



Science Arts & Métiers (SAM)

is an open access repository that collects the work of Arts et Métiers Institute of Technology researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: <https://sam.ensam.eu>
Handle ID: <http://hdl.handle.net/10985/24971>

To cite this version :

Bruno CABY, Guillaume BATAILLE, Florence DANGLADE, Jean-Rémy CHARDONNET, Frédéric MERIENNE - RemoteAR : Reconstruction d'environnement pour la collaboration en Réalité Augmentée - In: IHM'24 - 35e Conférence Internationale Francophone sur l'Interaction Humain-Machine, France, 2024-03-25 - IHM'24 - 35e Conférence Internationale Francophone sur l'Interaction Humain-Machine - 2024

Any correspondence concerning this service should be sent to the repository

Administrator : scienceouverte@ensam.eu



RemoteAR : Reconstruction d'environnement pour la collaboration en Réalité Augmentée

RemoteAR : Environment reconstruction for collaborative Augmented Reality

BRUNO CABY, Orange - Arts et Metiers Institute of Technology, LISPEN, France

GUILLAUME BATAILLE, Orange, France

FLORENCE DANGLADE, Arts et Metiers Institute of Technology, LISPEN, France

FRÉDÉRIC MERIENNE, Arts et Metiers Institute of Technology, LISPEN, France

JEAN-RÉMY CHARDONNET, Arts et Metiers Institute of Technology, LISPEN, France

La restitution d'environnements réels au sein d'un espace virtuel partagé est un enjeu majeur pour la collaboration en réalité étendue. L'objectif de cette recherche est de proposer un système de captation d'un environnement réel grâce aux seuls capteurs d'un casque de réalité augmentée, afin de le transmettre et de le restituer au sein d'un espace virtuel partagé. Dans cet article, nous explorons différentes méthodes de captation et de restitution de l'environnement, afin d'établir un système simple, mais de qualité suffisante pour faciliter les échanges. Nous avons réalisé une pré-expérience dans le but d'identifier les axes d'amélioration importants, avec pour objectif une future expérience utilisateur centrée sur les interactions entre les collaborateurs.

The restitution of real environments within a shared virtual space is a major challenge for collaboration in extended reality. The objective of this research is to design a system for capturing a real environment using only the sensors of an OST-HMD, in order to transmit it and restore it within a shared virtual space. In this paper we explore the different methods of capture and transmission of an environment to establish a simple system but of sufficient quality to facilitate exchanges. We carried out a pre-experiment focusing on important areas for improvement, with the aim of a future user experience centered on the interactions between the actors of the collaboration.

CCS Concepts: • **Computing methodologies** → **Reconstruction**; *Matching*; • **Human-centered computing** → *Mixed / augmented reality*; *Computer supported cooperative work*; • **Software and its engineering** → *Real-time systems software*; Software performance; • **Applied computing** → Telecommunications; • **Computer systems organization** → Peer-to-peer architectures.

Mots Clés et Phrases Supplémentaires: Reconstruction, Projection, Collaboration, Réalité Augmentée.

Additional Key Words and Phrases: Reconstruction, Projection, Collaboration, Augmented Reality.

Reference:

Bruno Caby, Guillaume Bataille, Florence Danglade, Frédéric Merienne, and Jean-Rémy Chardonnet. 2024. RemoteAR : Reconstruction d'environnement pour la collaboration en Réalité Augmentée. *IHM '24 : Actes étendus de la 35^e conférence Francophone sur l'Interaction Humain-Machine, March 25–29, 2024, Paris, France*

1 INTRODUCTION

Dans le milieu industriel, l'essor du télé-travail, la rationalisation de la main d'oeuvre "spécialisée", ou la séparation des équipes sur des sites distincts nous poussent à repenser les méthodes de travail, et donc les méthodes de collaboration. Le développement des technologies immersives 3D, réunies dans la suite sous le terme de **Réalité Étendue (RE)** : la **Réalité Augmentée (RA)**, **Réalité Virtuelle (RV)**, ou **Réalité Mixte (RM)**, offre de nouvelles perspectives en la

matière, avec la possibilité de collaborer à distance dans des espaces partagés. Les interactions peuvent alors se faire en temps réel (synchrone), ou différé (asynchrone), tout en restant libre de ses mouvements.

La restitution d'environnements réels dans un **Espace Virtuel Partagé (EVP)** pour une collaboration synchrone est un sujet encore peu étudié [2]. En effet, on rencontre principalement l'utilisation de la vidéo 360° pour immerger un utilisateur dans l'environnement d'un autre [9, 11]. Nous avons plus rarement rencontré l'utilisation de reconstructions 3D comme la reconstruction dense par nuage de points [4], ou par maillage [14]. L'ensemble de ces solutions repose généralement sur une technologie asymétrique : la réalité augmentée pour l'utilisateur local, et la réalité virtuelle pour l'utilisateur distant. L'utilisation de la réalité augmentée pour l'ensemble des collaborateurs (symétrique) pourrait permettre une collaboration plus empathique [2], réduire le cybermalaise [7] et éviter à l'utilisateur de percuter des éléments de son espace. Ce qui rend la solution plus adaptée pour une utilisation régulière en milieu industriel.

Cette étude se focalise sur la reproduction des environnements distants dans l'EVP, autour de la **problématique** suivante : *"Comment reproduire les environnements réels en Réalité Augmentée pour permettre la réalisation de tâches collaboratives synchrones ?"*. Nous proposons une approche pour la captation, et la restitution d'un environnement de manière dynamique en temps réel, pour la collaboration en Réalité Augmentée.

2 ÉTAT DE L'ART

Pour reproduire un environnement en temps réel dans un EVP, il faut dans un premier temps capter et restituer les données volumétriques, puis choisir une méthode pour les texturer.

2.1 Capture et restitution de l'environnement

La vidéo 360° est une méthode couramment utilisée. Lee et al. [9] l'utilisent pour comparer l'intérêt de la vision indépendante. Dans le cas de la vision dépendante, le champ de vision de l'utilisateur immergé en RV (ci-après : "utilisateur distant") est le même que l'utilisateur RA, qui porte une caméra 360° fixée à son casque. Dans le cas de vision indépendante, l'utilisateur distant peut contrôler son champ de vision en rotation, mais garde le point de vue de l'utilisateur RA. Rhee et al. [11] utilisent la même méthode, mais cette fois la caméra 360° est fixée sur un trépied, l'utilisateur distant a alors son propre point de vue mais est immobile dans la scène.

La reconstruction 3D est un autre moyen de restituer l'environnement. Gao et al. [4] utilisent un nuage de point 3D statique. Cependant, l'aspect statique indique qu'une modification de l'environnement réel ne sera pas retransmise dans l'EVP, et n'est donc pas compatible avec une collaboration en temps réel. Plus récemment, l'approche de Tian et al. [13] combine à la fois une capture préalable du nuage de point, et la capture dynamique afin de mettre à jour les positions des objets. Cette méthode permet notamment de créer des répliques virtuelles des objets réels afin que l'utilisateur distant puisse les manipuler. L'utilisation d'un maillage 3D est également possible. Dans l'étude de Zilner et al. [14] le maillage de la scène est unique et statique, mais l'utilisateur distant peut manipuler des objets grâce à un système d'extraction de maillage.

Ces travaux sont intéressants car ils explorent des techniques de captation et de restitution différentes, leurs résultats montrent la complexité des compromis sur le sujet. La vidéo 360° limite l'autonomie des utilisateurs par rapport à une reconstruction 3D, mais offre une meilleure compréhension de l'attention de son interlocuteur [9], tout en proposant une méthode simple de capture dynamique [12]. Toutefois, quelle que soit la méthode, nous pensons que l'utilisation de capteurs supplémentaires, ou de la VR, ne permet pas de s'adapter facilement, car elles nécessitent une installation particulière [13], ou encore un espace approprié pour une utilisation sans risque. De plus, la diversification des équipements pourrait entraîner des coûts supplémentaires, ainsi qu'une demande organisationnelle plus importante.

2.2 Affectation des textures

Plusieurs méthodes permettent de texturer des données volumétriques [3] :

- la combinaison d’images : choisir un certain nombre d’images de bonne qualité pour chaque face, et les combiner pour obtenir une texture des faces du modèle,
- la projection d’images : associer à chaque face une image appropriée et projeter cette image sur la face,
- la déformation d’image : introduire des déformations locales de l’image pour qu’elle corresponde au modèle 3D.

Lorsque l’on utilise une texture seule, un effet flou peut être présent sur l’image. Jeon et al. [5] ont combiné la projection et la combinaison d’images, afin d’optimiser le rendu et réduire ces effets de flou. De même Fu et al. [3] ont utilisé cette technique en ajoutant un processus d’optimisation locale pour éviter les discontinuités entre les faces texturées par des images différentes.

Ces techniques sont intéressantes de par la qualité du rendu. Cependant, celles-ci sont exécutées seules sur un ordinateur, avec un temps de calcul de l’ordre de la seconde. Elles seraient donc peu adaptées à une reconstruction en continu, embarquée dans l’Hololens.

3 APPROCHE

Notre étude se concentre sur la reproduction d’environnements distants pour la réalisation de tâches collaboratives synchrones en RA. Nous tentons donc de répondre aux questions de recherche suivantes :

- **Q1** : Quelles méthodes simples de reproduction de l’environnement envisager pour permettre la collaboration en RA ?
- **Q2** : Les performances et la qualité de la méthode choisie sont-elles suffisantes pour permettre une collaboration fluide et efficace ?

3.1 Implémentation

Ici, nous nous intéressons principalement aux briques technologiques de la reconstruction de l’environnement et l’affectation des textures. Concernant la méthode de transmission, nous avons choisi une transmission par UDP sans compression.

3.1.1 Captation et reconstruction de l’environnement.

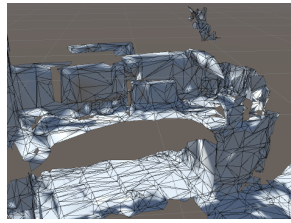
Nous avons dans un premier temps cherché à utiliser la reconstruction automatique de l’Hololens. Le Spatial Awareness est exécuté en continu par l’Hololens et recrée un maillage de l’environnement de l’utilisateur, Fig.1a. Cette méthode présente l’avantage d’être exécutée automatiquement, sans calculs supplémentaires. Cependant, l’ensemble de l’environnement étant gardé en mémoire, beaucoup de données récupérées sont inutiles. Par ailleurs, le taux de rafraîchissement (>1 seconde) ne convenait pas pour permettre une collaboration fluide (~ 0.1 seconde).

Pour contourner cette contrainte logicielle, le mode recherche du casque permet d’accéder aux données brutes des capteurs et notamment celles du capteur de profondeur. Nous avons alors rendu chaque point par un carré. Le rendu de la projection de texture sur ces carrés n’étant pas satisfaisant, nous avons donc choisi de leur affecter une couleur unique. À chaque appel, le capteur retournant au maximum 13000 points, cela revenait à avoir une représentation de notre environnement sur 13000 pixels, ce qui rendait la lisibilité très approximative et inadaptée, Fig.1b.

Notre solution de nuage de point n’étant pas adaptée pour obtenir une restitution nette de l’environnement, nous avons cherché à trouver le meilleur compromis entre les deux solutions précédentes en reconstruisant un maillage et en projetant la texture sur celui-ci. Plusieurs techniques existent pour la reconstruction de triangles à partir d’un nuage de

point, comme la triangulation de Delaunay [1] ou l’algorithme de Marching Cubes [10]. Cependant, nous souhaitons une méthode rapide à implémenter et d’exécution, notre liste de points étant ordonnée nous avons choisi d’utiliser un algorithme plus simple, consistant à trouver les retours à la ligne dans cette liste, puis à créer des faces entre les points de proche en proche. Bien que quelques irrégularités soient présentes sur le maillage, Fig.1c, celles-ci n’étaient pas visibles à l’utilisation.

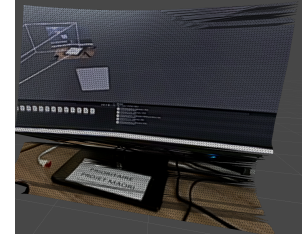
Nous avons résumé les différentes méthodes dans un tableau comparatif Table 1. La solution retenue est la reconstruction de maillage, car elle offrait le meilleur compromis entre le temps de rafraîchissement et la qualité de rendu.



(a) Spatial Awareness : Reconstruction d’un bureau



(b) Nuage de point : Reconstruction d’un écran (Unity)



(c) Maillage : Reconstruction d’un écran (Unity)

FIG. 1 – Méthodes de Captation

Méthode	Qualité	Temps de rafraîchissement	Quantité de données (bytes)
Spatial Awareness	Texture	> 1s	$\sim 10^6$
Nuage de point	13000 pixels	$\sim 10ms$	$\sim 10^5$
Reconstruction du maillage	Texture	$\sim 100ms$	$\sim 10^6$

TABLE 1 – Comparatif des solutions

3.1.2 Affectation des textures.

Le moyen le plus simple pour faire correspondre la texture avec les données volumétriques était de projeter la texture courante sur l’environnement courant. Pour chaque point récupéré, nous faisons correspondre ce point à sa position dans la texture, qui correspond alors aux coordonnées UV du point, Fig.2a.

Pour ce faire, nous utilisons la matrice de projection qui sert à transformer les coordonnées de points de l’espace en coordonnées sur l’écran, ainsi que la matrice de vue de l’espace à la caméra servant à transformer les coordonnées globales en coordonnées locales, relatives à la caméra. La matrice de projection dépendant de la position de la caméra, ainsi que de la taille du champ de vision, nous avons calibré ces valeurs de manière empirique, aidés de valeurs fournies par Microsoft pour l’Hololens.

3.2 Pré-Expérience

Dans le cadre de notre étude, nous avons choisi de réaliser une étude préliminaire qualitative sur la solution de reconstruction de maillage. Le protocole est le même que celui prévu pour une future expérience quantitative, centrée

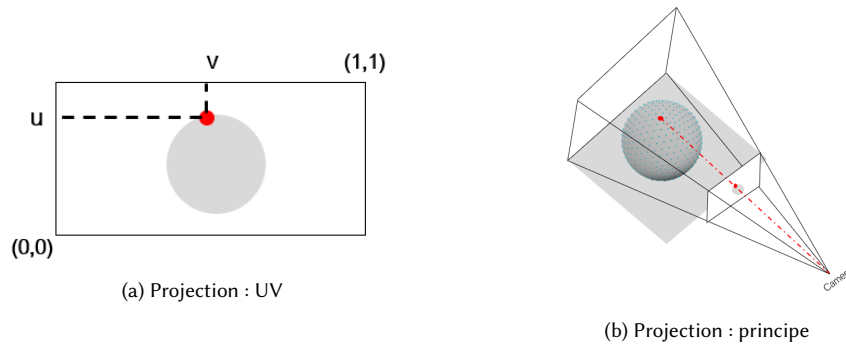


FIG. 2 – Projection de la texture

sur les différents aspects de la collaboration. L'objectif ici est d'identifier les principaux axes d'amélioration sur les solutions choisies, afin d'obtenir les résultats les plus justes possible lors de l'expérimentation finale.

Notre protocole se base sur le cas d'usage du technicien augmenté. Aidé par un expert à distance, le technicien devra réaliser le raccordement d'une pièce à un réseau internet domestique au moyen d'une maquette Fig.3a. Les participants travaillent donc en binôme et sont chacun équipés du casque de RA Hololens 2. Différents objets nécessaires au raccordement sont disposés sur une table et une étagère. Lors de cette pré-expérience, nous souhaitons savoir si l'expert réussissait à se repérer dans l'EVP et à identifier les objets réels.

Pour évaluer les performances de notre application, nous mesurons la fréquence de rafraîchissement de l'environnement, le temps d'envoi des données (vertices, triangles, uv et texture) et la fréquence de rafraîchissement de l'écran. Pour identifier les axes d'amélioration, nous avons enregistré les passages et recueilli les avis des participants les concernant.

4 RÉSULTATS

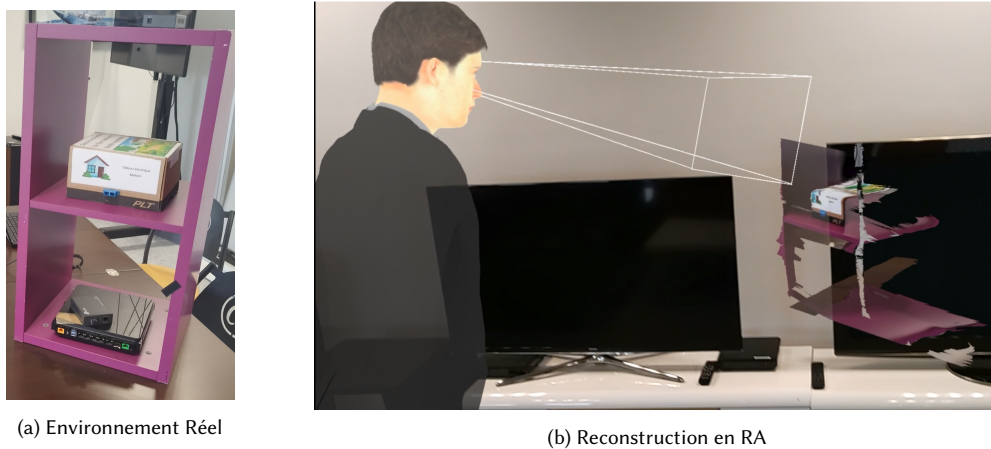


FIG. 3 – Résultat du rendu

Le temps moyen pour le traitement et la transmission d'une trame de donnée était de 334 ms, avec un écart-type de 0,045s, l'écart relatif entre le temps minimum et le maximum étant de 155%.

Concernant les pistes d'amélioration, les participants ont remonté que l'absence de rémanence du maillage induisait un mouvement constant des éléments de l'environnement. En effet, dès lors que l'utilisateur local tournait la tête, l'utilisateur distant pouvait ne plus retrouver la reconstruction actuelle, demandant un effort important à l'utilisateur distant pour se situer dans l'EVP et pour l'utilisateur local qui devait au maximum garder la tête immobile.

De plus, faute de place dans la salle "à distance", les éléments de l'EVP étaient souvent situés dans les murs, entraînant des difficultés à changer d'angle de vue ou à pointer des éléments dans l'EVP. Nous avons donc mis en place un système de relocalisation permettant de se "téléporter" à côté de son interlocuteur.

Enfin, la technique de projection ne donnant pas un résultat parfait à chaque fois, les utilisateurs distants étaient gênés par les effets de décalages entre les faces, Fig.3, affectant la compréhension et empêchant toute action déictique.

5 CONCLUSION ET PERSPECTIVES

Lors de cette étude, nous avons tenté de répondre à la problématique "*Comment reproduire les environnements distants pour améliorer la réalisation de tâches collaboratives synchrones en réalité augmentée?*". Nous avons fait le choix de n'utiliser que les capteurs de l'Hololens pour réaliser la captation de notre environnement et choisi d'utiliser des solutions assez légères en termes d'infrastructures et de développement.

Nous avons exploré plusieurs approches de captation et de rendu. Bien que la quantité de données à transmettre soit significativement plus élevée, la reconstruction de maillage est la méthode qui a offert la meilleure qualité de rendu. Pour la coloration de notre nuage de point ou de notre maillage, nous avons utilisé la technique de projection d'images. Cette méthode est intéressante de par sa simplicité. Néanmoins, elle nécessite un ajustement minutieux dont les paramètres dépendent de la distance avec les objets captés.

Les perspectives qu'offre cette étude sont nombreuses, nous identifions trois améliorations majeures à implémenter avant d'envisager notre étude quantitative sur les effets de la reconstruction 3D sur la collaboration.

Premièrement, l'absence de rémanence des données a été un des principaux facteurs de rejet de l'application durant la pré-expérience. En effet, l'utilisation d'algorithmes de type fusion d'images [6] adaptés à un environnement 3D pourrait permettre de conserver l'affichage des précédentes données sans risquer une mauvaise superposition.

Ensuite, l'amélioration de la technique de projection d'images pourrait engendrer une meilleure acceptation. Plusieurs pistes peuvent être explorées, comme l'établissement d'une démarche de calibration précise, ou l'utilisation de l'Intelligence Artificielle ou de segmentation [8] pour délimiter les textures et les associer aux objets 3D correspondant.

Enfin, la réduction du nombre de points dans le nuage permettrait de limiter la sollicitation du réseau et d'obtenir un meilleur taux de rafraîchissement. Nous envisageons d'extraire des primitives à partir du nuage de point.

RÉFÉRENCES

- [1] Jean-Daniel Boissonnat and Bernhard Geiger. 1993. Three-dimensional reconstruction of complex shapes based on the Delaunay triangulation. In *Biomedical image processing and biomedical visualization*, Vol. 1905. SPIE, 964–975.
- [2] Barrett Ens, Joel Lanir, Anthony Tang, Scott Bateman, Gun Lee, Thammathip Piumsomboon, and Mark Billinghurst. 2019. Revisiting collaboration through mixed reality : The evolution of groupware. *International Journal of Human-Computer Studies* 131 (Nov. 2019), 81–98. <https://doi.org/10.1016/j.ijhcs.2019.05.011>
- [3] Yanping Fu, Qingan Yan, Long Yang, Jie Liao, and Chunxia Xiao. [n.d.]. Texture Mapping for 3D Reconstruction With RGB-D Sensor. ([n.d.]).

- [4] Lei Gao, Huidong Bai, Rob Lindeman, and Mark Billinghurst. 2017. Static local environment capturing and sharing for MR remote collaboration. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications on - SA '17*. ACM Press, Bangkok, Thailand, 1–6. <https://doi.org/10.1145/3132787.3139204>
- [5] Junho Jeon, Yeongyu Jung, Haejoon Kim, and Seungyong Lee. 2016. Texture map generation for 3D reconstructed scenes. *The Visual Computer* 32, 6-8 (June 2016), 955–965. <https://doi.org/10.1007/s00371-016-1249-5>
- [6] Harpreet Kaur, Deepika Koundal, and Virender Kadyan. 2021. Image fusion techniques : a survey. *Archives of computational methods in Engineering* 28 (2021), 4425–4447.
- [7] Andras Kemeny, Jean-Rémy Chardonnet, and Florent Colombet. 2020. Getting rid of cybersickness. *Virtual Reality, Augmented Reality, and Simulators* (2020).
- [8] Sungyoung Kim, Soyoun Park, and Minhwan Kim. 2003. Central Object Extraction for Object-Based Image Retrieval. In *Image and Video Retrieval*, Gerhard Goos, Juris Hartmanis, Jan van Leeuwen, Erwin M. Bakker, Michael S. Lew, Thomas S. Huang, Nicu Sebe, and Xiang Sean Zhou (Eds.). Vol. 2728. Springer Berlin Heidelberg, Berlin, Heidelberg, 39–49. https://doi.org/10.1007/3-540-45113-7_5 Series Title : Lecture Notes in Computer Science.
- [9] Gun A. Lee, Theophilus Teo, Seungwon Kim, and Mark Billinghurst. 2018. A User Study on MR Remote Collaboration Using Live 360 Video. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Munich, Germany, 153–164. <https://doi.org/10.1109/ISMAR.2018.00051>
- [10] William E Lorensen and Harvey E Cline. 1998. Marching cubes : A high resolution 3D surface construction algorithm. In *Seminal graphics : pioneering efforts that shaped the field*. 347–353.
- [11] Taehyun Rhee, Stephen Thompson, Daniel Medeiros, Rafael dos Anjos, and Andrew Chalmers. 2020. Augmented Virtual Teleportation for High-Fidelity Telecollaboration. *IEEE Transactions on Visualization and Computer Graphics* 26, 5 (May 2020), 1923–1933. <https://doi.org/10.1109/TVCG.2020.2973065>
- [12] Theophilus Teo, Louise Lawrence, Gun A. Lee, Mark Billinghurst, and Matt Adcock. 2019. Mixed Reality Remote Collaboration Combining 360 Video and 3D Reconstruction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–14. <https://doi.org/10.1145/3290605.3300431>
- [13] Huayuan Tian, Gun A Lee, Huidong Bai, and Mark Billinghurst. 2023. Using Virtual Replicas to Improve Mixed Reality Remote Collaboration. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023), 2785–2795.
- [14] Jakob Zillner, Erick Mendez, and Daniel Wagner. 2018. Augmented Reality Remote Collaboration with Dense Reconstruction. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 38–39. <https://doi.org/10.1109/ISMAR-Adjunct.2018.00028>