgicales et de mettre en évidence une corrélation entre les mesures subjectives effectuées et une mesure objective utilisant l'information structurelle de l'image (métrique SSIM), d'autre part. Ceci permet de prédire la qualité telle qu'elle est perçue par les observateurs humains.

La deuxième contribution de ce travail se situe sur le plan technique et est liée à la construction d'une plateforme de **transmission en temps réel** de vidéos chirurgicales entre le CHU de Nancy et l'école de chirurgie. Les flux transmis ont été compressés avec le standard H.264 à des taux de compression allant jusqu'à 90 :1. Cette plate-forme est opérationnelle et actuellement utilisée, à des fins d'enseignement, lors des sessions de

DIU (Diplôme inter-universitaire) de Chirurgie Robotique organisées à l'école de chirurgie de Nancy (formation des chirurgiens).

Ces résultats ouvrent de multiples perspectives parmi lesquelles : à court terme le télé-enseignement et le telementoring et à long terme la possibilité d'interventions chirurgicales à distance dans des conditions réalistes mais aussi d'autres perspectives relatives aux différents sujets abordés dans cette thèse.

D'abord, il serait intéressant d'utiliser une méthodologie suivant un protocole d'essais subjectifs permettant d'évaluer la perception des chirurgiens vis-à-vis d'autres contraintes liées à la chirurgie robotisée à distance : le temps de latence, la synchronisation des flux stéréoscopiques, le choix d'équipements de restitution. En effet, dans cette thèse, nous avons fait le choix de compresser une seule vue du flux stéréoscopique en nous basant sur les travaux de [Stelmach *et al.*, 2000] qui a démontré qu'en compressant fortement l'une des vues du flux stéréoscopique, la qualité de la vision 3D dépendait du flux le moins compressé. Il serait, cependant, utile d'étudier l'effet de la **compression conjointe des deux flux sétéréoscopiques** sur la reconstruction 3D.

L'utilisation des standards de compression MPEG-2 et H.264 a été motivée par la maturité de ces normes et leurs performances débit/distorsion. Cependant, une deuxième perspective de ce travail peut concerner la méthode de compression vidéo elle-même. Celle-ci peut prendre en compte la nature des vidéos chirurgicales (dominance de la couleur rouge, peu de mouvements dans une fenêtre d'observation réduite, l'importance de l'absence d'artefacts sur les embouts des instruments chirurgicaux). Par ailleurs, des travaux peuvent être envisagés sur des **méthodes de compression de vidéos stéréoscopiques dans le contexte chirurgical** afin d'exploiter les disparités entre les deux flux et trouver un compromis entre la compression et le temps de latence. Un tel choix doit forcément prendre en compte la problématique de la synchronisation des deux flux stéréoscopiques et évaluer **la tolérance des chirurgiens aux défauts de synchronisation**.

Quant aux défauts liés à la compression des vidéos, il serait opportun d'identifier les artefacts les plus présents dans les vidéos chirurgicales. En effet, le contenu de ces vidéos est connu a priori par le chirurgien, qui concentre son attention le plus souvent sur une zone précise du champ opératoire. De plus, le mouvement dans ce type de vidéos est

très faible notamment en raison de la fenêtre de visualisation assez réduite, il est très probable que les artefacts temporels dus à la compression n'induisent que peu de gêne dans la perception des chirurgiens.

Concernant la transmission des flux vidéo sur réseaux de télécommunications, il conviendrait d'étudier les **effets de la latence et de la perte de paquets sur la performance du chirurgien** afin de concevoir des solutions autorisant une qualité de service (débit constant) et une sécurité de haut niveau ainsi que respectant la confidentialité des données. Des recherches doivent porter sur l'identification des diverses perturbations inhérentes aux réseaux, qui risquent d'empêcher le bon déroulement des opérations chirurgicales à distance, de rendre difficile le travail du chirurgien.

Enfin, une étude portant conjointement sur les paramètres de compression, de transmission et de la qualité de service de bout en bout de la chaîne serait utile pour trouver un bon compromis taux de compression/latence.

Comme en témoignent ces nombreuses perspectives, nos travaux s'inscrivent dans un contexte pluridisciplinaire où la maîtrise de tous les paramètres pourrait aboutir à long terme à la pratique en routine de la chirurgie robotisée à distance dont pourront bénéficier les chirurgiens et les patients.

Annexe A

Consignes pour les essais subjectifs en chirurgie

Cher(e) participant(e),

Vous avez accepté de participer à une session d'essais subjectifs visant à déterminer la qualité des images de téléchirurgie, ce dont nous vous remercions vivement.

Nous allons vous présenter plusieurs scènes (extraits d'opérations), afin que vous jugiez de leur qualité technique.

La session d'essais subjectifs dure 60 minutes environ, pendant lesquelles vous allez noter environ une soixantaine de présentations d'images numériques. Les trois premières présentations sont destinées à vous entraîner.

A.1 Principe de l'essai

Chaque test (ou présentation) est constitué(e) par l'enchaînement décrit dans la figure B.1). Les écrans gris vous permettront de différencier chaque présentation ainsi que leurs différentes séquences (A, B). Afin de conforter votre jugement, chaque paire sera présentée à l'identique deux fois consécutives. Pour chaque paire, une séquence (A ou B) peut comporter des dégradations, alors que l'autre est toujours intacte. Vous allez donc visionner deux fois consécutives la même scène au contenu identique mais dont la qualité d'image pourra différer entre les deux séquences. Attention : la séquence non dégradée

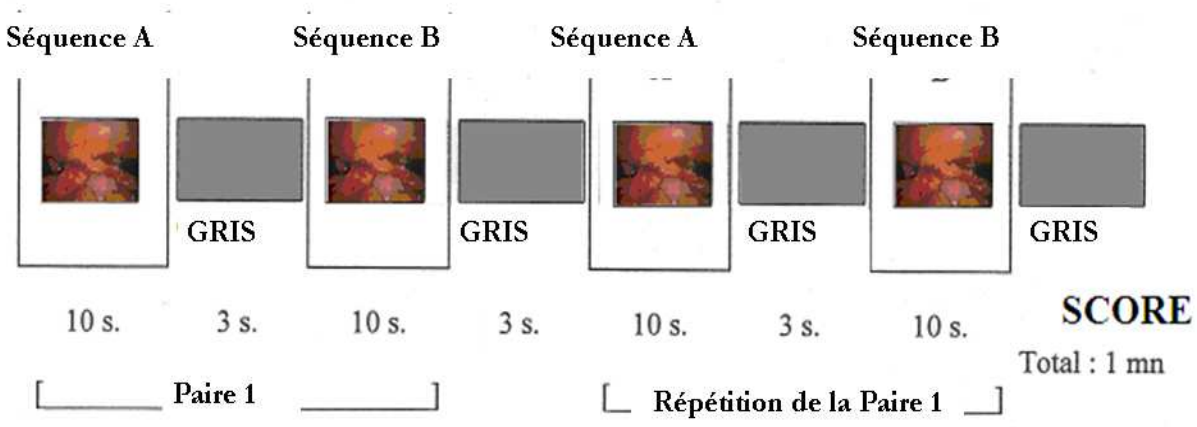


FIGURE A.1. *Enchaînement d'une présentation*

n'est pas toujours présentée en première position. L'ordre peut changer au cours des présentations.

A.2 Procédure de vote

Vous disposez d'une grille de notation pour chaque séquence. Chaque séquence présente deux échelles côte à côte (une pour la séquence A et l'autre pour la séquence B). Nous vous demandons de les utiliser pour noter la qualité des deux séquences A et B. Après la présentation de la première paire (séquences A et B), vous pouvez commencer à voter ou bien attendre la seconde présentation (11secondes vous seront disponibles à la fin des deux présentations). Pour voter, il vous suffit de marquer d'une croix la graduation correspondant au mieux à votre jugement de façon continue.

A.3 Quelques conseils

Votre jugement ne doit concerner que l'aspect technique de la qualité de l'image et faire abstraction de l'aspect artistique ou chirurgical du contenu des scènes. Vous êtes susceptibles de rencontrer les défauts, parmi lesquels :

- petites taches ;
- contours flous (dégradation des contours et de certaines structures de l'image) ;

– bruit (fourmillements, neige) ;

– défaut de restitution des mouvements.

Mais il peut aussi arriver que certains défauts soient imperceptibles ! Nous comptons sur vous car vos résultats serviront de base à une exploitation statistique.

Evitez de répondre au hasard. Sachez qu'il n'y a ni bonne, ni mauvaise réponse mais que seules vos appréciations subjectives et personnelles nous intéressent. Avant de commencer, avez-vous des questions à poser ?

Bibliographie

[Aksay *et al.*, 2007] AKSAY, A., PEHLIVAN, S., KURUTEPE, E., BILEN, C., OZCELEBI, T., AKAR, G., CIVANLAR, M. et TEKALP, A. (2007). End-to-end stereoscopic video streaming with content-adaptive rate and format control. *Signal Processing : Image Communication*, 22(2):157–168.

[Alpert et Evain, 1997] ALPERT, T. et EVAIN, J. (1997). Subjective quality evaluation : the sscqe and dscqe methodologies. *EBU Technical Review*, pages 12–20.

[Arnaud *et al.*, 2009] ARNAUD, J., NÉGRU, D., SIDIBÉ, M., PAUTY, J. et KOUMARAS, H. (2009). Adapted iptv service within novel ims architecture. *In Proceedings of the 5th International ICST Mobile Multimedia Communications Conference*, page 43. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[Bonnet, 1986] BONNET, C. (1986). Manuel pratique de psychophysique (paris : Armand colin).

[Bosi, 1997] BOSI, Marina ; Brandenburg, K. Q. S. F. L. A. K. F. H. D. M. (1997). Iso/iec mpeg-2 advanced audio coding. *J. Audio Eng. Soc*, 45(10):789–814.

[Butner et Ghodoussi, 2003] BUTNER, S. et GHODOUSSI, M. (2003). Transforming a surgical robot for human telesurgery. *Robotics and Automation, IEEE Transactions on*, 19(5):818–824.

[Cagnazzo *et al.*, 2007] CAGNAZZO, M., CASTALDO, F., ANDRÉ, T., ANTONINI, M. et BARLAUD, M. (2007). Optimal motion estimation for wavelet motion compensated video coding. *Circuits and Systems for Video Technology, IEEE Transactions on*, 17(7):907–911.

[Car, 2010] CAR (2010). Normes de la car en matière de compression irréversible pour l'imagerie numérique diagnostique en radiologie.

[Chen *et al.*, 2006a] CHEN, G., YANG, C., PO, L. et XIE, S. (2006a). Edge-based structural similarity for image quality assessment. *In Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 2, pages II–II. IEEE.

[Chen *et al.*, 2006b] CHEN, G., YANG, C. et XIE, S. (2006b). Gradient-based structural similarity for image quality assessment. *In Image Processing, 2006 IEEE International Conference on*, pages 2929–2932. IEEE.

[Corriveau *et al.*, 1999] CORRIVEAU, P., GOJMERAC, C., HUGHES, B. et STELMACH, L. (1999). All subjective scales are not created equal : The effects of context on different scales. *Signal processing*, 77(1):1–9.

[Crété-Roffet, 2007] CRÉTÉ-ROFFET, F. (2007). *Estimer, mesurer et corriger les artefacts de compression pour la télévision numérique*. Thèse de doctorat.

[Darazi *et al.*, 2009] DARAZI, R., GOUZE, A. et MACQ, B. (2009). Lifting scheme-based method for joint coding 3d stereo digital cinema with luminace correction and optimized prediction. *In Proceedings of SPIE*, volume 7257, page 72570J.

[Desurmont *et al.*, 2007] DESURMONT, X., BRUYELLE, J., RUIZ, D., MEESSEN, J. et MACQ, B. (2007). Real-time 3d video conference on generic hardware. *Real-Time Image Processing*.

[Engeldrum, 2000] ENGELDRUM, P. (2000). *Psychometric scaling : a toolkit for imaging systems development*. Imcotek Press, Winchester, Mass.

[Fabrlzio *et al.*, 2000] FABRLZIO, M., LEE, B., CHAN, D., STOIANOVICI, D., JARRETT, T., YANG, C. et KAVOUSSI, L. (2000). Effect of time delay on surgical performance during telesurgical manipulation. *Journal of endourology*, 14(2):133–138.

[Farias *et al.*, 2004] FARIAS, M., MOORE, M., FOLEY, J. et MITRA, S. (2004). Perceptual contributions of blocky, blurry, and fuzzy impairments to overall annoyance.

[Gaudeau et Moureaux, 2009] GAUDEAU, Y. et MOUREAUX, J. (2009). Lossy compression of volumetric medical images with 3D dead-zone lattice vector quantization. *Annals of telecommunications*, 64(5):359–367.

[Heinzelmann *et al.*, 1995] HEINZELMANN, M., SIMMEN, H., CUMMINS, A. et LARGIADER, F. (1995). Is laparoscopic appendectomy the new'gold standard'? *Archives of Surgery*, 130(7):782.

[hhi, 2010] HHI (2010). H.264/avc reference software.

[ISO, 2000a] ISO (2000a). Technologies de l'information - codage générique des images animées et du son associé : Données vidéo.

[ISO, 2000b] ISO (2000b). Technologies de l'information - codage générique des images animées et du son associé : Systèmes.

[ITU-R, 2000] ITU-R (2000). Recommendation 500-10 ; Methodology for the subjective assessment of the quality of television pictures. *ITU-R Rec. BT. 500*, 10.

[Kent et Atkinson, 1998] KENT, S. et ATKINSON, R. (1998). Rfc 2401 : Security architecture for the internet protocol, nov. 1998. *Status : Proposed Standard*.

[Marescaux *et al.*, 2002] MARESCAUX, J., LEROY, J., RUBINO, F., SMITH, M., VIX, M., SIMONE, M. et MUTTER, D. (2002). Transcontinental robot-assisted remote telesurgery : feasibility and potential applications. *Annals of surgery*, 235(4):487.

[Mills, 1992] MILLS, D. (1992). Network Time Protocol (Version 3) specification, implementation and analysis. *Network*, 1305.

[Muthukfishnall et Malis, 2000] MUTHUKFISHNALL, K. et MALIS, A. (2000). core mpis ip vpn architecture.

[Ninassi, 2009] NINASSI, A. (2009). *De la perception locale des distorsions de codage à l'appréciation globale de la qualité visuelle des images et des vidéos. Apport de l'attention visuelle dans le jugement de la qualité*. These, Université de Nantes.

[Norenzayan *et al.*, 2002] NORENZAYAN, A., SMITH, E., KIM, B. et NISBETT, R. (2002). Cultural preferences for formal versus intuitive reasoning. *Cognitive Science*, 26(5):653–684.

[Nouri *et al.*, 2010] NOURI, N., ABRAHAM, D., MOUREAUX, J., DUFAUT, M., HUBERT, J. et PEREZ, M. (2010). Subjective MPEG2 compressed video quality assessment : Application to Tele-surgery. *In Biomedical Imaging : From Nano to Macro, 2010 IEEE International Symposium on*, pages 764–767. IEEE.

[NTIA, 2011] NTIA (2011). Ntia general video quality metric (vqm) software.

[Oh *et al.*, 2004] OH, S., LEE, Y. et WOO, W. (2004). Scalable stereo video coding for heterogeneous environments. *Interactive Multimedia and Next Generation Networks*, pages 72–83.

[P.910, 1999] P.910, I.-T. R. (1999). Subjective video quality assessment methods for multimedia applications itu-t p.910.

[Puri *et al.*, 1997] PURI, A., KOLLARITS, R. et HASKELL, B. (1997). Basics of stereoscopic video, new compression results with MPEG-2 and a proposal for MPEG-4. *Signal Processing : Image Communication*, 10(1-3):201–234.

[Richardson, 2008] RICHARDSON, I. (2008). *H. 264 and MPEG-4 video compression*. Wiley Online Library.

[Schelkens *et al.*, 2003] SCHELKENS, P., MUNTEANU, A., BARBARIEN, J., GALCA, M., GIRO-NIETO, X. et CORNELIS, J. (2003). Wavelet coding of volumetric medical datasets. *Medical Imaging, IEEE Transactions on*, 22(3):441–458.

[Schulzrinne *et al.*, 2003] SCHULZRINNE, H., CASNER, S., FREDERICK, R. et JACOBSON, V. (2003). Rtp : A transport protocol for real-time applications (rfc 3550). *Internet Engineering Task Force*.

[Schulzrinne *et al.*, 1996] SCHULZRINNE, H., CASNER, S., FREDERICK, R., JACOBSON, V. *et al.* (1996). RTP : A transport protocol for real-time applications.

[Soper *et al.*, 1992] SOPER, N., STOCKMANN, P., DUNNEGAN, D. et ASHLEY, S. (1992). Laparoscopic Cholecystectomy The New'Gold Standard' ? *Archives of surgery*, 127(8):917.

[Stelmach *et al.*, 2000] STELMACH, L., TAM, W., MEEGAN, D. et VINCENT, A. (2000). Stereo image quality : effects of mixed spatio-temporal resolution. *Circuits and Systems for Video Technology, IEEE Transactions on*, 10(2):188–193.

[Stevens, 1957] STEVENS, S. (1957). On the psychophysical law. *Psychological Review*, 64(3):153.

[Theobalt *et al.*, 2007] THEOBALT, C., AHMED, N., ZIEGLER, G. et SEIDEL, H. (2007). High-quality reconstruction from multiview video streams. *Signal Processing Magazine, IEEE*, 24(6):45–57.

[VideoLAN, 2010] VIDEOLAN (2010). x264 free software library.

[VQEG, 2008] VQEG (2008). Final report from the video quality experts group on the validation of objective models of multimedia quality assessment, phase i. Rapport technique, VQEG.

[Vuori *et al.*, 2004] VUORI, T., OLKKONEN, M., P

"OL

"ONEN, M., SIREN, A. et H

"AKKINEN, J. (2004). Can eye movements be quantitatively applied to image quality studies ? *In Proceedings of the third Nordic conference on Human-computer interaction*, pages 335–338. ACM.

[Wang et Bovik, 2002] WANG, Z. et BOVIK, A. (2002). A universal image quality index. *Signal Processing Letters, IEEE*, 9(3):81–84.

[Wang *et al.*, 2004a] WANG, Z., BOVIK, A., SHEIKH, H. et SIMONCELLI, E. (2004a). Image quality assessment : From error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612.

[Wang *et al.*, 2004b] WANG, Z., LU, L. et BOVIK, A. (2004b). Video quality assessment based on structural distortion measurement. *Signal processing : Image communication*, 19(2):121–132.

[Watkinson, 2006] WATKINSON, J. (2006). *The MPEG handbook : MPEG-1, MPEG-2, MPEG-4*. Focal Press.

[Wharton et Howorth, 1967] WHARTON, W. et HOWORTH, D. (1967). *Principles of television reception*. Pitman.

[Winkler, 1999] WINKLER, S. (1999). A perceptual distortion metric for digital color images. *In Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*, pages 399–403. IEEE.

[Wolf et Pinson, 2002] WOLF, S. et PINSON, M. (2002). Video quality measurement techniques. *2002.*

[Yuen et Wu, 1998] YUEN, M. et WU, H. (1998). A survey of hybrid mc/dpcm/dct video coding distortions. *Signal Processing*, 70(3):247–278.

[Zhang *et al.*, 2006] ZHANG, B., SUN, L. et CHENG, X. (2006). Video QoS Monitoring and Control Framework over Mobile and IP Networks. *Advances in Multimedia Information Processing-PCM 2006*, pages 714–721.

Annexe B

Autorisation de Soutenance

Nancy-Université
INPL

AUTORISATION DE SOUTENANCE DE THESE
DU DOCTORAT DE L'INSTITUT NATIONAL
POLYTECHNIQUE DE LORRAINE

o0o

VU LES RAPPORTS ETABLIS PAR :
Monsieur Benoît MACQ, Professeur, Université Catholique de Louvain
Monsieur Marc ANTONINI, Directeur de Recherche, Université Sophia Antipolis

Le Président de l'Institut National Polytechnique de Lorraine, autorise :

Madame NOURI Nedia

à soutenir devant un jury de l'INSTITUT NATIONAL POLYTECHNIQUE DE LORRAINE, une thèse confidentielle à huis clos :

en vue de l'obtention du titre de :

DOCTEUR DE L'INSTITUT NATIONAL POLYTECHNIQUE DE LORRAINE

Spécialité : « **Automatique, Traitement du Signal et des Images, Génie Informatique** »

Fait à Vandoeuvre, le 05 septembre 2011
Le Président de l'I.N.P.L.,
F. LAURENT

Institut National Polytechnique de Lorraine
inpl@inpl-nancy.fr - www.inpl-nancy.fr

2 av. de la Forêt de Haye - BP 3
54501 Vandœuvre-lès-Nancy Cedex
FRANCE
Tél. : +33(0)3 83 59 59 59
Fax : +33(0)3 83 59 59 55

Résumé

L'évolution des techniques chirurgicales, par l'utilisation de robots, permet des interventions mini-invasives avec une très grande précision et ouvre des perspectives d'interventions chirurgicales à distance, comme l'a démontré la célèbre expérimentation «Opération Lindbergh» en 2001. La contrepartie de cette évolution réside dans des volumes de données considérables qui nécessitent des ressources importantes pour leur transmission. La compression avec pertes de ces données devient donc inévitable. Celle-ci constitue un défi majeur dans le contexte médical, celui de l'impact des pertes sur la qualité des données et leur exploitation. Mes travaux de thèse concernent l'étude de techniques permettant l'évaluation de la qualité des vidéos dans un contexte de robotique chirurgicale. Deux approches méthodologiques sont possibles : l'une à caractère subjectif et l'autre à caractère objectif. Nous montrons qu'il existe un seuil de tolérance à la compression avec pertes de type MPEG2 et H.264 pour les vidéos chirurgicales. Les résultats obtenus suite aux essais subjectifs de la qualité ont permis également de mettre en exergue une corrélation entre les mesures subjectives effectuées et une mesure objective utilisant l'information structurelle de l'image. Ceci permet de prédire la qualité telle qu'elle est perçue par les observateurs humains. Enfin, la détermination d'un seuil de tolérance à la compression avec pertes a permis la mise en place d'une plateforme de transmission en temps réel sur un réseau IP de vidéos chirurgicales compressées avec le standard H.264 entre le CHU de Nancy et l'école de chirurgie.

Mots-clés: compression vidéo, chirurgie robotisée à distance, qualité objective et subjective, transmission

Abstract

The digital revolution in medical environment speeds up development of remote Robotic-Assisted Surgery and consequently the transmission of medical numerical data such as pictures or videos becomes possible. However, medical video transmission requires significant bandwidth and high compression ratios, only accessible with lossy compression. Therefore research effort has been focussed on video compression algorithms such as MPEG2 and H.264. In this work, we are interested in the question of compression thresholds and associated bitrates are coherent with the acceptance level of the quality in the field of medical video. To evaluate compressed medical video quality, we performed a subjective assessment test with a panel of human observers using a DSCQS (Double-Stimuli Continuous Quality Scale) protocol derived from the ITU-R BT-500-11 recommendations. Promising results estimate that 3 Mbits/s could be sufficient (compression ratio aroundthreshold compression level around 90 :1 compared to the original 270 Mbits/s) as far as perceived quality is concerned. Otherwise, determining a tolerance to lossy compression has allowed implementation of a platform for real-time transmission over an IP network for surgical videos compressed with the H.264 standard from the University Hospital of Nancy and the school of surgery.

Keywords: Video, Encoding, Robotic-Assisted Surgery, Quality assessment,Transmission