

# La visio-conférence holographique : Pourquoi ? Comment ?

## Francis Felix

Labo LSIS / Arts & Métiers Paritech (ENSAM)  
2 Cours des Arts et Métiers  
13100 Aix-en-Provence

## Thierry Henocque

AIP-Primeca Dauphiné Savoie / Grenoble INP  
740 rue de la piscine  
38100 Saint Martin d'hères

## Résumé

*La visio-conférence holographique s'appuie sur des technologies permettant de capter, transmettre et restituer en 3D un environnement, un objet ou des personnes.*

*La projection holographique permet de distribuer, à chacun, une scène en 3D, relative à sa position, devant, ou mieux, autour de l'espace projeté. Elle ajoute la perception de la direction du regard, du langage corporel et des gestes destinés à des cibles distantes. Estompant les difficultés d'une visio-conférence traditionnelle, en 2D (fond clair, fuite du regard, gestes limités).*

*L'évolution des briques technologiques (captation, transmission, restitution) permet d'innover en terme de communication. Des réseaux haut-débits, des systèmes multi-cameras, la photométrie, photogrammétrie, l'interopérabilité des différentes couches logicielles, la compression des données et les protocoles utilisés, définissent les qualités, usages et procédés nécessaires à la visio-conférence holographique.*

*Les éléments constitutifs de telles solutions, sont encore difficiles à maîtriser. La captation d'une scène implique, pour être restituée, de nombreux instruments qui enregistrent différents niveaux d'informations. Leur transmission induit un encodage et des encapsulations de données dans les flux transmis, si possible, à haut débit ou sur des couches connexes aux flux vidéos. Il sera ainsi possible de restituer images, sons, données 3D, comme on manipule des sous-titres dans les flux vidéos traditionnels. La restitution de tels environnements, pour le spectateur distant, lui fera percevoir une réalité induite par ses nouveaux niveaux d'informations disponibles, sur la base de procédés de projection comparés, utilisés et étudiés.*

*L'état de l'art sur ces technologies, la nature innovante de ces solutions, et l'interopérabilité des systèmes sont l'objet de notre étude afin d'envisager les meilleures solutions en présence, y compris au travers des travaux de communautés dynamiques sur les différentes articulations de tels systèmes.*

## Mots-clefs

*visioconférence, holographie, vision stéréoscopique, voxel, immersion, reconnaissance d'objets, capture de mouvements, perception spatiale, modélisation tridimensionnelle, OpenGL, OpenCV, OpenNI.*

## 1 Introduction

La qualité d'une visioconférence a un impact très fort sur la qualité des informations échangées lors de la réunion.

Il est possible de considérer trois étapes fondamentales qui peuvent toutes avoir un impact fort sur le résultat obtenu qui sont dans l'ordre la captation, la transmission et la restitution. Les défauts liés à chaque étape ne peuvent faire que s'aggraver en fonction des problèmes rencontrés lors des étapes suivantes.

La qualité de la captation est primordiale et dépend très fortement d'un grand nombre de paramètres liés à l'environnement. Parmi les plus importants, on peut mentionner la qualité du son, la qualité des images captées, et le comportement des participants vis-à-vis de la caméra.

Le réseau utilisé pour le transport et la transmission entre les sites concernés doit être correctement dimensionné et la qualité de service correctement configurée.

La qualité de la restitution et en particulier l'impact de la sensation d'immersion [1] des participants dans la scène distante a un impact très fort sur la fatigue, les capacités de concentration des participants et la richesse des informations échangées lors de la réunion.

En guise de réponse à la question « pourquoi ? », nous allons étudier l'impact de l'usage d'un système holographique sur chacune de ces étapes et à la question « comment ? », nous répondrons en présentant quelques pistes techniques pour la réalisation d'un tel dispositif.

## 2 Impacts d'un système holographie sur la visioconférence

### 2.1 Définition étendue de l'holographie

L'holographie est un procédé photographique qui consiste à enregistrer sur une plaque les franges d'interférence entre une source cohérente de lumière, produite par un laser, et la réflexion de cette même source de lumière sur un objet (Figure 1). Dénes Gábor a découvert en 1948 que si cette plaque (l'hologramme) est à nouveau éclairée par la même source de lumière, la diffraction de la lumière sur les lignes d'interférence reconstitue visuellement et virtuellement une image tridimensionnelle de l'objet (Figure 2).

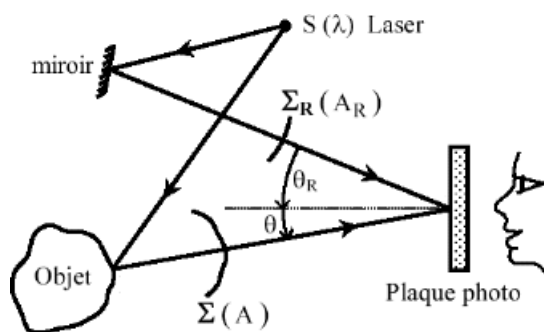


Figure 1 - Enregistrement d'un hologramme

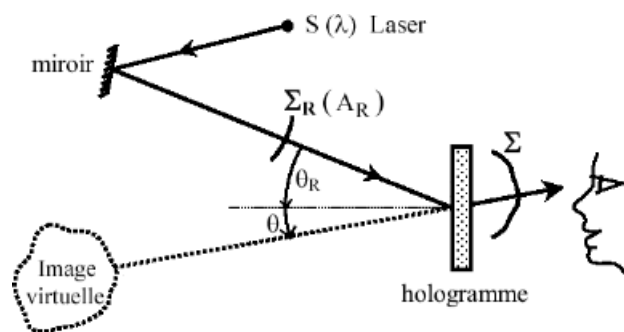
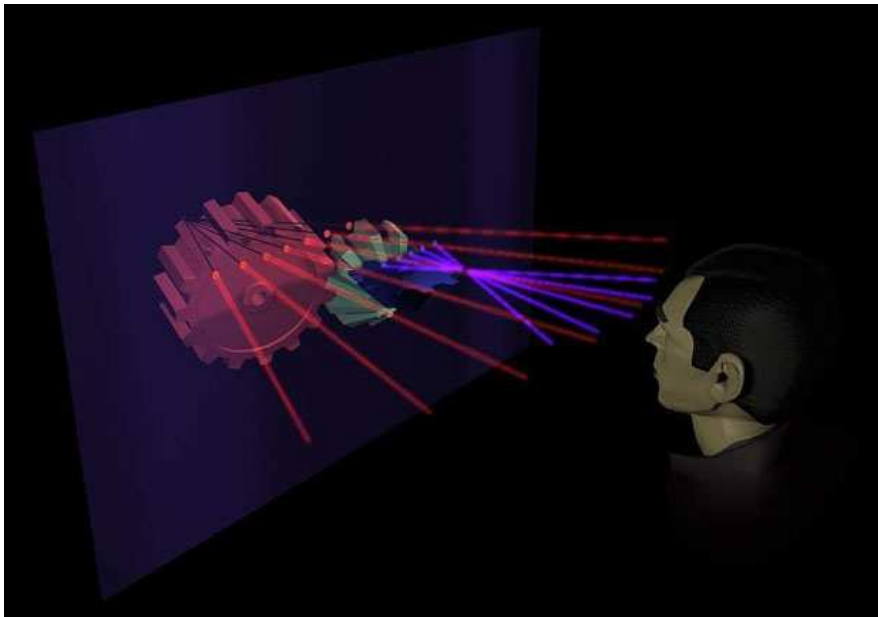


Figure 2 - Lecture d'un hologramme

Par extension, nous appelons « système holographique » tout mécanisme optique qui permet de générer l'image tridimensionnelle d'un objet qui n'est réellement présent, de l'observer sans accessoires supplémentaires (sans lunettes par exemple) et de pouvoir changer de point de vue en tournant autour de l'objet. C'est le cas en particulier du système Hologvizio (brevet codétenu Holografica Sony) qui par le jeu d'un grand nombre de vidéoprojecteur placés derrière une dalle diffractante (figure 3) recalcule les images nécessaires pour que la scène soit vue comme si les objets étaient à leur position dans celle-ci.

Cette technologie se distingue des autres technologies actuellement disponibles par le fait qu'elle projette en couleur toute scène ou animation au format Open-GL et qu'elle délivre un nombre de points de vue de la scène tel qu'on a l'impression d'une continuité lors du déplacement autour de celle-ci.



*Figure 3 - Principe de fonctionnement de l'écran Hologvizio*

## 2.2 L'immersion totale

Qui dit visioconférence holographique dit forcément restitution holographique de l'environnement et des personnes distantes. C'est donc uniquement dans la phase de restitution et dans l'immersion que cette technologie procure que le bénéfice d'une telle architecture s'exprime.

La vision en relief apporte plusieurs choses importantes.

D'abord, il n'existe plus de problèmes de distinction entre le fond de la scène et les participants puisqu'ils se trouvent spatialement séparés et les quelques degrés d'ombrages et de luminosité qui les séparent forcément font que notre cerveau fera parfaitement la différence entre eux, même s'ils sont tous de la même couleur. D'ailleurs, puisqu'on parle d'habillement, peu importe qu'un des participants porte une chemise à carreaux, qui attire trop l'attention dans une visioconférence traditionnelle, puisque le cerveau n'a plus besoin de ce critère pour essayer de repérer des perspectives qui lui permettrait de reconstituer mentalement la volumétrie de la scène qu'il observe.

Ensuite, puisqu'il y a des repères spatiaux, l'observateur a une perception très précise de la direction des mouvements du corps. Il est par exemple possible de pointer du doigt un objet et que chaque participant ait une perception précise de l'objet pointé, et ce, quel que soit sa position (locale ou distante) par rapport à l'objet et par rapport à la personne qui fait le geste.

Cet aspect fondamental de perception vectorielle des mouvements s'applique aussi aux mouvements de tête et pour la direction du regard.

En clair, on se pose dans un environnement d'immersion total qui fait qu'il n'y a plus, comme dans la vraie vie, de contraintes d'environnement et d'habillement, et qu'il n'y a plus non plus de problèmes de comportement liés à ce que l'on appelle la fuite du regard et à l'interprétation des mouvements des participants.

## 2.3 Volumétrie des données nécessaire

Le volume de données à transmettre est un point crucial dans un système de visioconférence. Une vue naïve pourrait être de considérer que l'on a à transmettre des voxels<sup>1</sup> qui, pour un équivalent 2D de 1920x1080 pourrait représenter quelque chose comme 1920x1080x1080 soit environ  $2.10^9$  voxels. Une autre approche peut consister à considérer qu'il faut environ 20 points de vue, avec des techniques de photogrammétrie, pour reconstituer par corrélation et extrapolation une scène en 3 dimensions. Dans ce cas, la volumétrie serait simplement 20 fois plus importante qu'une visioconférence traditionnelle.

<sup>1</sup> Voxel : contraction de volumetric pixel, est un pixel en 3D

Une troisième approche consisterait à transmettre, une scène avec des objets mobiles, en OpenGL<sup>2</sup> par exemple. Cette technique offre plusieurs avantages. D'une part la restitution d'une scène décrite sous la forme d'objets tridimensionnels peut être restituée facilement avec n'importe quel point de vue et sans utiliser plus de ressources que celles de la carte vidéo du périphérique utilisé pour la restitution.

L'inconvénient de cette technique est que la volumétrie de la scène dépend de sa complexité, donc difficile à évaluer à priori. On sait néanmoins que si on a à décrire une forêt, il n'est pas nécessaire d'intégrer en tant qu'objets tous les arbres de celle-ci car de toute façon, au-delà de cinq à six mètres, notre cerveau n'utilise plus que les lignes de fuites et les perspectives pour interpréter les volumes. Il suffit donc d'un fond et de 2 ou 3 arbres décrit en tant qu'objets pour que le rendu soit correct.

Cette technique devrait donc être globalement rentable du point de vue de la volumétrie et du point de vue de la rapidité de restitution.

Il reste bien sûr dans ce cas à reconnaître et à modéliser à la volée les objets de la scène.

### 3 Solutions techniques envisageable pour la captation

On sait qu'il est particulièrement difficile de reconnaître et de modéliser par les techniques de reconnaissance de forme avec une webcam un objet nouveau sur lequel on n'a aucune information. Par contre, il est assez facile d'identifier des objets ou des éléments d'un objet que l'on connaît bien et pour lequel on a beaucoup d'informations. Par exemple, avec les bibliothèques de traitement d'image d'OpenCV<sup>3</sup>, il est aisé de reconnaître des lignes et de reconstruire des perspectives, de suivre le mouvement des yeux d'un visage, de reconnaître si une personne est triste ou gaie, de suivre un objet en mouvement qui possède une forme ou une couleur caractéristique etc.

Par ailleurs, l'évolution des techniques de suivi de personnes a énormément progressé avec l'apparition de la Kinect, le périphérique de la console de jeu Xbox360. Ce périphérique a rapidement été détourné de sa console de jeu pour être utilisé à partir d'un ordinateur avec l'apparition d'un environnement de développement open-source (OpenNI<sup>4</sup>) suivi rapidement de la mise à disposition par l'éditeur Microsoft d'un environnement de développement libre (Kinect SDK).

En conséquence de ce détournement d'usage, Microsoft a sorti un modèle spécifique pour Windows avec des caractéristiques techniques améliorées, en particulier une meilleure définition, par rapport à la version de base prévue pour console de jeu.

Ce système, comme les systèmes de « tracking » (suivi de mouvements) professionnels, utilise l'émission de motif infra-rouge et l'analyse de la réflexion de ceux-ci sur l'environnement, pour évaluer les formes et les distances entre le périphérique et les objets qui constituent la scène. Il utilise ensuite la reconnaissance de point particuliers de l'anatomie pour identifier la position du corps et les expressions du visage. Il utilise 17 points (articulations) caractéristiques du squelette pour identifier la position et les mouvements du corps et, 85 points caractéristiques du visage, pour reconnaître l'individu et ses expressions. Cela signifie que transmettre les mouvements et expressions des personnes impliqués dans une visioconférence pourrait se limiter à la transmission d'une centaine de positions par personne dans l'espace.

### 4 Conclusion

Les systèmes de reconnaissance par reconnaissance d'image et par suivi de mouvement évoluent très vite et sont d'ores et déjà suffisamment mature pour envisager de capturer et de modéliser sous la forme d'objets tridimensionnels en quasi temps réel un ensemble de personnes et leur environnement dans le cadre d'une visioconférence. Le volume de données à transférer serait très réduit si on se contente de transférer d'une part les formes et texture et ensuite uniquement les mouvements et expressions des personnes impliquées dans la visioconférence.

Les solutions de restitution holographique existent, même si elles sont encore très limitées mais ne devrait pas tarder à se développer aux vues de l'intérêt qu'elles apportent en terme de communication. La question subsidiaire « quand ? » trouve ici sa réponse : Lorsque les technologies de visualisation holographique seront démocratisées.

---

<sup>2</sup> OpenGL : Open Graphics Library

<sup>3</sup> OpenCV : Open Computer Vision

<sup>4</sup> OpenNI : Open Natural Interaction

