

# Télévision à point de vue libre et système de capture à plusieurs caméra omnidirectionnelles

Thomas MAUGEY<sup>1</sup>, Cédric LE CAM<sup>1</sup>, Laurent GUILLO<sup>2</sup>

<sup>1</sup>Inria Rennes Bretagne Atlantique  
Campus Beaulieu, 35042 Rennes, France

<sup>2</sup>CNRS, IRISA UMR 6074  
Campus Beaulieu, 35042 Rennes, France

thomas.maugey@inria.fr, cedric.le-cam@inria.fr  
laurent.guillo@irisa.fr

**Résumé** – La télévision à point de vue libre a pour objectif de permettre à l'utilisateur qui observe un contenu multimédia de changer son angle de vue à tout instant, en temps réel. Cette application rencontre des verrous scientifiques importants. Le tout premier concerne la capture de la scène. Celle-ci doit permettre d'enregistrer suffisamment de contenu d'une scène pour permettre une restitution de bonne qualité du côté de l'utilisateur. Dans le cadre du projet ATeP porté par l'Inria, nous nous intéressons à une nouvelle approche de capture utilisant un système à plusieurs caméras omnidirectionnelles (à la place des traditionnelles caméras perspectives). Dans cet article, nous expliquons pourquoi un tel système peut permettre une meilleure couverture de l'espace. Nous présentons les verrous qu'un tel système de capture peut rencontrer, et nous proposerons une manière de résoudre le principal d'entre eux, à savoir la calibration du système. Les résultats montrent que le "modèle sphérique unifié" permet de modéliser correctement la projection de point 3D sur les images omnidirectionnelles.

**Abstract** – Free viewpoint television aims at enabling users who watch a multimedia content to change their angle of view in real time whenever they want. This application raises several major scientific issues: on top of them, the scene capture. It has the task to register a sufficient amount of the scene content to make a good restitution to the users possible. In the context of the Inria project ATeP, we are interested in a new acquisition approach using a system with multiple omnidirectional cameras (instead of traditional perspective cameras). In this paper, we explain why such a system can lead to a more complete acquisition of the scene. We present all the issues raised by such a capture system, and we propose to solve the most important of them: the calibration. Results show that the "unified spherical model" leads to a good modeling of the projection of a 3D point onto the omnidirectional images.

## 1 Introduction

La Télévision à Point de Vue Libre (TPVL) est une application émergente dans laquelle un contenu multimédia est transmis aux utilisateurs de manière à ce que chacun puisse choisir et changer son angle de vue en temps réel [1]. Autrement dit, l'utilisateur a la possibilité d'observer une scène du point de vue qu'il souhaite : soit en choisissant parmi un ensemble de vues présélectionnées (Fig.1(a)), soit en naviguant librement où bon lui semble, *i.e.*, selon un angle de vue différent de celui des caméras d'acquisition (Fig.1(b)). Dans ce dernier système, le rendu peut se faire par un affichage (2D ou stéréoscopique) classique, ou dans le cadre de système de réalité virtuelle.

Les applications cibles de la TPVL sont nombreuses : transmission d'événement sportifs, d'événement culturels, mais également dans le domaine industriel (*e.g.*, «usine du futur»), de l'éducation et de la santé [1]. Aujourd'hui, l'émergence de systèmes de TPVL est limitée par un certain nombre de verrous portant sur l'ensemble de la chaîne de transmission : capture, représentation, codage, transmission, décodage et rendu [1]. Dans cet article, nous nous concentrerons sur les problé-

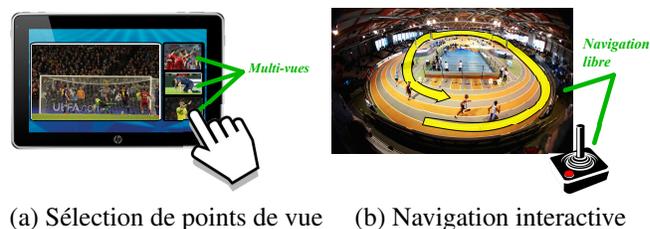


FIGURE 1 – Deux approches pour la TVPL.

matiques de capture. L'acquisition de scènes pour la TPVL présente de nombreuses problématiques liées à leur grande taille. En effet, pour permettre une navigation de bonne qualité, il faut que l'utilisateur ait accès à une navigation parmi un ensemble de point de vues suffisamment important pour couvrir la totalité (ou une grande partie) de la scène. Cette contrainte rend très coûteuse une telle acquisition et pose des questions fondamentales telles que : combien de caméras faut-il utiliser ? Où doit-on les placer ?

Dans les systèmes actuels, l'acquisition est généralement faite en utilisant deux types de capteurs : des caméras vidéos

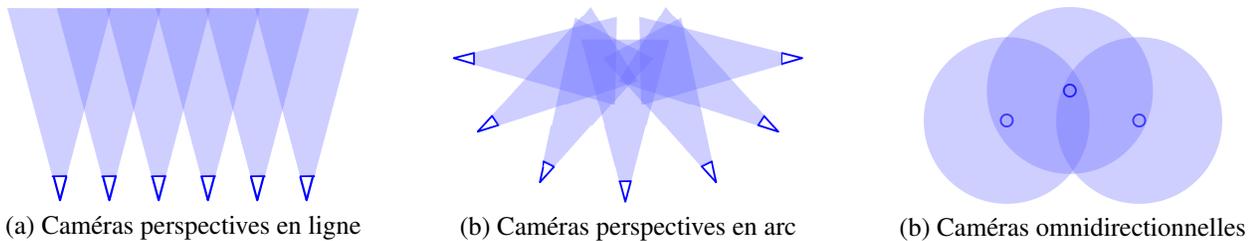


FIGURE 2 – Capture d’une scène avec différents types de caméra. Les zones colorées correspondent aux points de la scène visibles depuis les caméras avec une résolution acceptable.

"classiques" et parfois des caméras de profondeur pour enregistrer l’information géométrique de la scène. Bien que présentant des résolutions de plus en plus grandes, ces caméras utilisent une projection perspective qui limite le champ de vision. Un grand nombre de caméras doit ainsi être utilisé pour couvrir une scène. Les séquences disponibles aujourd’hui pour la TVPL ne comprennent malheureusement qu’une centaine de vues maximum, et pour des scènes restreintes en taille [2].

L’objectif de nos travaux est de changer les modalités de captures en utilisant un système à plusieurs caméras *omnidirectionnelles*. Ces caméras présentent le grand avantage de pouvoir enregistrer à une position donnée, les rayons lumineux provenant de toutes les directions. Ces caméras sont aujourd’hui grandement utilisées pour des applications de réalité virtuelle ou tout simplement par des utilisateurs désirant enregistrer une scène et la restituer avec un sentiment d’immersion. Bien que de plus en plus populaires, ces caméras n’ont été utilisées que dans des contextes où *une seule* enregistrerait la scène. La nouveauté de nos travaux porte ainsi sur une utilisation conjointe de plusieurs caméras pour enregistrer et ensuite restituer une grande partie de la scène. Dans cet article, nous expliquons pourquoi, par principe, une telle approche permettra une meilleure capture de la scène qu’une utilisation de caméras 2D classiques.

Cette idée soulève un certain nombre de problématiques comme la calibration, la synchronisation, l’équilibrage des couleurs, l’estimation de géométrie, la synthèse de vue virtuelle, la compression. Dans cet article, nous nous intéresserons à la *calibration*, consistant à estimer i) le modèle de construction de l’image enregistrée et ses paramètres, et ii) les positions et orientations relatives des caméras par rapport aux autres. Pour cela, les algorithmes de traitements traditionnellement utilisés pour la calibration de réseaux de caméras par exemple, nécessitent d’être adaptés/repensés puisqu’ils auront désormais affaire avec des modèles de caméras différents des caméras perspective classiques.

Dans cet article, nous proposons une stratégie de calibration qui consiste tout d’abord à valider un modèle de caméras, dit le *modèle sphérique unifié*, et l’utiliser pour calibrer plusieurs caméras entre elles. Nous montrons en particulier que le modèle sphérique unifié est valide pour les caméras "*oeil de poisson*" utilisées.

L’article est structuré comme suit. D’abord en Section 2 nous expliquerons pourquoi les caméras omnidirectionnelles

permettent la capture d’une scène de grande ampleur. Ensuite, en Section 3 nous rappellerons les fondements de la géométrie des caméras omnidirectionnelles ainsi que le modèle sphérique unifié adopté. En Section 4, nous présenterons la stratégie de calibration proposée. Ensuite, nous donnerons les résultats expérimentaux obtenus pour la calibration de nos systèmes en Section 5. Enfin, en Section 6, nous conclurons et détaillerons les prochains verrous scientifiques à soulever avant de pouvoir utiliser concrètement un tel système multi-caméras omnidirectionnelles pour de la TPVL.

## 2 Intérêt des caméras omnidirectionnelles pour la Télévision à point de vue libre

Le champ de lumière (également appelé fonction plénoptique) est une fonction  $p(x, y, z, \theta, \phi, t)$  décrivant l’information de couleur à un instant  $t$ , en un point de coordonnées  $[x, y, z]$  et selon une orientation  $[\theta, \phi]$ . L’objectif de la TPVL est de permettre à l’utilisateur d’observer cette fonction en n’importe quel point (*i.e.*, point de vue libre). Le système d’acquisition doit donc capter l’information nécessaire à cette navigation libre au récepteur. Naturellement, il est impossible de mesurer  $p(x, y, z, \theta, \phi, t)$  en chacun des points d’une scène. Cette fonction est en pratique échantillonnée, par des mesures en certains points. On utilise usuellement des caméras à géométrie perspective. Afin de capter la fonction plénoptique en un maximum de points, celles-ci sont souvent disposées en ligne (Figure. 2 (a)) ou en arc de cercle (Figure. 2 (b)). Avec de tels systèmes de capture, la navigation de l’utilisateur se résume principalement à évoluer le long de la ligne ou de l’arc sur lequel sont disposées les caméras. Bien que nouvelle, l’impression d’immersion donnée aux utilisateurs reste limitée. De plus, ce système de capture nécessite un grand nombre de caméras pour une couverture de la scène qui reste modeste.

L’approche que nous nous proposons d’étudier ici repose sur l’utilisation de caméras omnidirectionnelles. Celles-ci permettent en un point  $[x, y, z]$  de capter la fonction plénoptique  $p(x, y, z, \theta, \phi, t)$  provenant de tous les angles  $[\theta, \phi]$ . Autrement dit, l’échantillonnage du champ de lumière se résume à *enregistrer l’information omnidirectionnelle en plusieurs positions* comme indiqué dans la Figure. 2 (c). On voit bien comment cette approche, permet de couvrir une plus grande partie de

la scène à partir d'un ensemble de caméra bien plus restreint que dans le cas de caméras "classiques". Cette approche est récemment devenu possible grâce à l'arrivée caméras omnidirectionnelles sur le marché grand public. Bien sûr, les caméras omnidirectionnelles comportent un certain nombre de désavantages (faibles résolutions, visibilité des caméras dans les autres vues capturées) soulevant des compromis tels que résolution vs couverture ou densité de l'échantillonnage vs pollution de la scène par le matériel de capture.

Cette approche soulève un certain nombre de problèmes nouveaux. Celui que nous nous proposons d'étudier dans ce papier est la calibration du système.

### 3 Modèle d'imagerie omnidirectionnelle

Dans un premier temps, nous rappelons les règles élémentaires de projection perspective. Prenons un point  $\mathbf{p} = [x, y, z, 1]^T$  (coordonnées homogènes exprimées dans le repère *monde*) et une caméra placée à une position  $\mathbf{c}$  par rapport au centre du repère *monde* et orientée selon une rotation  $\mathbf{R}$  par rapport au même repère *monde*. On dit alors que la matrice de projection *extrinsèque* est :

$$\mathbf{P}_{\text{ext}} = \begin{bmatrix} \mathbf{R} & -\mathbf{R}\mathbf{c} \end{bmatrix}. \quad (1)$$

On note  $f$  la distance focale de la caméra (*i.e.*, la distance entre le centre de la caméra et le plan de projection),  $(x_0, y_0)$  les coordonnées du point principal,  $(k_x, k_y)$  la résolution spatiale de l'image, et  $s$  le paramètre de déformation. Alors, la matrice de projection *intrinsèque* s'écrit :

$$\mathbf{P}_{\text{int}} = \begin{bmatrix} k_x f & s & k_x x_0 \\ 0 & k_y f & k_y y_0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (2)$$

Ainsi, un point 3D  $\mathbf{p}$  est projeté sur une image en coordonnées  $[u, v]^T$  grâce aux équations suivantes :

$$\begin{cases} [u', v', z]^T = \mathbf{P}_{\text{int}} \mathbf{P}_{\text{ext}} \mathbf{p} \\ u = u'/z \text{ and } v = v'/z \end{cases}. \quad (3)$$

Le processus de calibration plusieurs caméras perspectives consiste à estimer les paramètres décrivant les matrices intrinsèques et extrinsèques de chacune d'entre elle.

Pour des caméras omnidirectionnelles, les matrices extrinsèques ont la même expression. En revanche, la projection sur l'image ne se fait plus linéairement. En effet, une distorsion radiale intervient, à savoir que la projection dépend de l'angle avec lequel le rayon "rencontre" virtuellement l'image. Les équations de projection dépendent du processus de production de l'image omnidirectionnelle. Autrement dit, suivant qu'elle soit générée grâce à un miroir hyper catadioptrique, para catadioptrique ou par des lentilles dites "oeil de poisson", les règles de projection varient.

Néanmoins, le modèle sphérique unifié [3] permet de modéliser l'ensemble des projections omnidirectionnelles existantes.

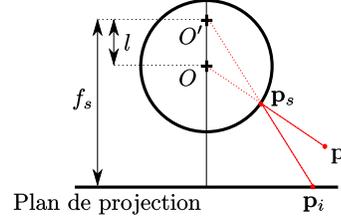


FIGURE 3 – Projection selon le modèle sphérique unifié.

Ce modèle est schématisé dans la Figure 3. Etant donné un point  $\mathbf{p} = [x, y, z]$  dans l'espace 3D, le modèle sphérique unifié décrit comment celui-ci est projeté sur l'image (au pixel  $[u, v]$ ). Tout d'abord, une première projection de  $\mathbf{p}$  sur la sphère de centre  $O$  est effectuée. On obtient un point 3D  $\mathbf{p}_s$  qui vérifie la relation suivante :

$$\mathbf{p}_s = \frac{\mathbf{p}}{\|\mathbf{p}\|} \quad (4)$$

Si  $O$  est pris comme origine du repère. Ensuite, une projection perspective (comme décrite plus haut) du point  $\mathbf{p}_s$  sur le plan de projection est effectué. La spécificité de cette projection est qu'elle est faite depuis le centre de  $O'$ , translaté de  $O$  d'une valeur de  $l$ . Si on considère que les coordonnées mondes sont celles du centre de la sphère, on a :

$$\begin{cases} \begin{bmatrix} u' \\ v' \\ z \end{bmatrix} = \begin{bmatrix} k_x f_s & s & k_x x_0 \\ 0 & k_y f_s & k_y y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{x}{l\sqrt{x^2+y^2+z^2+z}} \\ \frac{y}{l\sqrt{x^2+y^2+z^2+z}} \\ 1 \end{bmatrix} \\ u = u'/z \text{ and } v = v'/z \end{cases}. \quad (5)$$

Bien qu'appelé "unifié", ce modèle sphérique ne correspond pas à tout type de système de capture omnidirectionnelle. En effet, les caméras dites "oeil de poisson", utilisent des lentilles à distorsion radiale. Cette distorsion est généralement une fonction inconnue plutôt modélisable par un polynôme à fort degré ( $\geq 8$ ). Dans ces travaux, nous nous intéressons plus particulièrement au cas de caméras "oeil de poisson" car ce sont les plus répandues parmi les caméras omnidirectionnelles disponibles sur le marché. Tout l'enjeu de ces travaux est de prouver si ce système sphérique unifié permet une bonne modélisation du système de projection d'un point 3D sur une caméra omnidirectionnelle grand public (et donc de type "oeil de poisson").

### 4 Calibration

L'objectif d'un algorithme de calibration, consiste à trouver les paramètres des équations de projection (*cf.* (3) pour les caméras perspectives et (5) pour les caméras omnidirectionnelles). Cela revient à poser un système d'équations dans lequel, les inconnues sont ces paramètres et les valeurs connues sont des coordonnées de points 3D  $[x, y, z]^T$  et leur projetés correspondants  $[u, v]^T$  [4].

Dans nos travaux, nous utilisons une méthode développée dans [3]. La mire utilisée est un damier de forme rectangulaire dont les cases sont carrées. Comme on peut le voir dans la Fi-

gure 4, ce damier est déformée par la distorsion radiale (typique de la vision omnidirectionnelle). L’algorithme de détection automatique des correspondances de  $[x, y, z]^T$  et  $[u, v]^T$ , utilise une estimation d’intersection de parallélogrammes afin de détecter les coins des cases du damier (cf. la fonction `findChessboardCorners()` d’OpenCV). Ensuite, celui-ci prend l’ensemble des jeux de points détectés et trouve les paramètres qui minimisent la distance entre les points  $[u, v]^T$  observés et ceux projetés avec les paramètres et le modèle sphérique unifié. La minimisation se fait grâce à une algorithme de type RANSAC.



FIGURE 4 – Image utilisée lors de la calibration d’une vue.

La calibration de plusieurs caméras utilise les valeurs obtenues par les calibrations indépendantes des caméras, en utilisant conjointement la même mire observée. Les paramètres "extrinsèques relatifs" donnant la position de la caméra par rapport à la mire à chaque instant, sont ensuite triangularisés afin d’en déduire les positions relatives des caméras les unes par rapport aux autres.

## 5 Résultats expérimentaux

Les résultats expérimentaux suivants sont menés dans le contexte d’un projet nommé ATeP (Acquisition Traitement et Partage de données vidéo multi-vues pour la Télévision à Point de Vue Libre). Dans ce projet, nous possédons 40 caméras de type "oeil de poisson" et de modèle "Samsung Gear 360" et désirons capturer des scènes intérieures et extérieures. Les questions que nous nous posons, sont les suivantes : le modèle sphérique unifié est-il un bon modèle pour calibrer un système basé sur le type de caméra utilisé ? Si on effectue la calibration de chacune des caméras indépendamment, obtenons-nous des paramètres intrinsèques similaires, éloignés ou complètement différents ? Nous avons effectué le test avec deux caméras et donc 4 images omnidirectionnelles, car les deux lentilles sont calibrés séparément. Nous observons en pratique que les paramètres obtenus sont très similaires entre les caméras. De plus, comme nous le voyons dans le Tableau 1, chacun d’eux permet une erreur quadratique moyenne entre le pixel détecté et projeté qui est relativement faible (autour de 1 pixel) et qui ne varie pas beaucoup lorsqu’un jeu de paramètre moyen est utilisé. La première conclusion de nos travaux est donc que le *modèle sphérique unifié est bien valide* pour les caméras que nous utilisons, et qu’un jeu de paramètre intrinsèque commun à l’ensemble des lentilles peut être utilisé (sans nécessiter la

	Cam. 1 avant	Cam. 1 arrière	Cam. 2 avant	Cam. 2 arrière
(cam)	0,85	1,13	1,37	1,45
(moy)	1,11	1,36	1,50	1,57

TABLE 1 – Erreur Quadratique moyenne calculée entre la position des pixels mesurée et celle estimée grâce aux paramètres de calibration pour : un jeu de paramètres estimé *par lentille* sur 50 images (cam) et pour un jeu de paramètre *moyen* estimé sur l’ensemble des 50 images des quatre lentilles (moy).

$\Delta x$ réel (mètre)	1	2	3	4	5
écart-type $\Delta x$ estimé (mètre)	0,02	0,03	0,07	0,03	0,09

TABLE 2 – Ecart type moyen des  $\Delta x$  estimés en fonction des  $\Delta x$  réels entre les caméras, lorsque les autres paramètres extrinsèques sont nuls.

calibration de chacune des 80 lentilles de notre système).

Dans le test suivant, nous positionnons deux caméras de manière à ce que seulement les coordonnées  $x$  varient. Nous estimons les positions relatives des caméras pour différentes distances de 1 à 5 mètres. Le Tableau 2 montre les écart-types sur les distances estimées, et permet de conclure que la précision de la calibration est acceptable (entre 2% et 5%).

## 6 Conclusion et travaux futurs

Dans cet article, nous expliquons comment une capture basée sur des caméras omnidirectionnelles rend un système de TPVL envisageable en pratique. Nous nous intéressons à la calibration d’un système de caméras omnidirectionnel de type "oeil de poisson". Nous montrons que le modèle sphérique unifié permet une calibration très fiable et que les comportements intrinsèques des caméras peuvent être considérés semblables, ce qui permet une calibration extrinsèque plus simple.

Dans la suite, nous étendrons ces résultats prometteurs pour développer une calibration des 40 caméras entre elles. Ensuite, nous utiliserons les paramètres de calibrations et les redondances entre les images pour estimer les cartes de profondeurs pour chacune des vues. Enfin, nous utiliserons ces informations géométriques pour effectuer de la synthèse de vues virtuelles, et développer ainsi un système de TVPL concret.

## Références

- [1] M. Tanimoto, "FTV : Free-viewpoint television," *IEEE Signal Processing Magazine*, vol. 27, no. 6, pp. 555–570, Jul. 2012.
- [2] M. Domański, A. Dziembowski, A. Grzelka, and D. Mieloch, "Optimization of camera positions for free-navigation applications," in *International Conference on Signals and Electronic Systems (ICSES)*, Krakow, Poland, Sept 2016, pp. 118–123.
- [3] C. Mei and P. Rives, "Single view point omnidirectional camera calibration from planar grids," in *IEEE International Conference on Robotics and Automation*, Roma, Italy, Apr 2007, pp. 3945–3950.
- [4] B. Li, L. Heng, K. Koser, and M. Pollefeys, "A multiple-camera system calibration toolbox using a feature descriptor-based calibration pattern," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nov 2013, pp. 1301–1307.