

# La réalité augmentée



## Gilles SIMON

**CV** Gilles Simon, ancien élève de l'École Supérieure d'Informatique et Applications de Lorraine (aujourd'hui Télécom Nancy) a soutenu en 1999 une thèse de troisième cycle d'informatique portant sur le positionnement robuste en réalité augmentée à l'Université Henri Poincaré, Nancy-I. Après un séjour post-doctoral à l'Université d'Oxford dans le groupe Visual Geometry, où il a travaillé sur des problèmes de vision par ordinateur en collaboration avec la société Oxford Metrics (2d3, Vicon), il est recruté comme maître de conférences à l'UHP en 2000. Ses travaux au sein de l'équipe MAGRIT du Loria (Laboratoire lorrain de recherche en informatique et ses applications) concernent la calibration et le suivi de caméra, la reconstruction 3D et la réalité augmentée en vision par ordinateur. Il intervient sur ces sujets dans plusieurs formations de l'Université de Lorraine (master Interaction et Nouvelles Images, licence et master d'informatique, Télécom nancy, Supelec Metz, ...). Il est l'auteur d'une trentaine de publications internationales dans le domaine de la réalité augmentée, et d'un ouvrage paru aux Éditions Dunod, «Intégrer images réelles et images 3D» destiné aux professionnels de l'audiovisuel et à tout public intéressé par le sujet.



Figure 1 : Intégration d'un bonhomme géant sur la terrasse du musée des Beaux-arts de Stockholm.

L'équipe MAGRIT de l'INRIA Nancy Grand-Est se consacre depuis près de deux décennies à résoudre des problèmes scientifiques et techniques posés par la réalité augmentée (RA). Pour autant, cette technologie n'en est qu'à ses balbutiements. Le grand public la découvre tout juste, à travers des jeux vidéos, des parcs d'attractions et des sites Internet. Mais en permettant à l'être humain de superposer à la réalité des informations ou des images tridimensionnelles jusqu'ici inaccessibles ou présentées de manière dissociée, la RA est aussi un outil précieux pour l'aide à la décision et à la compréhension, dans des domaines à forts enjeux sociétaux tels que l'industrie, l'urbanisme, l'aide à la mobilité, l'éducation et la santé. Dans cet article, nous présentons certains problèmes scientifiques soulevés par cette thématique, et tâchons de dissiper quelques doutes ou idées reçues sur le potentiel actuel et futur de la RA, conduisant parfois à des réactions excessivement hostiles ou au contraire, un peu trop enthousiastes, autour de cette technologie.

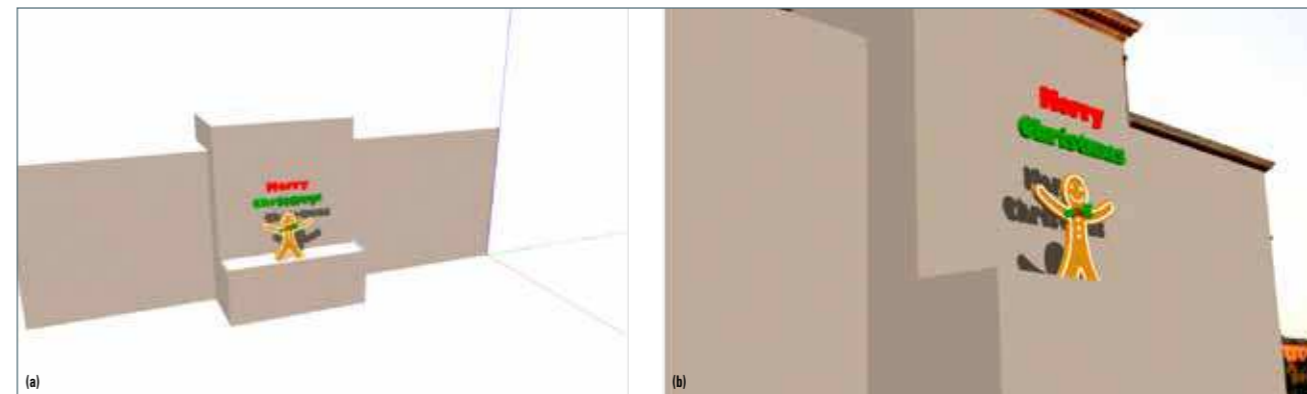


Figure 2 : Aligement des scènes réelle et virtuelle. (a) La scène virtuelle est positionnée relativement à un modèle 3D de la scène. (b) Le modèle 3D est superposé à la scène, avant d'être camouflé pour ne laisser apparaître que les éléments ajoutés.

## Problématique

Commençons par dissiper un premier doute : la RA ne permet pas de faire apparaître des images tridimensionnelles directement dans notre environnement. En croisant des rayons d'énergie, il est physiquement possible de matérialiser des photons à l'intérieur de certains gaz, mais cela est impossible dans l'air. La RA ne fait qu'augmenter notre propre perception de la réalité. Pour que cela soit possible, il est nécessaire de passer par un dispositif d'affichage bidimensionnel, qui peut être monoscopique (écran ou vitre) ou stéréoscopique (paire de lunettes). Dans le cas des vitres et des lunettes semi-transparentes, la réalité est perçue directement à travers les verres, sur lesquels sont projetées les images virtuelles. Dans le cas des écrans (ordinateurs, téléphones ou lunettes opaques) la réalité est filmée à l'aide d'une ou deux caméras, et les images obtenues sont combinées avec les images virtuelles avant d'être affichées sur les écrans.

Pour que cette combinaison bidimensionnelle soit physiquement cohérente, il faut d'une certaine manière comprendre ce que l'on voit : les mouvements, les formes, les couleurs, la lumière. Si pour la plupart des êtres humains cette compréhension est immédiate, l'acquisition automatique des informations nécessaires à une superposition cohérente des deux mondes pose de sérieux problèmes. D'autant que ces informations évoluent au cours du temps : les déplacements de l'observateur, ses mouvements de tête, le passage d'objets mobiles dans l'environnement et les modifications d'éclairage transforment à chaque instant les images à augmenter.

Un exemple permettra de mieux circonscrire la problématique de la RA : supposons que nous désirions voir un bonhomme de pain d'épice géant à la terrasse du musée des Beaux-arts de Stockholm (figure 1). Pour cela, il faut résoudre plusieurs problèmes :

- 1 Exprimer le fait que le bonhomme (du monde virtuel) est situé sur la terrasse (du monde réel),
- 2 Représenter le bonhomme de manière à ce que sa position et son orientation soient cohérentes avec la position et l'orientation de la terrasse dans la photographie,
- 3 Faire en sorte que les pieds du bonhomme soient grignotés afin d'apparaître comme occultés par le parapet,
- 4 Représenter l'ombre du bonhomme sur la façade du musée.

Il est en fait possible de résoudre le problème 1 de telle sorte que les problèmes 2, 3 et 4 se résument à un seul et unique problème. Cela nécessite de disposer d'un modèle tridimensionnel de la scène observée. La position et l'orientation des éléments virtuels, ainsi que la position des sources lumineuses, peuvent alors être définies par rapport à ce modèle (figure 2.(a)), et l'on se ramène à un problème d'alignement entre un modèle 3D et son observation (figure 2.(b)). En pratique, le modèle est camouflé dans l'image augmentée, de telle sorte que l'on ne puisse voir que les effets des interactions spatiales (occultations, collisions) et photométriques (ombres portées, reflets) entre le modèle et les objets virtuels (figure 1).



# La réalité augmentée

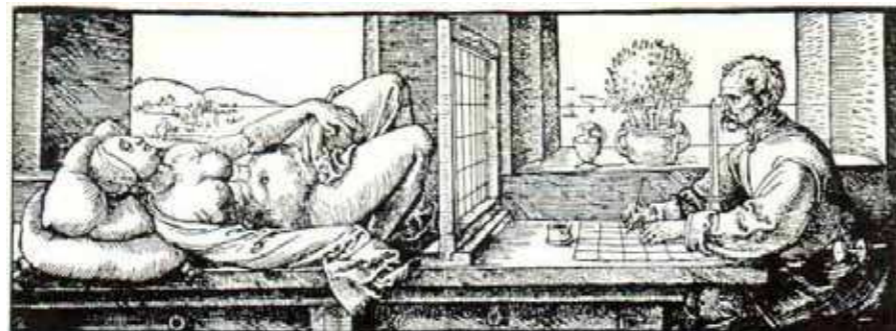


Figure 3: Albrecht Dürer, Instructions pour mesurer (Le dessinateur de la femme couchée), 1527.

## Alignement des scènes réelle et virtuelle

Le modèle utilisé pour dessiner la scène virtuelle est celui de la perspective conique. Ce modèle a été proposé à la Renaissance afin de représenter le monde conformément à notre propre perception visuelle. Il repose sur l'hypothèse que la lumière se propage en ligne droite, rebondissant éventuellement sur des objets avant d'atteindre notre rétine. La couleur des objets correspond ainsi à notre perception subjective des fréquences d'ondes non absorbées par les objets rencontrés. Une gravure d'Albrecht Dürer (figure 3) montre comment dessiner selon les lois de la perspective conique. Le principe est de regarder la scène à travers une grille, depuis un point de vue d'observation fixe. Chaque case de la grille correspond à une case du dessin, et la position des éléments de la scène relative aux bords des cases est reproduite sur le dessin. Cette technique est encore utilisée de nos jours par de nombreux peintres.

Supposons que, son travail accompli, le peintre décide le lendemain de dessiner un drap sur le buste de son modèle. Nous sommes alors dans une situation analogue à celle de la RA, le drap remplaçant la scène virtuelle et le tableau de la veille l'image à augmenter. Le peintre devra alors faire revenir son modèle, le couvrir d'une pièce de tissu supplémentaire, puis lui demander de poser exactement de la même manière que la veille et dans des conditions d'éclairage identiques. Pour que les traits ajoutés au dessin soient correctement alignés avec les éléments apparaissant dans la première version du tableau, il devra aussi positionner la grille et le repère d'observation aux mêmes endroits que la veille. Le problème de l'alignement est donc en fait un problème de positionnement. Dans le cas d'un flux d'images à traiter en temps réel, il est nécessaire de connaître à chaque instant le point de vue de l'observateur ou de la caméra par rapport à la scène observée ou filmée.

Ce problème de positionnement est commun à la RA et à la robotique. Pour le résoudre, il est possible d'utiliser des capteurs physiques de position et d'orientation : en intérieur, des capteurs magnétiques, optiques ou acoustiques peuvent être utilisés en appliquant le principe de triangulation. La précision de ces capteurs est bonne, mais leur portée est limitée à quelques mètres et leur utilisation nécessite d'installer un dispositif onéreux et complexe à manipuler. En extérieur, une boussole couplée à des accéléromètres et à un gyroscope permet de mesurer une orientation dans le repère terrestre; un capteur tel que le GPS (Global Positioning System) fournit par ailleurs une position (latitude et longitude) dans ce même repère. Il est donc possible de résoudre le problème de l'alignement dans des images capturées depuis un dispositif muni de tels capteurs. Malheureusement, la précision de ces capteurs est faible et sensible aux conditions environnementales et météorologiques. De l'ordre d'une dizaine de mètres pour le GPS et de quelques degrés pour les capteurs d'orientation, elle se traduit fréquemment par des décalages de plusieurs dizaines de pixels dans les images vidéo. Cette précision est suffisante pour des applications visant à afficher des informations textuelles par-dessus les lieux ciblés, mais insuffisante pour d'autres utilisations telles que l'étude d'impact en architecture ou l'aide à la mobilité. À l'heure actuelle, le seul moyen d'obtenir une précision plus importante en milieu extérieur est d'utiliser la vision par ordinateur<sup>(1)</sup>.

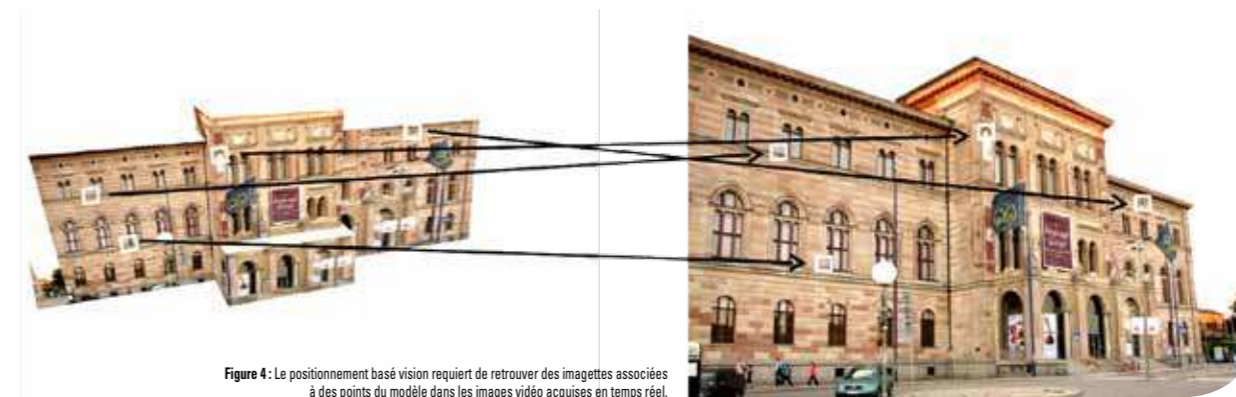


Figure 4: Le positionnement basé vision requiert de retrouver des imagerie associées à des points du modèle dans les images vidéo acquises en temps réel.

## Positionnement basé vision

Puisqu'on cherche à aligner un modèle 3D dans une image, il semble naturel de considérer que l'image elle-même peut aider à résoudre le problème. En effet, si l'on parvient à identifier dans une image vidéo des zones d'ancrage du modèle, il est possible de calculer le point de vue d'observation (position + orientation de la caméra) correspondant à cette image. Ces zones d'ancrage sont la plupart du temps des points (coins ou autres types de points saillants), mais on peut aussi considérer les arêtes du modèle qui correspondent généralement à des gradients forts (contours) de l'image. Si l'on parvient à identifier six points du modèle dans l'image, le point de vue de la caméra associé à cette image est calculable. En pratique, un plus grand nombre de points est utilisé afin de réduire la sensibilité au bruit des mesures.

La principale difficulté réside aujourd'hui dans la capacité à retrouver les primitives du modèle dans les images vidéo. Cette tâche peut être aisée si l'on place des marqueurs artificiels dans la scène, dont les positions sont définies par rapport au modèle. On sait reconnaître des marqueurs artificiels en temps réel depuis une quinzaine d'années, et la majorité des applications de RA utilisent ce procédé. Cependant, l'utilisation de marqueurs n'est pas adaptée aux environnements de grande taille.

Le positionnement sans marqueur repose généralement sur l'utilisation d'un modèle texturé de l'environnement. Le problème revient à retrouver des imagerie associées à des points du modèle dans les images vidéo (figure 4). Malheureusement, les distorsions géométriques dues à la perspective et les variations d'éclairage peuvent rendre les imagerie difficilement reconnaissables dans les images vidéo. Si l'on dispose d'une connaissance approximative du point de vue (par exemple, en utilisant des données capteurs), les déformations dues à la perspective peuvent être atténuées en synthétisant une image du modèle tel qu'il serait observé depuis le point de vue hypothétique. Cependant, il n'est pas toujours possible d'obtenir cette connaissance. Beaucoup de travaux ont visé ces dernières années à retrouver dans une image des imagerie ayant subi une forte distorsion géométrique et/ou photométrique. En fait, plutôt qu'à des imagerie, on s'est intéressé à des descripteurs de points, c'est-à-dire des vecteurs dont les coefficients sont liés à la distribution des valeurs de l'image autour des points détectés. L'objectif étant que ces coefficients soient le plus invariants possible aux changements de points de vue et d'éclairage.



(1) La précision du GPS peut être améliorée notablement en utilisant un GPS différentiel : un réseau de stations fixes de référence transmet au GPS l'écart entre les positions indiquées par les satellites et sa position réelle. Du fait de son coût prohibitif, cette solution n'est cependant pas accessible au grand public.

# La réalité augmentée



Figure 10



Figure 11



Figure 5



Figure 6



Figure 8

## Applications actuelles et à venir

La RA est aujourd'hui utilisée dans de nombreux domaines : héritage culturel, urbanisme, e-commerce, santé, éducation, industrie, divertissement, art, etc. Chaque nouvelle technologie de positionnement a apporté son lot de nouvelles applications. Ainsi, il y a plus de dix ans, l'Université de Canterbury a mis à disposition du public un kit de développement appelé ARToolkit, qui permet de construire des applications de RA basées sur l'utilisation de marqueurs artificiels. Des centaines, voire des milliers d'applications reposant sur ce kit ont vu le jour depuis. Il s'agit essentiellement de jeux (figure 5) ou d'applications un peu gadget, parfois utilisées dans le cadre d'opérations de marketing. Mais des applications très intéressantes ont aussi pu voir le jour, comme des ouvrages pédagogiques dont les illustrations apparaissent en RA. Les marqueurs pouvant être facilement manipulés, il est aussi possible de les utiliser pour interagir physiquement avec un environnement virtuel, comme nous l'avons proposé dans une application d'aide à la compréhension des phénomènes lunaires réalisée en collaboration avec l'IUFM de Lorraine (figure 6). Les capteurs physiques de mouvement ont rapidement intéressé les médecins, pour leur précision et leur robustesse en environnements contrôlés. Ils permettent de visualiser des images

médicales tridimensionnelles directement par-dessus le corps du patient ou dans des images issues d'une autre modalité d'acquisition telles que les images endoscopiques (figure 7). Ces capteurs sont aussi utilisés pour aider à réaliser certaines tâches de maintenance ou d'assemblage : l'utilisateur est guidé pas à pas grâce à des indications visuelles se superposant aux pièces à manipuler (figure 8). La commercialisation à l'échelle planétaire de smartphones équipés de GPS et de capteurs d'orientation a rendu la RA accessible au plus grand nombre. Les premières applications ayant vu le jour permettent d'afficher des informations géoréférencées (stations de métro, sites touristiques et autres points d'intérêts) par-dessus les images capturées en temps réel par la caméra équipant le téléphone (figure 9). Des applications plus dédiées ont suivi, notamment dans les domaines de l'astronomie (figure 10) et de l'aménagement urbain (figure 11). Les avancées des recherches en vision par ordinateur ont permis d'obtenir des applications sans marqueurs ni capteurs, utilisables en environnements confinés ou sur des parties du corps humain. On peut par exemple essayer des lunettes de vue en RA sur le site Internet de Ray-Ban, à l'aide d'une simple webcam.



Figure 9



Figure 12

En revanche, il faudra encore attendre un peu avant de pouvoir utiliser des méthodes purement visuelles pour la RA en extérieur urbain. Les raisons sont multiples : changements d'apparence de la scène observée entre l'acquisition des modèles et leur utilisation (suivant les saisons, l'heure de la journée, l'état des devantures, l'encombrement des rues, etc.), diversité des caméras utilisées (résolution, mise au point, ...), très grands changements de points de vue d'observation probables, présence de nombreux motifs répétés (toutes les fenêtres se ressemblent !). Coupler les données images à des données capteurs permet de restreindre l'espace de recherche des primitives du modèle et de lever certaines ambiguïtés, mais ne résout pas toutes les difficultés. Des travaux sont en cours pour tenter de résoudre ces divers problèmes. En dehors des applications basées marqueurs, toutes ces applications reposent sur la connais-

sance d'un modèle 3D de la scène. Les modèles peuvent être obtenus de diverses manières (mesures laser, photogrammétrie, images satellitaires etc.), mais le fait que l'on ne puisse pas s'en passer est une contrainte importante qui exclut d'utiliser la RA dans des environnements non connus a priori. Cela nous prive actuellement d'un grand nombre d'applications de proximité (aménagement de jardins particuliers par exemple) ou visant des environnements éphémères (comme l'affichage d'effets spéciaux immédiatement durant la phase de tournage d'un film). Pour que ce type d'applications puisse voir le jour dans les prochaines années, des chercheurs travaillent sur des méthodes d'acquisition in situ de l'environnement, permettant d'obtenir facilement et rapidement des modèles de scènes, juste avant de démarrer l'application de RA ou en même temps que celle-ci opère (figure 12).

## Conclusion et perspectives

Dans cette brève introduction à la réalité augmentée, nous avons essayé de montrer que si cette technologie permet en effet d'ajouter tout et n'importe quoi à notre champ visuel (l'infographie offre aujourd'hui autant de liberté que la peinture), elle ne permet en revanche pas de le faire en tout lieu et dans n'importe quelles conditions. Le choix de la méthode utilisée pour le positionnement (capteurs physiques, marqueurs artificiels, vision pure ou méthodes hybrides) dépendra de la précision escomptée et du type de scène considérée (intérieur, extérieur, présence ou absence de textures, conditions d'éclairage, etc.). On peut s'attendre à une évolution et peut-être à une convergence des différentes méthodes utilisées aujourd'hui. D'autres dispositifs de visualisation sont aussi en cours d'étude, tels que des lentilles de contact qui permettront un jour d'afficher des éléments infographiques au plus près de nos rétines. Mais augmenter la réalité restera un acte volontaire, nécessitant de se munir d'un dispositif dédié pour voir apparaître les objets ajoutés.

La vision par ordinateur continuera à jouer un rôle important dans les avancées futures, ne serait-ce que pour permettre l'acquisition des modèles. Nous assistons aujourd'hui à l'édification communautaire de bases de données gigantesques de monuments et de constructions modélisés en 3D et déposés quotidiennement sur des SIG (Systèmes d'Information Géographique) tels que Google Earth. Proposer des outils plus efficaces pour aider les contributeurs à modéliser leur environnement proche permettra d'accélérer le processus en cours. Par ailleurs, les imprimantes/scanners de documents plats seront bientôt remplacés par des imprimantes / scanners d'objets solides. Nous pourrions donc bientôt acquérir en 3D nos objets quotidiens, et les dupliquer à volonté.

Dans cet article, nous avons considéré le problème de la RA comme chercheurs en vision par ordinateur. Mais la réalité augmentée fait l'objet d'études dans bien d'autres domaines : robotique, électronique, optique, télécommunication, ergonomie, psychologie cognitive, psychophysiology, philosophie, économie, etc. C'est que la vue est une des fonctions les plus utiles à l'évolution de l'espèce humaine, il convient donc de considérer ses extensions avec le plus grand sérieux !