

Dispositif de mapping vidéo interactif pour le spectacle vivant

Antoine Villeret antoine.villeret@gmail.com*

Cyrille Henry ch@chnry.net

Annie Leuridan annieleuridan@gmail.com

13 septembre 2012
Révisé le 10 novembre 2013

Introduction Cet article présente le dispositif de vidéo-projection interactif développé au sein du groupe de recherche DRii de l'EnsaLab, laboratoire de recherche de l'École Nationale Supérieure des Arts Décoratifs de Paris. Ce dispositif est utilisé pour le spectacle *Les Fuyantes* de la compagnie de cirque contemporain *Les choses de rien*¹. Ce spectacle a été créé en octobre 2011 à Bonlieu, scène nationale d'Annecy.

Ce dispositif propose différentes stratégies afin de projeter de la vidéo sur des objets en mouvement (mapping dynamique). Il a été développé pour l'utilisation en spectacle vivant avec les contraintes de fiabilité, robustesse, de facilité de mise en œuvre et de réactivité temps-réel que ce cadre impose.

Nous présenterons dans un premier temps les principes d'interaction imaginés et le dispositif technique déployé en conséquence. Finalement, nous détaillerons les aspects techniques d'implémentation mis en place pour chaque type d'interaction.

État de l'art La projection d'images sur une surface afin d'en changer les paramètres (couleur, texture, éclairage, ...) est un champ d'investigation artistique depuis de nombreuses années.

Des artistes comme Bertrand Planes², effectuent cela à partir d'images fixes. Le collectif AntiVJ a développé cette technique, pour créer des effets d'animation sur des façades d'immeuble. Dans ce deuxième cas, même si la projection générée est dynamique, le support est statique, permettant ainsi de réaliser une vidéo en amont, avec des outils de modélisation 3D génériques. Adrien Mondot utilise

FIGURE 1 – *Bump It!* – Bertrand Planes 2008



quant à lui les principes de mapping dynamique en spectacle vivant grâce à son logiciel eMotion³, mais uniquement pour capter quelques points de contrôle afin de générer du mouvement. Dans le spectacle *Apparition* de Klaus Obermaier et Ars Electronica Futurelab⁴, c'est une projection sur les corps des danseurs qui est effectuée.

Nous allons montrer comment nous avons pu généraliser ces dispositifs techniques, afin de rassembler les différentes possibilités, et d'en offrir de nouvelles.

Principes d'interaction Plusieurs principes de mapping vidéo ont été développés afin de permet-

1. <http://www.leschosesterien.com>

2. <http://www.bertrandplanes.com/bumpit/>

3. <http://www.adrien.net/emotion/>

4. <https://www.youtube.com/watch?v=-wVq41Bi2yE&feature=related>

tre :

- la projection de texture sur des éléments de décor mobiles,
- la projection de texture sur des éléments de décor déformables,
- la projection de texture sur des corps en mouvement,
- une "poursuite" automatique des acteurs,
- la création d'une fausse ombre de corps en mouvement.

Ces principes d'interactions sont basés soit :

- sur une analyse d'image afin de repérer la position et d'identifier les objets suivis pour permettre la création de nouvelles images en fonction de ces positions,
- sur une projection de l'image filmée, après transformation,
- sur un mélange de ces deux techniques.

1 Dispositif technique

Présentation du dispositif Afin de pouvoir projeter des images sans interférer avec la captation, nous utilisons une caméra infra rouge (IR). La scénographie a été réalisée à base de tissu lycra gris, ne réfléchissant pas les IR et nous utilisons des marqueurs IR pour générer les informations que va capter la camera. Ces marqueurs sont de deux types :

- marqueur actif : des LED émettent de la lumière IR *cf. 3(a)*,
- marqueur passif : nous utilisons du tissu réfléchissant l'IR *cf. 3(b)*.

Nous avons choisi des LED de 1 W générant une lumière suffisamment importante pour être vue comme une tâche blanche par la caméra IR *cf. 3(a)*.

Pour les marqueurs passifs, nous avons utilisé du tissu rétro-réfléchissant, qui a la particularité de réfléchir la lumière dans l'axe de l'éclairement. Ce matériau réfléchit donc une quantité importante de lumière à la caméra, lorsqu'une source lumineuse se trouve à proximité de son axe focal. Ce dispositif nous permet donc d'obtenir un fort taux de contraste entre le décor et les objets trackés nous permettant d'éviter les traitements lourds d'image tout en rendant le dispositif peu sensible aux perturbations extérieures (fuite de lumière, reflet...).

Le tissu rétro-réfléchissant réagit de manière inhabituelle à la lumière et c'est donc un parti pris

FIGURE 2 – Images issue de la camra IR



(a) marqueur actif



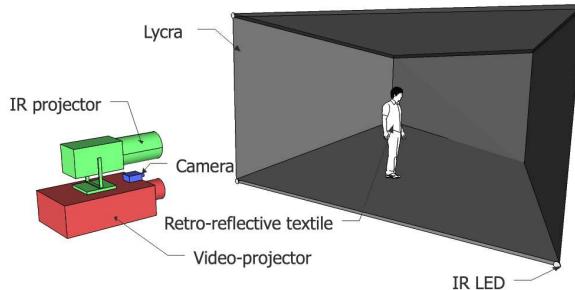
(b) marqueur passif

esthétique très fort que de l'utiliser pour les costumes du spectacle.

La caméra nous donne une position en deux dimensions alors que les éléments sur lesquels on projette se déplacent en trois dimensions. La projection vidéo se fait également en deux dimensions. Afin de pouvoir faire correspondre parfaitement ces deux images dans toute la profondeur de l'espace de projection, il est nécessaire de faire coïncider les deux centres optiques. En pratique, la camera est fixée au support du vidéo-projecteur afin de limiter au mieux la distance entre les deux centres optiques. Ce consulat n'est pas parfait, la profondeur crée

donc une légère erreur, qui est d'autant plus visible qu'on utilise des focales courtes. De plus les focales courtes des vidéo projecteurs ont souvent une profondeur de champ plus faible et donc la projection est plus floue qu'avec une focale longue.

FIGURE 3 – Schéma du dispositif



Pré-requis : calibration Pour que l'image filmée et l'image projetée coïncident parfaitement, il faut aussi que les deux objectifs de la caméra et du vidéo-projecteur soient parfaitement identiques. En pratique d'une part, les deux objectifs ont des focales, des ouvertures (conditionnant la profondeur de champ) et des mises au point différentes et d'autre part, ils ne sont pas parfaits et engendrent des déformations de l'image.

Pour autant, deux calibrations sont nécessaires :

- la calibration des paramètres intrinsèques à la caméra permettant de corriger les déformations liées à l'objectif (*fish-eye*) mais aussi à la fabrication de la caméra (non parallélisme du capteur par rapport au plan image de l'objectif),
- la calibration de la caméra par rapport au vidéo-projecteur.

La calibration des paramètres intrinsèques se fait facilement grâce à l'implémentation de la fonction OpenCV `cvCalibrateCamera2`⁵.

La déformation qui lie l'image de la caméra à celle du vidéo-projecteur est une transformation de perspective. Celle-ci peut être déterminée facilement en connaissant quatre couples de points correspondant entre chaque point de vue. Une implémentation simple consiste à déformer une texture selon ces quatre angles.

⁵. https://pure-data.svn.sourceforge.net/svnroot/pure-data/trunk/externals/pix_opencv

En pratique ces deux calibrations ne sont pas suffisantes. Nous avons constaté des décalages par endroits de plus de 30 cm pour une image de 12 m de base : les erreurs de calibration de la caméra cumulées aux déformations apportées par l'objectif du vidéo-projecteur ne sont pas négligeables. Nous avons donc implémenté la possibilité d'apporter une déformation arbitraire à l'image. En pratique, l'image est déformée à partir d'une fonction bicubique de 5 × 5 points de contrôle. Cette déformation permet ainsi de faire le lien entre un pixel de la caméra et le pixel correspondant du vidéo-projecteur.

Paramètres de calibration Les paramètres de calibration ne doivent être déterminés qu'une seule fois, lors du déploiement du dispositif. En ce qui concerne la calibration des paramètres intrinsèques de la caméra, celle-ci est effectuée automatiquement en filmant un damier selon la méthode décrite par Zhang [3].

La déformation de perspective nécessite le déplacement de quatre points en deux dimensions. Cela s'effectue rapidement manuellement, il n'est pas nécessaire de chercher à automatiser ce processus, bien que ce soit possible avec une méthode similaire à celle de la calibration de la caméra [1]. Cette méthode automatique nécessite entre-autre que la caméra voie l'image vidéo, ce qui n'est pas possible avec une caméra infrarouge.

En revanche la calibration de la caméra avec le vidéo-projecteur est plus délicate. En pratique, nous n'avons pas non plus cherché à l'automatiser. Cette calibration s'effectue manuellement à l'aide d'une mire infra-rouge. Dans le cadre du dispositif pour *les Fuyantes* un grand écran avec des petits carrés de tissu rétro-réfléchissant a été confectionné. En projetant l'image filmée de la mire sur celle-ci, il est possible de déformer l'image de façon interactive afin de faire coïncider l'image des marqueurs IR sur les marqueurs eux-mêmes. Lorsque l'image filmée correspond parfaitement à l'image projetée, le système est correctement calibré. Cette déformation pourrait permettre de corriger également les défauts de la caméra, mais alors la calibration serait plus difficile à effectuer de part le cumul des défauts.

La latence La latence est le temps qui sépare l’acquisition d’une image par la caméra de la projection de cette image par le vidéo-projecteur. Cette latence est très critique pour que l’image projetée suive au plus près les mouvements des objets. Plusieurs stratégies ont été mises en œuvre afin de réduire au mieux cette latence.

Il est techniquement impossible de synchroniser l’acquisition, le traitement et la projection. La latence totale dépend donc de la fréquence de ces trois étapes. Elles doivent donc toutes s’effectuer le plus rapidement possible.

L’utilisation d’une caméra à haute cadence d’images permet de diminuer la latence de l’acquisition. Pour *les Fuyantes*, nous utilisons une caméra AVT Guppy à 60fps.

Le traitement de l’image s’effectue à 100 Hz, cette fréquence est limitée par la puissance de calcul de l’ordinateur.

La latence à l’affichage est limitée par le vidéo-projecteur. Nous utilisons un Sanyo PDG-DET100L cadencé à 85 Hz.

Le tableau 1 montre la latence mesurée dans trois configurations différentes :

- double buffer et synchronisation du changement de l’image sur le balayage de l’écran (mode par défaut, le seul permettant de se passer de tout artefact visuel),
- double buffer, sans la synchronisation sur l’écran (vblank sync),
- simple buffer, sans synchronisation sur l’écran.

Nous avons choisi la dernière configuration.

Le dispositif utilisé pour effectuer ces mesures est basé sur un patch Pure Data, fonctionnant sur un ordinateur sous Linux (Ubuntu 10.04). Le matériel n’est pas le même que celui utilisé en spectacle pour des raisons de disponibilité. La caméra est une Sony Playstation Eye, connectée en USB, fonctionnant à 60fps. L’écran fonctionne aussi à 60Hz. La génération d’images s’effectue à 100Hz. Pour mesurer la latence, l’écran affiche à la fois l’heure système de l’ordinateur, ainsi que l’image de la camera qui filme ce compteur. En enregistrant l’image générée, il est possible par soustraction des valeurs de déterminer la latence d’affichage. En pratique, nous effectuons une moyenne sur une dizaine de mesures. L’écart type de ces valeurs a aussi été calculé. Afin de tester la validité de ces mesures à travers l’heure système, nous testons aussi la fréquence et l’écart type de la génération des images affichées. Nous

constatons que la fréquence d’images générées correspond bien à ce que nous demandons au logiciel : 16,6 ms entre deux images (*i.e.* 60fps) lorsque la synchronisation sur l’écran est activée, 10 ms (*i.e.* 100fps) lorsque qu’elle est désactivée. L’écart type est assez faible, permettant de valider la fiabilité des mesures effectuées. Nous constatons ainsi que la latence la plus faible s’effectue sans surprise lorsque la synchronisation avec l’écran est désactivée. Le passage en simple buffer permet d’améliorer la répétabilité de la mesure.

Afin de réduire le temps entre le calcul de l’image et son affichage, nous avons donc déconnecté la synchronisation avec l’écran, ainsi que le double buffering. Cela peut engendrer quelques artefacts visuels, mais est un facteur important de réduction de latence.

En condition de spectacle, la camera connectée en Firewire nous permet d’obtenir une latence plus faible. Ce n’est pour autant pas suffisant pour les mouvements rapides des acteurs.

Implémentation et optimisations Nous utilisons un ordinateur Apple Macpro à 4 processeurs (8 coeurs) cadencés à 2,8GHz, équipé de 3 Go de RAM et d’une carte graphique ATI Radeon 5770. Le prototypage du dispositif a été fait en grande partie avec Max/MSP/Jitter⁶. Le développement de la version pour la compagnie les Choses de Rien s’est fait dans Pure Data⁷ (aka Pd) avec Gem⁸, sous Linux (Ubuntu 10.04).

Afin d’optimiser l’implémentation, nous avons cherché à exploiter au mieux les capacités de calcul de l’ordinateur. La calibration des paramètres intrinsèques de la caméra se fait grâce à la bibliothèque OpenCV, sur le CPU de l’ordinateur. Un thread⁹ est dédié à ce processus, ainsi qu’à tous les autres traitements lourds en CPU (suivis et reconnaissance par caméra notamment). En majorité les calculs d’affichage se font sur la carte graphique (GPU), afin d’exploiter son importante puissance de calcul. La déformation bicubique de l’image se fait grâce un shader GLSL¹⁰.

6. <http://www.cycling74.com>

7. <http://puredata.info>

8. Gem est l’environnement graphique associé à Pure Data <http://gem.iem.at>

9. [http://fr.wikipedia.org/wiki/Thread_\(informatique\)](http://fr.wikipedia.org/wiki/Thread_(informatique))

10. <http://www.opengl.org/documentation/glsl/>

TABLE 1 – Mesure de latence

	durée d'une image (ms)		latence (ms)	
	moyenne	écart type	moyenne	écart type
double buffer, vsync	16,6	1	110	10
double buffer, no vsync	10	1,5	69	12
single buffer, no vsync	10	2,5	66	3

Le déploiement du dispositif Ce dispositif est amené à être utilisé régulièrement, dans des lieux différents, parfois à des cadences soutenues. Une attention toute particulière a donc été portée à la facilité de mise en place du dispositif.

Dans le cas des *Fuyantes*, la position du vidéoprojecteur est définie grâce aux plans fournis par les théâtres. Une fois sur place, son installation et celle de l'ordinateur prend environ trente minutes. La calibration de la caméra se fait en quelques minutes, et quinze minutes suffisent pour calibrer le vidéoprojecteur. Ces temps d'installation ont pu être réduits grâce au développement des outils de calibration détaillés plus haut.

2 Dispositif artistique

La projection de texture sur des éléments de décor mobile Nous utilisons la bibliothèque AR-Toolkit¹¹ pour reconnaître des patterns imprimés sur un élément plan du décor. Pour éviter que le pattern ne perturbe l'image projetée, il faut le masquer au maximum. Pour cela on peut le dessiner avec une peinture visible seulement par la caméra. En pratique, le textile rétro-réfléchissant peut être brodé sur un autre textile de couleur identique. Cela permet à la caméra, munie d'un projecteur infrarouge, de pouvoir discerner très clairement le pattern dans la scène. Mais la broderie est tout de même visible en lumière ambiante et la projection vidéo la révèle encore plus.

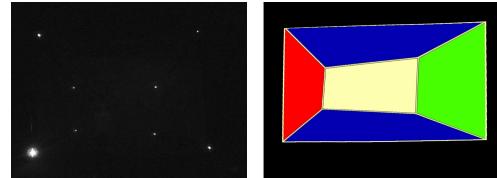
Cette technique permet d'identifier indépendamment chaque élément du décor marqué et plusieurs éléments peuvent être suivis en même temps.

¹¹ <http://www.hitl.washington.edu/artoolkit/> et son implémentation dans Gem pix_artoolkit

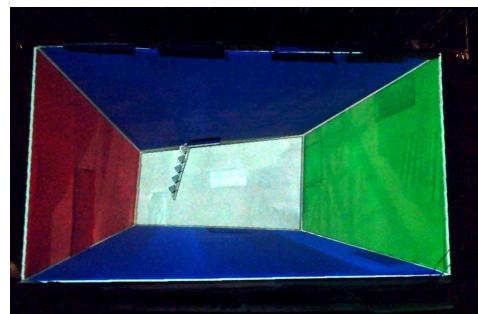
Il serait certainement possible de trouver d'autres techniques pour pouvoir créer un marqueur visible uniquement en IR, mais cette piste n'a pas été développée.

La projection de texture sur des éléments de décor déformables Cette autre méthode consiste à placer des LED infrarouges sur les éléments de décor, typiquement dans les coins.

FIGURE 4 – Tracking du décor équipé de LED IR



(a) Image de la caméra (b) Image envoyée au vidéoprojecteur



(c) Rendu final

Cette méthode permet de faire du mapping sur des surfaces non planes, comme dans le cas de d'Extended View Toolkit [2], mais en utilisant la

caméra pour modifier dynamiquement les points de contrôle.

Cette technique a l'avantage d'être beaucoup plus discrète que la précédente même s'il faut néanmoins pouvoir intégrer l'alimentation des LED dans le décor (pile ou transformateur). À l'inverse, elle ne permet pas de reconnaître facilement les éléments marqués les uns des autres. Pour cela on pourrait dessiner des formes géométriques reconnaissables avec les LED. Mais cette option semble assez lourde en calcul et souffre toujours des problèmes de masquage : pour reconnaître un élément de décor, il faut voir toutes ses LED en même temps.

Une autre option pour permettre l'identification des éléments serait de faire clignoter les LED selon un pattern connu mais dans ce cas le temps de reconnaissance est rallongé (même s'il est possible de synchroniser les clignotements avec le reste de la chaîne de captation).

Cette méthode a été testée dans deux cas de figure :

- reconnaissance de planches manipulées par les interprètes,
- suivi des mouvements d'un décor.

Sur les planches, un interrupteur permettait d'allumer les LED placées aux quatre coins d'une face. Ainsi les interprètes pouvaient choisir à quel moment leur planche serait suivie. Les quatre LED étaient directement utilisées comme points de contrôle des angles d'une image.

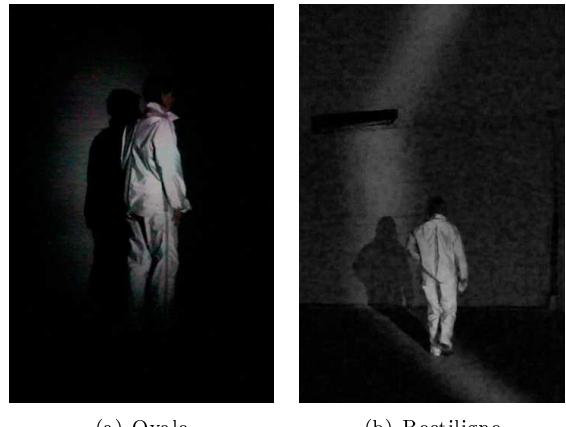
Dans *les Fuyantes*, la scénographie est constituée d'un décor mobile. Ce dispositif permet de projeter des images en renforçant ou en atténuant les perspectives des volumes, tout en s'adaptant aux déplacements de ceux-ci. Le décor était divisé en cinq surfaces, délimitées par huit angles (et donc huit LED), *cf.* 3 Un algorithme spécifique a été développé afin de pouvoir associer les coordonnées des huit LED trackées aux huit points de contrôle de la projection.

La projection de textures sur des corps en mouvement La projection sur les corps en mouvement consiste à projeter de la vidéo sur les corps, ou plutôt sur les costumes des interprètes pendant leur évolution sur le plateau. Pour cela nous utilisons l'image filmée des interprètes comme masque de texture. Dans notre cas, les images captées sont suffisamment contrastées 3(b) pour pouvoir se

FIGURE 5 – Projection sur les corps



FIGURE 6 – Poursuite automatique



(a) Ovale

(b) Rectiligne

passer d'algorithmes de soustraction de fond et de segmentation complexes.

La "poursuite" automatique des acteurs Le suivi des interprètes, qui permet de créer par exemple un effet de poursuite automatique, nécessite de suivre un point particulier de l'image. Pour cela il est possible d'utiliser une LED infrarouge dissimulée dans le costume, ou selon notre dispositif utiliser le fait que seul le costume réfléchisse l'IR, et suivre l'intégralité du costume. La limitation de cette méthode réside dans l'absence d'identification des différents acteurs.

FIGURE 7 – Fausse ombre



Création d'une fausse ombre de corps en mouvement Pour créer une ombre interactive, on extrait la silhouette d'un personnage et on la projette transformée sur la scène. Une transformation de perspective peut permettre de créer un simili d'ombre mais la pauvre qualité de la silhouette affaiblit l'effet comme on le voit sur la fig.7. La déformation géométrique dynamique de l'ombre comme projetée sur l'eau ou brouillée par une transmission hertzienne imaginaire crée un effet qui la détache du mauvais simili et lui confère une valeur esthétique.

Ce dispositif nous a aussi permis de travailler sur un effet humoristique basé sur la temporalité : en utilisant un système de délai variable sur l'image générée, l'ombre peut prendre du retard ou rattraper l'interprète.

Mélange des différents processus Même si tout ces systèmes sont incompatibles entre eux, il est tout à fait possible de les utiliser quasi simultanément. En pratique, l'éclairage IR pour les réflecteurs, ainsi que les huit LED utilisées pour suivre la scénographie sont pilotés depuis la régie. Il est donc possible d'utiliser les réflecteurs passifs lorsque la scénographie est immobile en activant les bons dispositifs de captation.

La figure 8 montre deux principes de mapping interactif utilisés simultanément : des images sont utilisées pour créer des fausses perspectives sur la scénographie. En même temps, les corps des performeurs sont éclairés en blanc dans le cas de la figure 9(a). Alors que dans la figure 9(b), une image noire est projetée sur les corps afin de les faire

FIGURE 8 – Mélanges de plusieurs dispositifs de mapping interactif



(a) Éclairage



(b) Effacement

disparaître.

Perspectives Ce dispositif répond aux contraintes de flexibilité et de fiabilité nécessités par l'utilisation en spectacle vivant. Les aspects techniques ont pu être simplifiés afin de permettre une écriture simple utilisant tous ces principes d'interaction et facilitant également son exploitation. Malgré l'ensemble des possibilités offertes par ces dispositifs, pour certaines parties du spectacle, nous avons préféré effectuer un suivi manuel, essentiellement afin de limiter le temps de développement. La technique n'est pas forcement la meilleure solution à tous les problèmes. Ce dispositif technique offre de multiples possibilités qu'un seul spectacle ne peut pas explorer pleinement ni exploiter justement. Beaucoup de choses restent à tester et à expérimenter.

Références

- [1] Joan Massich Gabriel Falcao, Natalia Hurtos. Plane-based calibration of a projector-camera system. Master's thesis, VIBOT Master, 2008.
- [2] Cyrille Henry Winfried Ritsch Peter Venus, Marian Weger. Extended view toolkit. In *4th international Pure Data Convention 2011 Weimar - Berlin* http://www.uni-weimar.de/medien/wiki/PDCON:Conference/Extended_View_Toolkit.
- [3] Zhengyou Zhang. A flexible new technique for camera calibration. Technical report, Microsoft Research, 1998.

Crédits

La photo de la figure 1 (*Bump it!* – Nemo, Festival Nemo, Paris, Bertrand Planes 2008) a été reproduite avec l'aimable autorisation de Bertrand Planes.

Toutes les autres photos ont été prises par Antoine Villeret lors de répétitions ou de représentations du spectacle *Les Fuyantes* avec la compagnie Les Choses de Rien.