HAL
open science

Spatial Augmented Reality for Physical Drawing

Jérémy Laviole

To cite this version:

Jérémy Laviole. Spatial Augmented Reality for Physical Drawing. Human-Computer Interaction [cs.HC]. Université Sciences et Technologies - Bordeaux I, 2013. English. NNT : . tel-00935602v1

HAL Id: tel-00935602

https://theses.hal.science/tel-00935602v1

Submitted on 23 Jan 2014 (v1), last revised 29 Sep 2014 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

PRÉSENTÉE À

L'UNIVERSITÉ BORDEAUX 1

ÉCOLE DOCTORALE DE MATHÉMATIQUES ET INFORMATIQUE

Par Jérémy LAVIOLE

POUR OBTENIR LE GRADE DE

DOCTEUR

SPÉCIALITÉ : INFORMATIQUE

Interaction en Réalité Augmentée Spatiale pour le Dessin Physique

Directeur de thèse : Martin Hachet

Soutenue le : 5 Décembre 2013

Devant la commission d'examen formée de :

M. LECUYER, Anatole	Directeur de recherche	Inria	Rapporteur
M. DUBOIS, Emmanuel	Maître de conférences	Université de Toulouse	Rapporteur
M. HACHET, Martin	Chargé de recherche	Inria	Directeur de thèse
Mme. ROUILLON COUTURE, Nadine	Professeure	ESTIA	

Université Bordeaux 1
Les Sciences et les Techniques au service de l'Homme et de l'environnement

Interaction en Réalité Augmentée Spatiale pour le Dessin Physique

Résumé

Cette dissertation décrit le design, implémentation et évaluation de nouvelles applications en Réalité Augmentée Spatiale (RAS). Ces applications sont concentrées sur l'amélioration du dessin physique tel que les dessins au crayons ou peintures en projetant des outils numériques.

Tout d'abord, nous décrivons notre système de RAS et ses possibilités. Il prend en comptes les paramètres internes et externes d'un couple caméra/projecteur pour permettre une projection précise sur des feuilles de papier. De plus, il permet la détection du toucher des feuilles et de la position de la main au dessus de celles-ci grâce à une caméra de profondeur. Par conséquent, il permet la création d'écrans tactiles interactifs sur des feuilles de papier posées sur une table.

Ensuite, nous nous penchons sur la création d'art visuel, plus précisément sur les premières étapes de la création quand l'artiste créer la structure. Nous offrons la possibilité de créer et d'éditer des lignes de construction numériques (LCN) projetées sur le papier. Ces outils sont des outils de Réalité Augmentée (RA), c'est-à-dire qu'ils s'ajoutent aux outils existants: toute l'interface utilisateur est sur la table, et l'utilisateur n'utilise jamais ni une souris, ni un clavier ni un écran. En plus des simples LCN (lignes et courbes), nous proposons une spécialisation pour les dessins spécifiques tels que les dessins en perspective, les dessins de personnages ainsi que les dessins à partir d'un modèle 3D. Nous proposons de nouvelles méthodes pour afficher et interagir avec des objets 3D sur papier. Nous proposons également de créer des dessins mixtes: art visuel interactif qui tire parti à la fois des possibilités physiques et numériques.

Pour finir, nous décrivons des nouveaux usages pour notre système de RAS de nombreux contextes différents à travers des démonstrations publiques. L'acceptabilité de ce genre de système a été très bonne, et "magique" par la plupart des utilisateurs. Ils ont juste vu et interagis avec des feuilles de papier sans remarquer le système de projection et suivi.

Mots clés:

Réalité augmentée spatiale, projection, papier, interaction tangible, camera de profondeur, dessin, création artistique, multi-touch, code créatif, interaction spatiale

Spatial Augmented Reality for Physical Drawing

Abstract

This dissertation presents the design, implementation, and evaluation of novel Spatial Augmented Reality (SAR) applications. These applications focus on enhancing physical drawing such as pencil drawing or painting by the projection of digital tools.

First, we describe our tabletop SAR system and its possiblities. It takes into account the camera and projector internal and external parameters to enable precise projection onto tracked sheets of paper. In addition to this, we enable touch and hand detection over the table with a depth camera. Consequently, it enables the creation of interactive touch screens on the pieces of paper on a table.

Next, we focus on the creation of visual art, more precisely the first steps of creation when the artist structures it. We enable the creation and edition of digital construction lines (DCL) that are projected onto the paper. These tools are Augmented Reality (AR) tools, *i.e.*, they add to existing tools: the whole user interface is on the table and the user never uses a mouse, keyboard or screen. In addition to the simple DLC (lines and curves), we propose to specialize them for specific drawings such as perspective drawing, character drawing and drawing from a 3D model. We propose a novel way to display and interact with 3D objects on paper. We also propose to create mixed media drawing: it enables the creation of interactive visual art that takes advantage from both the physical and digital possiblities.

Finally, we describe new usages for our SAR system in many different contexts through public exhibitions. The acceptability of such systems was found very good and "magical" by most of the users. They only saw and interacted with interactive pieces of paper without noticing the projection and tracking system.

Keywords:

spatial augmented reality, projection, paper, tangible interaction, depth camera, drawing, artistic creation, multi-touch, creative coding, spatial interaction

Intitulé et adresse du labo... A élaborer

Laboratoire Bordelais de Recherche en Informatique
Unité Mixte de Recherche CNRS (UMR 5800)
351, cours de la Libération F-33405 Talence cedex

Centre de Recherche
Inria Bordeaux - Sud-Ouest
200, avenue de la Vieille Tour 33405
Talence cedex

Épigraphe et remerciments

Résumé substantiel en Français. À venir.

Contents

Contents **8**

I Augmented Reality research **15**

1 Augmented Reality **17**
1.1 Vocabulary . . . 17
1.2 Definition . . . 18
1.3 Visualization . . . 18
1.3.1 See through . . . 18
1.3.2 Spatial . . . 20
1.4 Tangible Interfaces . . . 21
1.4.1 Definition . . . 21
1.4.2 Tracking techniques . . . 22
1.5 Paper Interfaces . . . 24
1.5.1 Augmented Documents . . . 24
1.5.2 Augmented art . . . 25
1.6 General public . . . 27
1.7 Contributions . . . 27
1.8 Future of Augmented Reality . . . 27

2 SAR system **29**
2.1 System Design . . . 29
2.1.1 Drawing medium tracking . . . 30
2.1.2 Vision-based tracking . . . 30
2.1.3 Touch detection . . . 31
2.2 Spatial Augmented Reality . . . 33
2.2.1 Image formation model . . . 33
2.2.2 Extrinsic Calibration . . . 34
2.2.3 Calibrated Virtual Camera . . . 34
2.3 Multi-touch with Kinect . . . 37
2.3.1 Transformation from Kinect point of view to normalized coordinates . . . 37
2.3.2 Detection and tracking . . . 39

2.4 Performance . . . 40
2.4.1 Paper tracking . . . 40
2.4.2 Touch tracking . . . 41
2.4.3 Projection size . . . 42
2.4.4 Digital Screen on paper . . . 43

II Augmented Drawing **45**

3 Guiding the drawings **47**
3.1 Motivations . . . 47
3.1.1 Projection for drawing . . . 47
3.1.2 Preliminary work on projection for drawing . . . 48
3.2 Image filters . . . 48
3.3 Construction lines . . . 50
3.3.1 General purpose construction lines . . . 51
3.3.2 Augmented keypad . . . 51
3.3.3 Pointing . . . 52
3.4 User evaluation . . . 52
3.4.1 Task . . . 53
3.4.2 Participants . . . 54
3.4.3 Apparatus . . . 54
3.4.4 Results . . . 54
3.5 Dedicated construction lines . . . 58
3.5.1 Perspective construction lines . . . 58
3.5.2 Character drawing . . . 59

4 Drawing of a 3D digital model **63**
4.1 Motivations . . . 63
4.2 3D scene rendering on paper . . . 63
4.2.1 Problem . . . 63
4.2.2 Solution . . . 64
4.2.3 Limitations . . . 64
4.3 Interaction . . . 66
4.3.1 Tangible interface . . . 66
4.3.2 Touch interface . . . 68
4.3.3 General public use . . . 69
4.4 Shading for Drawing . . . 70
4.4.1 Simplification . . . 70
4.4.2 Composition . . . 71
4.5 Stereographic drawing . . . 71
4.5.1 Introduction . . . 71
4.5.2 Existing techniques . . . 72
4.5.3 Proposed system. . . . 72
4.5.4 Perspectives . . . 75

III Beyond Drawing **77**

5 Usage and Applications **79**
5.1 Palais de la Découverte . . . 79
5.2 Rencontres Inria Industries . . . 80
5.3 Semaine digitale . . . 81
5.4 Cap Sciences . . . 82
5.5 Marbotic . . . 84

6 Towards Mixed Media Creation **87**
6.1 Creative coding . . . 87
6.1.1 Processing . . . 87
6.1.2 Code for the paper . . . 88
6.1.3 Ideas of applications . . . 90
6.1.4 Implications on creativity . . . 90
6.2 Image acquisition for drawing . . . 90
6.2.1 Animation . . . 90
6.2.2 Proof-of-concept implementation . . . 90
6.2.3 Physical/Digital image composition . . . 91

Bibliography **97**

Introduction

"A journey of a thousand miles begins with a single step."

Lao-tzu, The Way of Lao-tzu.

Technology is an important part of our lives nowadays: we use the term ubiquitous computing to describe it. Ubiquitous means "found everywhere", in fact mobile phones are available worldwide, even in places without electricity. In our developed countries it is even more important with computers, cars, smart phones everywhere.

Ubiquitous computing is a strong word to describe our current digital uses. It is possible to access data from anywhere from the so called "cloud computing" but this information is always confined in a device. Today's possibilities would be better described by "accessible with any device" computing. In this thesis, we focus on how to get rid of the "device", at least from a user perspective. This is a new goal of Augmented Reality (AR): the historical AR goals were on information visualization on top of the physical world. Nowadays, we go towards the creation of "organic interfaces": interfaces that can take any shape, and more likely the shape of everyday objects.

There are many steps towards an "augmented world" in which the physical and digital elements mix seamlessly together. The first steps were paved by researchers in Computer Graphics, Computer Vision and Human Computer Interaction (HCI) for more than twenty years. Today's challenges are to understand to which extent the digital world fits the physical world, and consequently how to make the digital world fit the physical world. Our approach is to "augment" the physical world. We start from an existing task: the creation of physical visual art (drawing, painting...) and we extend its possibilities though augmented reality. In the first part of this document, we describe how our approach is inscribed in the related work in augmented reality. Additionally, we describe our interactive projection system that enable us to augment reality with digital elements.

We explored how we could enable the creation of physical drawings or paintings using computer graphics. We inserted an important constraint: the edition and manipulation of the augmentation are to be done directly on the drawing medium. We want to keep a tight link between the digital and physical elements so that everything could be manipulated using

traditional drawing tools and simple interactions. An example is our solution which enables the manipulation and lighting of a 3D scene using our touch and 3D pointing interface.

In this work, we want to provide new technologies with as less as possible new techniques to learn for the user. Although it is a dream of HCI researchers to create "natural", easy to use and efficient interfaces, in our case it is also a constraint due to the pre-existing physicality of the interface. Most of office workers work on a physical desktop and paper independently of their work: artists, engineers, CEOs it is all the same basic elements. We learned to use desktop space since our childhood and its use and organization is different from one person to another. Consequently, an easy customization of the mixed physical/digital world is crucial. We explored this for drawing as described in chapter 6. This raises a larger question: "How to program our everyday objects ?". We propose some explorations on programming sheets of paper but this question will be a challenge for the next decades.

Organization of this work

This thesis is divided in three different parts, each part contains two chapters. The organization is illustrated in the figure at the end of this introduction. The first chapter is "Previous and Related Works", it describes the inspiring related works in research and the current work related to this thesis. The second chapter is the technical chapter: its purpose is the replication of a similar system. We explain how to use a projector and a camera to do Spatial Augmented Reality, and how to calibrate the projection with a depth camera to make it interactive. In the following chapters some variations of the system are introduced and we consider that from the basic system, the extensions are easy to implement.

The second part is made of the two chapters on physical drawing tools. These chapters are the mature part of this thesis, they have been tested and evaluated. Chapter 3 describes a construction line tool that enables the projection of lines for the creation of drawings. In chapter 4, we describe our 3D rendering on paper and its applications for drawings as well as stereoscopic drawings.

The third part starts with a description of the main public exhibitions and the industrial applications of this thesis. It continues with a discussion on the creation of spatial augmented reality interfaces for rich applications such as a mixed reality drawing application. We also show some of the possibilities of the Software Development Kit (SDK) we developed for creative coding on sheets of paper. Finally, we demonstrate some applications of the capture/re-projection process to create animations.

Notes about this thesis

This thesis contains many figures and illustrations, each illustration complements the text and it is a way for me to transmit more precisely my intent. Chapter 4 requires 3D anaglyph glasses and we recommend to use good quality glasses such as the 3D Vision Discover from NVIDIA. Each chapter is relatively independent, with its own bibliography, motivations and perspectives. We made this choice to make the thesis easier to read, write and reuse. Although it is the ambition of every thesis to be read, here we enforce it with the English language. We do

this despite the history of rivalry between France and England, and the imperialist explanation of the strength of the English language nowadays. It is important to accept the reality that English is the current international language by far, and french may never come back as a major international language. Finally, please understand that my favorite choice would have been to write this in thesis in the "international language" (also called Esperanto) described more than 100 years ago by Dr. Zamenhof. Ethical and political progress is slow compared to science and technology; we can either choose to go towards what we believe will be the future or wait for it to happen.

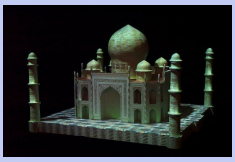
Organization of this thesis

PART I : Augmented Reality

Chapter 1

Previous and Related Work

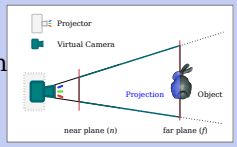
Spatial Augmented Reality
Tangible Interfaces
Camera-based Tracking



Chapter 2

Spatial Augmented Reality System

Image formation model
Camera-Projector calibration
Multi-touch with Kinect



PART II : Augmented Drawing

Chapter 3

Augmented Drawing Applications

Projection of photos
Digital construction lines
User evaluation

Chapter 4

Drawing of a 3D Digital Model

3D Rendering for Paper
Touch and Hovering
Stereoscopic Drawings

PART III : Beyond Drawing

Chapter 5

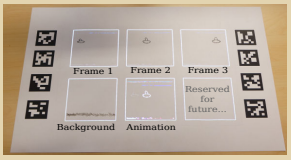
Usage and Applications

Palais de la Découverte
Cap Sciences
Marbotic

Chapter 6

Mixed Media Creation

Creative Coding
Drawing acquisition
Animation



Part I

Augmented Reality research

Chapter 1

Augmented Reality

Augmented reality (AR) started as a sub-field of *Virtual Reality* (VR) which started in the 70s with the beginning of computer graphics. VR was enabled by head worn devices, called *Head Mounted Displays* (HMD), created by Ivan Sutherland [Sutherland, 1968]. They were heavy devices placed on the user's head enabling the vision of virtual object on top of the physical world.

1.1 Vocabulary

The vocabulary evolved from the oxymoron *Virtual Reality* to the easier to understand term *Mixed Reality* (MR). This term was popularized by Milgram *et. al.* [Milgram and Kishino, 1994] and the Reality Virtuality contiuum, illustrated in Figure 1.1. The idea of Mixed Reality was quite a good definition at that time. However, nowadays it is more difficult to identify a reality distinct from the digital world, we call this "ubiquitous computing" or "ambient intelligence" as stated in the introduction. Let us take a simple example of ubiquity: Twitter. Is Twitter on a phone, tablet, TV, PC, (printed) on paper ? It is everywhere at the same time, and certainly real. Consequently, the virtual environment in opposition to a real environment does not apply anymore. The continuum view still applies, we propose a different vocabulary and slightly different definition in this thesis.

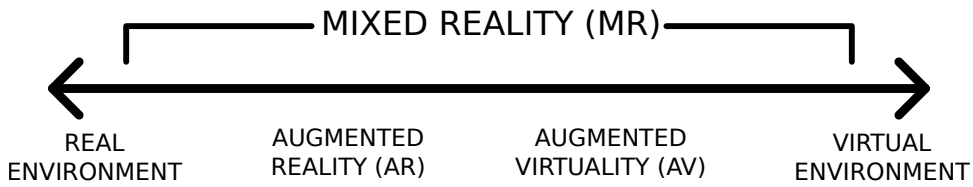


Figure 1.1: Reality-Virtuality continuum by Milgram and Kishino in 1994.

Real Environment is favored to *Physical Environment* or *Physical World*, and *Virtual Environment* is favored to *Digital World* or *Digital Environment*. This vocabulary shift is done to understand that the digital world is indeed part of the real world. The difference between physical and digital is at the heart of this work because we propose to create *physical drawings* using *digital tools*. Consequently, our contiuum expands from purely digital elements such as

a description of a 3D object, to physical element such as print of this 3D object. We consider that all the technology involved between these two sides are mixed reality. In chapter 6 we propose mixed creations that merges physical and digital elements, these creations fit to the whole Mixed Reality Continuum because part of it is physical, and part of it digital and the complete creation is both purely digital and purely physical.

Figure 1.2: Physical-Digital continuum used in this thesis.

1.2 Definition

> *"Augmented Reality lets people use familiar, everyday objects in ordinary ways. The difference is that these objects* also *provide a link to a computer network."*
>
> From *Augmented Reality: Linking real and virtual worlds. A new Paradigm for interacting with computers* [Mackay, 1998].

This definition of augmented reality is our goal in this work. It is in opposition with traditional computing, where the application is contained within the device and controlled by dedicated hardware. The idea of Augmented Reality is to merge the digital world with the physical world. It is quite easy to imagine a perfect way to add information to either though holographic projection as a "shared" display or retinal display for "individual" use as illustrated in Figure 1.3. In contrast the perfect interface is a more complex question and it may never get a better answer than: "an interface that behaves how it is expected to behave".

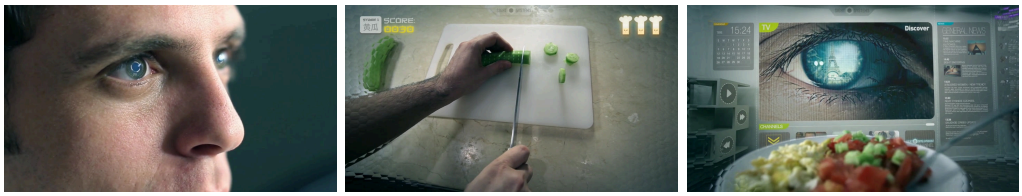
The ultimate augmented reality interface would make any everyday object interactive and acting as display. The final step of sensing and display technology was imagined by Sutherland in 1965 [Sutherland, 1965] with his "Ultimate Display" where the physical geometry and appearance of objects could also be programmed.

1.3 Visualization

1.3.1 See through

The visualization of the *augmentations* can take different forms. The most common in industry is *See-Through Augmented Reality* (STAR): the digital elements are seen through a screen, as illustrated in Figure 1.4 left. The screen could be a Head Mounted Display (HMD), or any

Augmented Reality environment. Left: no augmentation, right: Animated textures on walls.
From **Cloud Atlas** written, produced and directed by
Lana Wachowski, Tom Tykwer and Andy Wachowski.



Augmented Reality contact lenses.
From the short movie **Sight** by Eran May-raz and Daniel Lazo

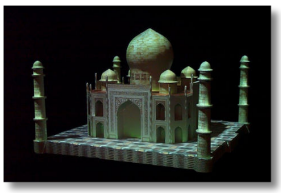
Figure 1.3: Artist's vision of Augmented Reality, on digital texturing (top) and retinal displays (bottom).

kind of screen. STAR has the advantages of the current screen technologies: high contrast and resolution, and possibility to make assumptions about the user's head position such as facing the screen. It has limitations on the number of users, e.g.: each user must have their own HMD. Moreover, for mobile devices the user has to carry it which obstacle the interaction space and it can be tiring over time depending on the device. It is also possible to combine a global display and a local display for each user, this was recently demonstrated by [Spindler et al., 2012]. In their approach they use sheets of paper as mobile screen to limit fatigue. This application was not demonstrated explicitly for AR purposes but it is also relevant for AR.

See-through Augmented Reality

Marker tracking and HMD calibration for a video-based augmented reality conferencing system
[Kato and Billinghurst, 1999]

Physical Augmentation



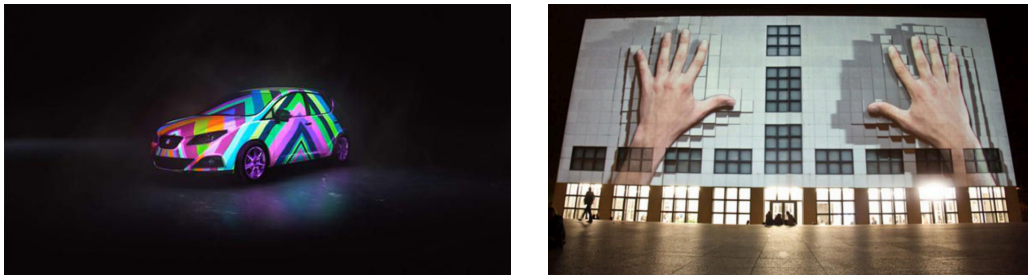
Shader lamps: Animating real objects
with image-based illumination
[Raskar et al., 2001]

Figure 1.4: Augmented Reality visualization technologies.

1.3.2 Spatial

The second visualization technology is based on projection. This is called *Spatial Augmented Reality* (SAR). The term and technology was first created by Raskar *et. al.* in 1998. The book he wrote with Bimber is the current reference for SAR [Bimber et al., 2005]. The Taj Mahal with a "virtual" (projected) illumination demonstrated the potential of the technology for visualization, as seen in Figure 1.4 right.

SAR for digital illumination is getting more and more popular in the entertainment industry, it is called *Projection Mapping* (Figure 1.5). It consists in projection on buildings at night, or objects made for the projection experience.



(a) Texturing a car (Seat commercial 2010)

(b) Digital elements and texturing a building 555 KubiK by Urban Screen (2009)

Figure 1.5: Projection Mapping enables digital texturing (a) and the projection of digital elements that provide geometry changing effects (b).

It is possible to separate the rendering in two complementary types. The first one is the projection of textures and appearances. The projected elements fits the physical objects. Consequently, the projection is correct from any point of view (see Figure 1.5 (a)).

The second type of rendering is the projection of digital elements on top of physical ones. These digital elements are rendered for a given point of view. For building projection, as illustrated in Figure 1.5 (b) this point of view can be large enough for multiple viewers. However, for small scale projection like in the Beamatron [Wilson et al., 2012] (Figure 1.8 bottom) the user's point of view is crucial.

In order to achieve the best results for rendering digital objects the physical geometry must be suitable for projection. Moreover, stereoscopic visualization enhances the depth perception of the digital elements [Benko et al., 2012] but makes the experience available to only one user. It is illustrated in Figure 1.6. The possibility to project digital objects for more than two points of view is the main limitation of the current projection technology. The two points of view are currently available using active shutter glasses in addition to polarized light projection.

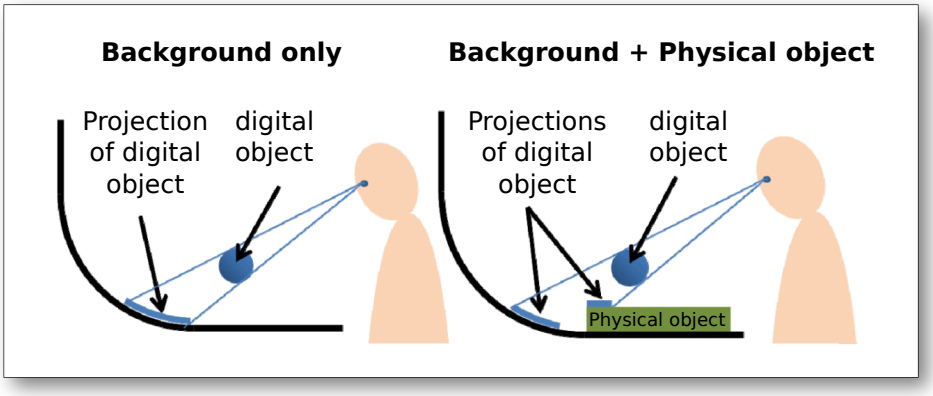


Figure 1.6: Projective texturing requires that the system takes into account the geometry of physical (real) objects on the tabletop at every frame in order to correctly present perspective 3D digital (virtual) information to the user's eye. Adapted from [Benko et al., 2012].

1.4 Tangible Interfaces

1.4.1 Definition

Tangible User Interfaces (TUI) constitutes a major part of the augmented reality interfaces. The word "tangible" means *Perceptible by touch*, it refers to interfaces where the touch capability has a meaning for the application. It is generally opposed to standard interface like mice. Whereas pointing devices are generally indirect, tangible interfaces provide a direct control on the objects. In the early works of TUI, it was called "Graspable Interfaces" in 1995 by Fitzmaurice [Fitzmaurice et al., 1995]; followed by a more comprehensive vision by Ishii *et. al.* in their view of *Tangible Bits*. Nowadays, with not only graspable devices but also actuated devices, the predicted next generations of devices would be "radical atoms" [Ishii et al., 2012]. Their vision is of the future of interfaces is though the embodiment of digital elements: "all digital information has physical manifestation" (Figure 1.7). In addition to the technological challenge of tangible bits and radical atoms, the physical representation of digital information is challenging. The main power of computing is its ability for dealing with large amount of data, and it is clearly the main limitation of the physical world and physical representation of information. This problematic is discussed in chapter 4, more specifically the question of creating menus or dedicated tangible elements and in which case.

This section is called "Tangible Interfaces" because the spatial intangible interactions constitute a small portion of the AR interfaces. In real life we rarely perform hand or body mid-air gestures for other activities than communication. Moreover, the description of tangible interfaces is complete and comprehensive. We can easily fit the other descriptions of AR user interfaces within it such as spatial and organic user interfaces. *Spatial User Interfaces* (SUI) is the set of interfaces that uses the 3 dimensions to control applications, tangible or not. *Organic User Interfaces* [Holman and Vertegaal, 2008]are the interfaces displayed on screens or objects that are not planar, i.e, curved, spherical, object, or even living beings. All these interfaces are subsets of the Augmented Reality interfaces possibilities, which we call Tangible Interfaces in this thesis.

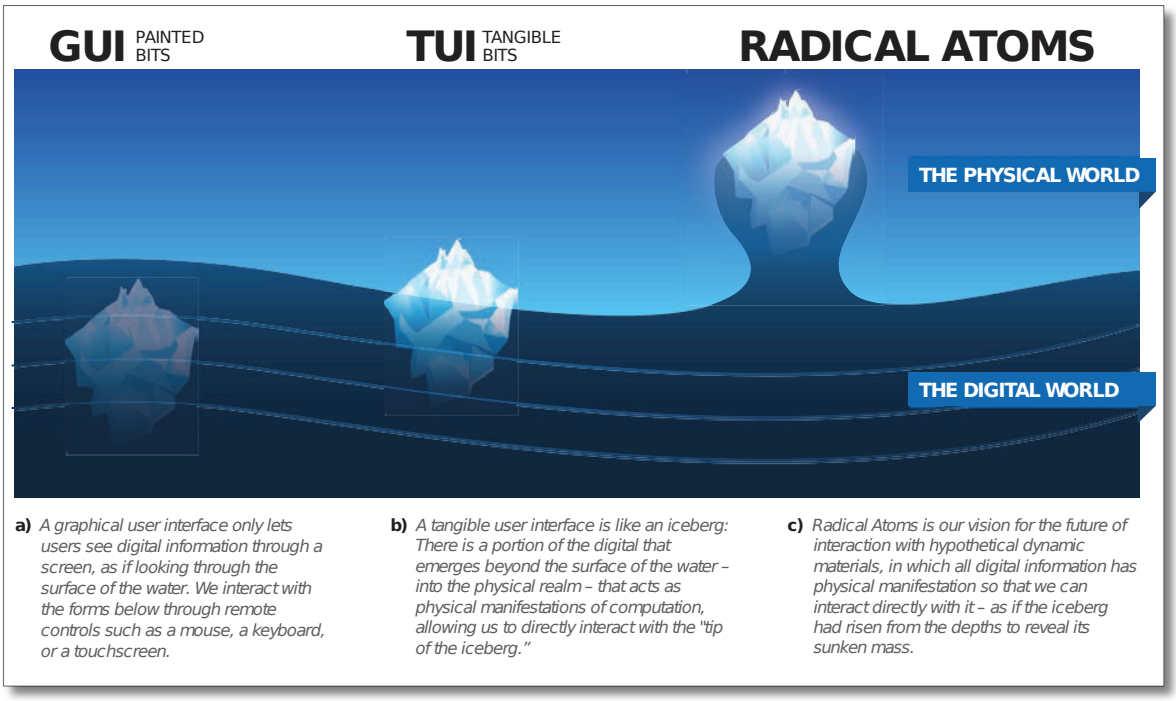


Figure 1.7: Iceberg metaphor—from (a) GUI (painted bits) to (b) TUI (tangible bits) to (c) Radical Atoms. From [Ishii et al., 2012].

1.4.2 Tracking techniques

Interactive AR and SAR firstly used the tracking techniques from Virtual Reality. There are three possibilities for tracking, it can be carried by the user, on the object to track, or in the environment (around the user and/or object). This is illustrated in Figure 1.8.

User Instrumentation

User instrumentation enable to use the system in many context. An interesting perspective on this is the "Sixth Sense" developed at MIT [Mistry and Maes, 2009]. They show many applications of embedded AR using projection. Another important example of embedded AR is the soon to come *Google Glasses*[1]. They imagine the same overlay of information for an individual use.

Object Instrumentation

The second possibility is to modify the objects to make them interactive. This is generally done using conductivity, as in Touché by Sato, Poupyrev and Harrison [Sato et al., 2012]. Their instrumentation is non intrusive and they use "swept frequency capacitive sensing" coupled with machine learning to enable the detection of the user's type of contact on many objects.

[1] http://www.google.fr/glass/start/what-it-does/

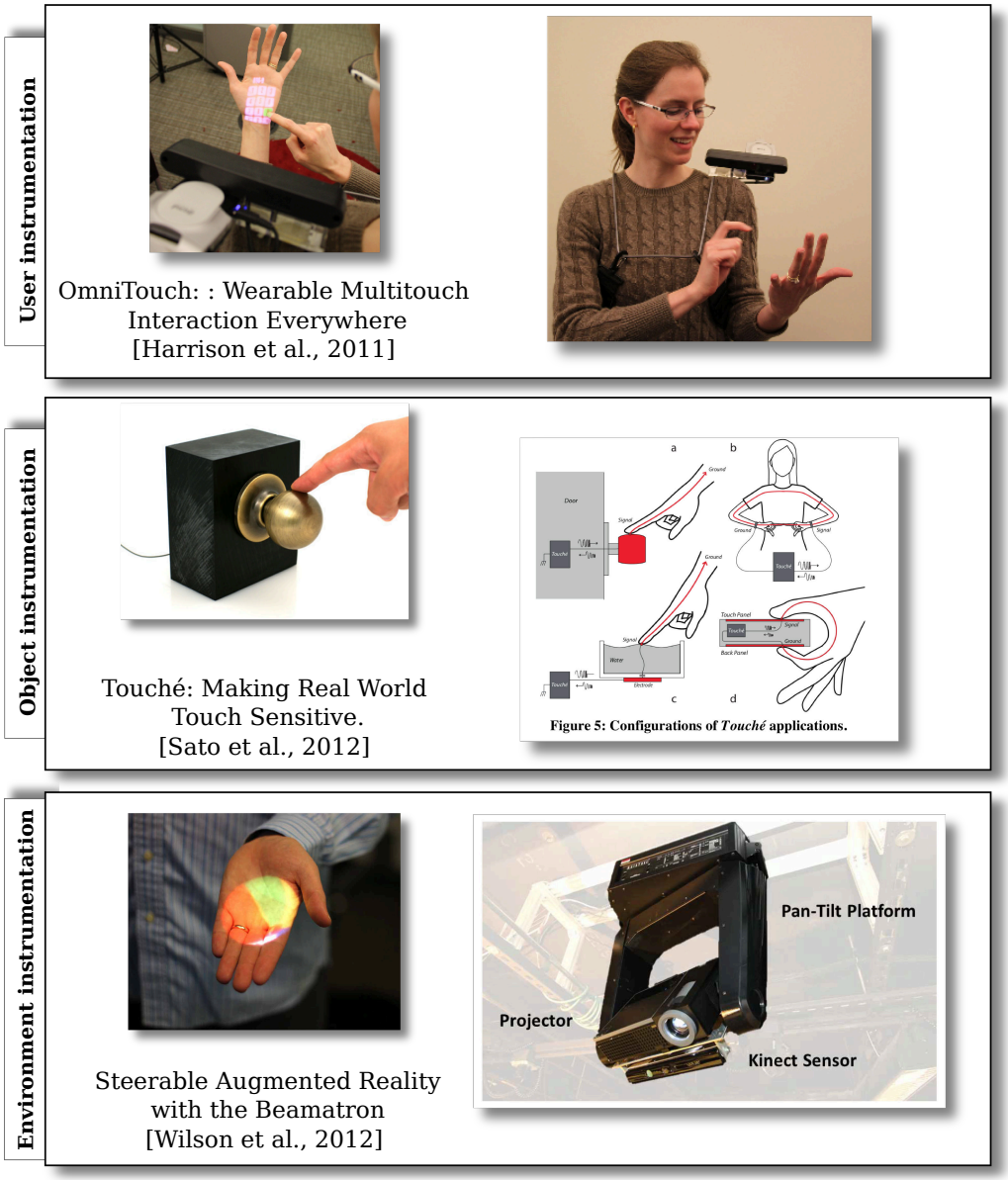


Figure 1.8: Augmented Reality sensing technologies.

Environment Instrumentation

The last possibility is instrumentation of the environment, this is common in research and the industry with infrared cameras tracking. For a few years now, environment instrumentation is available to the general public with the Kinect. When coupled with a projector it is very suited for Spatial Augmented Reality. An example of advanced environment instrumentation is the "Beamatron" [Wilson et al., 2012] where the sensing and projection is mounted on tilt/pan motors. Another important work on is the "Humantenna" [Cohn et al., 2012], it is neither user, nor environment instrumentation because it uses the existing electromagnetic noise from the environment to detect the user's pose and gestures.

1.5 Paper Interfaces

Paper is present everywhere in our lives. Paper was a major invention that lead to our current civilization and it is still extremely useful. Paper has been here for a long time and it may remain for a long time in our homes and offices [Sellen and Harper, 2003]. As the technology goes forward with advanced electronics the technological advances of paper over electronics are countless: it is lightweight, cheap, foldable and so on. However, if we compare paper documents and digital documents the digital version have many advantages: it does not occupy much space, it is easy to duplicate and share. With shared online documents, it is even better than physical paper for collaboration. The possibility to use physical paper to manipulate digital was originally described by Wellner in his seminal work the DigitalDesk [Wellner, 1993]. In this work, he proposes to augment the paper using projection and he created the basics for augmented paper with pen and touch interface and paper document recognition. Additionally he implemented and conceptualized a wide range of augmented paper applications such as a desktop digital calculator, translator, painting and collaborating applications with multiple digital desks.

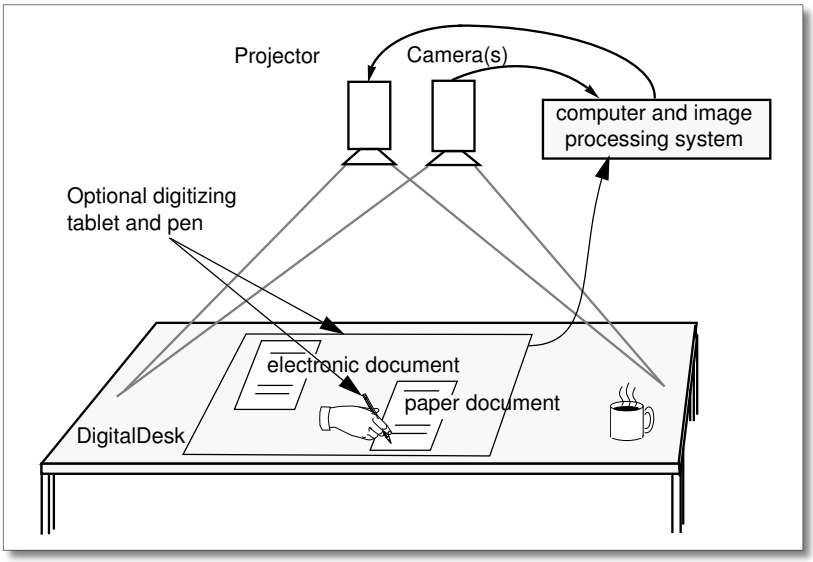


Figure 1.9: Schematic diagram of a DigitalDesk, in [Wellner, 1993]. Note that there is a cup of hot coffee on the table. It also present in two other illustrations and photos in the DigitalDesk article. It conveys the idea of augmented reality: the DigitalDesk is still a desk.

1.5.1 Augmented Documents

Augmented documents research have two research directions. The first one is the augmented reality approach: the physical documents are empowered with digital processing. An example is the annotation of paper; a paper document is printed, annotated using a pen and these annotations are captured and integrated in the digital document [Guimbretière, 2003]. The printed paper becomes an input for the digital world and the interface and interaction stays the same as usual.

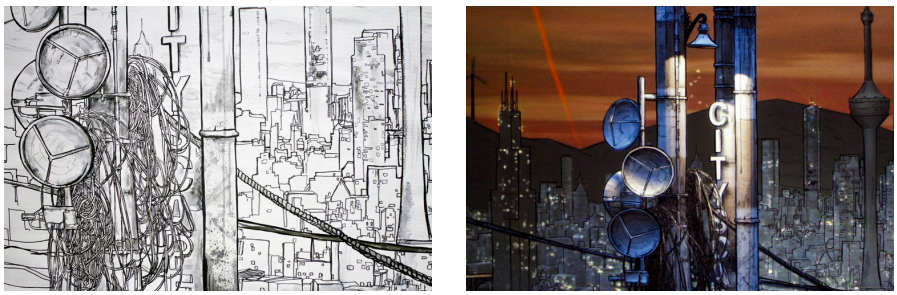
The second research direction is the use of the physical properties of paper for computing. In Paper Windows [Holman et al., 2005] Holman and colleagues propose a few interaction metaphors to interact with a *Graphical User Interface* (GUI) on paper. It is possible to interact with the paper directly, the hand on paper, and a stylus. Recently, Steimle and colleagues extended the interaction possibilities by modeling the deformations of the sheet of paper in Flexpad [Steimle et al., 2013]. The sheet of paper they use is either flexible (like usual paper) or shape-retaining. Each type of paper have its own physical capabilities which have implications for the application, *e.g.*, the shape-retaining paper can be held with one hand and not change shape, which have no consequence for the application.

1.5.2 Augmented art

In Projector-guided painting [Flagg and Rehg, 2006], Flagg and colleagues created a system dedicated to painting. They proposed solutions to assist novices during the whole painting process for: color mixing, brush stroke orientation, and layering of the painting. They demonstrated that their system successfully enabled novices to create high quality paintings.

Kim and colleagues recently proposed new interaction techniques to interact with projection for drawing in "AR pen and hand gestures: a new tool for pen drawings" [Kim et al., 2013]. They enable digital drawing, copy/paste operations and various guides to create physical drawings.

Mixed creations have already been created by a few artists. An example is the IceBook by David & Kristin McGuire [2]: they use the possibilities offered by cuts of paper to create visually rich dioramas. The projection on top of it consist mainly in videos and digital illumination. The use of physical elements bring more contrast and depth to the creations. Another remarkable example is Cityscape 2095 by Antivj [3] illustrated in Figure 1.10. The object could be self-sufficient to show the artist's idea and benefit from projection and interactivity to achieve its full potential.



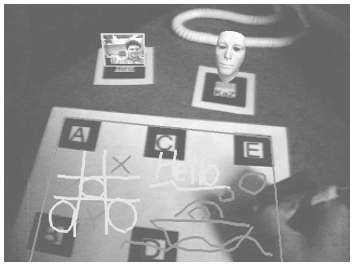
Painting. Projection on painting.

Figure 1.10: Mixed reality painting created by the antivj group. The projection is animated to display the life in the city and weather effects.

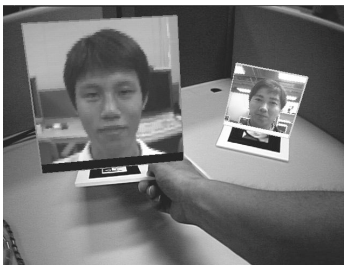
Another work is described in the introduction of chapter 6 and the chapter is dedicated to the creation of augmented art.

[2] http://www.davyandkristinmcguire.com/
[3] http://www.antivj.com/cityscape/

Using the Augmented Reality Interface.



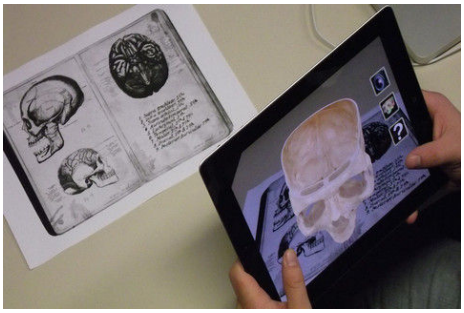
Virtual Shared White Board.



Remote user representation in the AR interface.

Marker tracking and HMD calibration for a video-based augmented reality conferencing system
[Kato and Billinghurst, 1999]

Wikitude World Browser on the iPhone 3GS uses GPS and a solid state compass (2009)



App iSkull, an augmented human skull for education (iOS OS) (2012)

Nintendo 3DS AR Cards (2011)
image from Nintendo website

PS Vita augmented Reality (2012)
image from Playstation website

Shape Recognition and Pose Estimation for Mobile Augmented Reality
[Hagbi et al., 2009]

Figure 1.11: Augmented Reality knew a big growth since the first marker-based applications (top). It is now on most mobile devices: phones, tablets, video games consoles (middle). The next steps of mobile AR enable a large variety of marker detection and tracking in real-time (bottom).

1.6 General public

Augmented reality has now reached the general public. The term in known by an important portion of the population[4] and there are many augmented reality applications on mobile devices, as illustrated in Figure 1.11. The first mobile applications were for navigation, they enable the visualization of points of interest though the phone's camera. In order to do this they use the compass, GPS, gyroscope and accelerometer embedded in the phone. The first application was Wikitude[5] released in late 2008. Another example is AR applications using dedicated "AR cards". It is derived from the first AR works with fiducial tracking [Kato and Billinghurst, 1999] (Figure 1.11, top part).

The expression "augmented reality applications" is excessive because even though the graphics show elements displayed on a table, in the physical world. The applications are video games or visualizations that do not complete and "augment" the physical world; they just provide graphics that takes place on the physical world and seen through a screen.

The limitations of the general public applications are linked to the limitations of the tracking performance and possibilities. More recent work, such as the one illustrated in Figure 1.11, bottom part by Hagbi and colleagues [Hagbi et al., 2009] enable the tracking of "natural" and drawn shapes. This kind of innovation will enable more versatile AR applications that do not the user to carry fiducial cards with him or her. Mobile augmented reality is a young field and it will grow rapidly only when it goes further than video games and start propose applications that are produce more result on the physical world.

1.7 Contributions

Augmented Reality research spreads in many fields. In order to better understand our contributions we represent them by chapter in the "Building blocks for augmented reality" figure proposed by Bimber, Raskar and Inami in their book "Spatial Augmented Reality" [Bimber et al., 2005] (Figure 1.12). Chapter 2 describes our system with hand and paper tracking, a precise projected display and the interactive possibilities offered by this system. Chapters 3, 4 and 6 describe our augmented and mixed drawing applications which lead to the design of dedicated interaction metaphors and rendering techniques. The drawing tools have been tested though a controlled user study described in chapter 3.

1.8 Future of Augmented Reality

As we stated before SAR is already in the industry through projection mapping exhibitions. The rendering technology are evolving quickly, the screens did a huge leap in the past 10 years to achieve an ideal resolution for some displays: higher than what the eye can perceive. In the same way, the projectors are getting smaller and cheaper every year. We can now separate three kinds of projectors (see Figure 1.13), sorted by size. The standard projectors are large, heavy (> 3kg) and noisy and provide an intense luminosity. However, the "new" projectors are smaller,

[4] Counting from french people who are science oriented, about a third knew the term from our estimation.
[5] http://www.wikitude.com

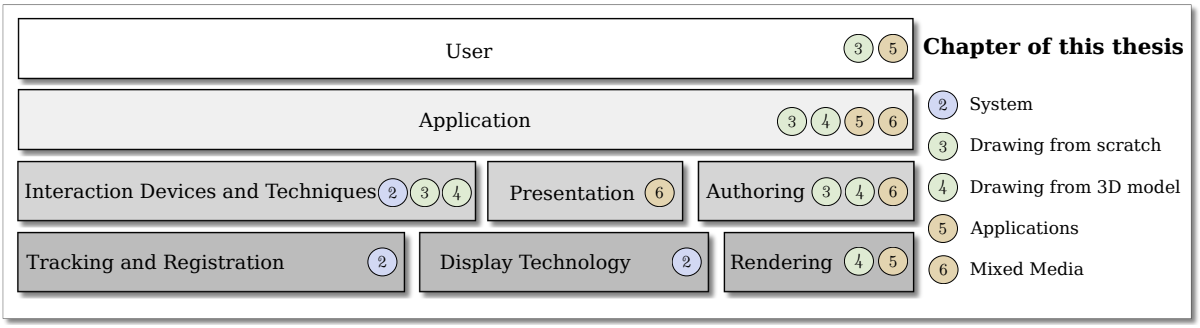


Figure 1.12: Building blocks for augmented reality from [Bimber et al., 2005] with the contributions of each chapter in the blocks.

featuring lower intensities they are easier to manipulate, and embeddable on other devices for the pico projectors. The first projector phones were created in 2010, nowadays projector for mobile phones propose low resolutions (320x240 px) and intensities (15 lumens). Even if the intensities stay low due to battery limitations, the resolutions will go up. The intensity limitations could be overcome with the additive nature of projection, by coupling multiple projector phones larger and brighter images could be achieved.

Type	**Intensity (lumens)**	**Usage**	**Price**
Standard	2500 - 5000	ligths on	300 - 2000 €
Mini	250 - 500	ligths off	500 - 2000 €
Pico	15 - 100	complete dark	80 - 150 €

Figure 1.13: Consumer level projection technologies.

While SAR seems to have a bright future, other VR technologies are getting consumer-ready. With the Google Glass and Occulus Rift, personal VR and AR are reaching the general public. These possible future are explored by artists and the images in Figure 1.3 may be available in a not so distant future.

Chapter 2

SAR system

In this chapter we describe our spatial augmented reality system. We firstly state the constraints for the design of a *Spatial Augmented Reality* (SAR) system for physical drawing. After relating the existing techniques we propose a technical description of the SAR system. This technical chapter is aimed at students and researchers that have the desire to create their own interactive SAR system. Both projection and tracking techniques have been successfully implemented by Master's students in programming or research projects. The implementation of such a system is not difficult *per se* for students in computer science. However the description of the underlying techniques from computer vision and computer graphics are rarely described together.

Indeed, HCI (system) papers usually focus on the system possibilities, and eventually how it performs compared to other equivalent systems. The technical implementation is generally not precisely described, because they are not relevant in the field of HCI. It is a pity for the sake of replication, especially for a replication by inexperienced researchers. This chapter requires some experience with computer graphics and computer vision and it sums up the technical experience required for the replication of our SAR system.

2.1 System Design

The system is for augmented reality, which means that we begin with the drawing task and we expand its possibilities. We do not want to hinder the artist's possibilities, we want to enhance them. Consequently, our system must enable any type of drawing medium: any color, size and texture. Thus, we opted for top projection: a bottom projection would impose transparent or semi-transparent medium.

The next question is how to interact on the piece of paper. A simple answer is to provide a pen dedicated to digital elements, in addition to the usual pens, colored pens and pencils. However, current interactive pen technologies have scalability issues, though their use for mixed creation (chapter 6) were promising. Finally, we opted for a touch input using finger detection on the table: the user does not require specific tool, nor hardware to interact with the system.

2.1.1 Drawing medium tracking

Object tracking is the estimation of an object's position. It is associated with pose estimation, the pose being the orientation of the object. For tracking and pose, the quality is evaluated by three criterion: latency, precision, and tracking rate. Tracking and registration is an extensively researched problem and solutions come at a wide variety of complexity, computational requirements and equipment.

The most common tracking techniques nowadays are vision-based trackers. The main limitation of cameras is occlusions, generally the occlusion problem is compensated by adding more cameras and maintaining a robust model of the objects tracked. In contrast, electromagnetic tracking enables precise location and pose estimation and the interference come only from the metallic objects nearby if any. Electromagnetic trackers are more expensive, they have a lower range of motion and require an important instrumentation. Another possibly is to use physical tracking: with strings attached to the drawing medium. This kind of instrumentation have scaling issues and provide fewer degrees of freedom, but the tracking frequency and precision are high.

Finally, we opted for vision-based tracking for its scalability and affordable prices.

2.1.2 Vision-based tracking

Object tracking using computer vision provides a large range of solutions, depending on the objects to track. In "Object Tracking: Survey" [Yilmaz et al., 2006], Yilmaz, Javed and Shah describe the general solutions to detect shapes and features, cluster elements, and create models for recognition. From this general approach we need to use techniques that estimate the position and orientation of an element. The general techniques for pose estimation of a known object in augmented reality is the use of fiducial markers, natural features or model fitting as described by Leptit and Fua in "Monocular Model-Based 3D Tracking of Rigid Objects: A Survey" [Lepetit and Fua, 2005].

We need to track paper, colored paper and canvas indiscriminately. It is complicated to create a model for each drawing medium. Consequently, we opted for fiducial tracking: a fiducial marker is added next to the canvas or paper to identify its position and orientation. In order to achieve this, we use the library ARToolKitPlus [Wagner and Schmalstieg, 2007] which enables small markers and multi-marker detection. It is based on Kato and Billinghurst's work on marker tracking for Head Mounted Displays [Kato and Billinghurst, 1999]. The main limitation of this kind of detection is the normal estimation of the marker. Indeed, the number of features tracked is low in one marker (4 corners). In order to limit this effect we place markers at each corners of the tracked canvas and all around for occlusion robustness, as illustrated in Figure 2.1.

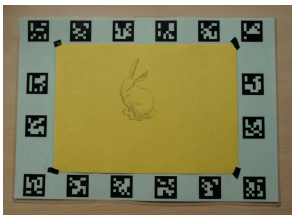


Figure 2.1: ARtoolKitPlus fiducial markers around a yellow piece of paper.

This tracking technique is "old-fashioned" already, the ideal would be tracking the drawing canvas and maintain an incremental model of the "canvas + drawing" evolving over time using natural features. A first step toward this is initiated by Audet *et. al.* in [Audet et al., 2010] :

they enable tracking of a planar surface knowing the projected elements and the printed elements on it.

Standard computer vision techniques provide a good and reliable solution for planar surface tracking. However, they are generally not suited for hand tracking without hand or finger instrumentation using colored gloves [Wang and Popović, 2009]. It is also possible to detect one hand's pose using wrist-mounted apparels as in Digits by Kim *et. al.* [Kim et al., 2012], but it is not suited for touch interactions and require calibration for each user. Usually the color camera impose predefined posture to detect touch and touch precision is limited as in [Audet et al., 2012]. Nowadays, touch detection is done using depth cameras, in the next section we describe the related works in touch detection using depth cameras.

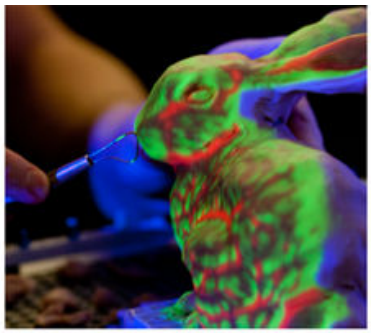
2.1.3 Touch detection

Depth cameras such as the Microsoft® Kinect® provide a new set of solutions for computer vision problems that are considered as difficult for standard color camera approaches. The main technologies for sensing depth are structured light scanners (SL) (Kinect, Asus Xtion) time-of-flight cameras (TOF) (Kinect 2) and stereoscopic vision (SV) (LEAP Motion). In one hand the SL and TOF cameras rely on the projection of light and can have trouble dealing with ambient illumination such as sunlight. On the other hand SV cameras achieve generally lower resolutions, with the exception of the LEAP motion that also projects light to achieve their high speed and quality results.

Depth sensing has been proved to be suited for touch detection on most surfaces since 2007 by Wilson [Wilson, 2007]. Since then, the sensing technology using depth cameras has known a "golden age" on developing touch interfaces in a wide variety of setups:

- Table, hand and wall data manipulation [Wilson and Benko, 2010]
- Hand and arm, surfaces and walls interactions [Harrison et al., 2011], Figure 2.2 (middle).
- Hand reconstruction to interact with virtual objects [Hilliges et al., 2012], and distant collaboration [Benko et al., 2012]
- Hand guidance for rehabilitation [Sodhi et al., 2012]
- Water-based interfaces (in bath) [Takahashi et al., 2012], Figure 2.2 (bottom).
- Physical sculpture [Rivers et al., 2012]. Figure 2.2 (top).
- Object Reconstruction and tracking [Izadi et al., 2011]
- Room Reconstruction - room size interactive surface [Molyneaux et al., 2012]
- Behind a transparent screen [Lee et al., 2013]
- Semi-rigid sheets of paper [Steimle et al., 2013]

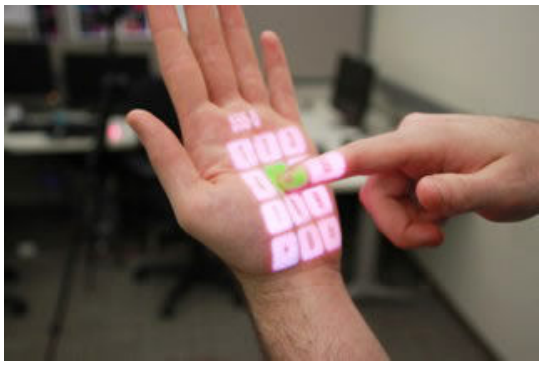
Target 3D model

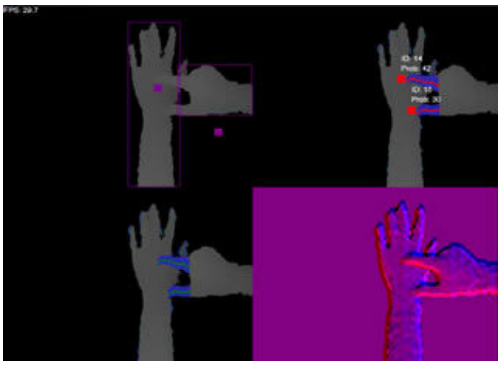


Guidance projected onto material

Sculpted physical replica

Sculpting by Numbers
[Rivers et al., 2012]





Omnitouch : Wearable Multitouch Interaction Everywhere
[Harrison et al., 2011]

Usage scenario

Fluid Space : Interactive Water Surface Display for Viewing Information in a Bathroom
[Takahashi et al., 2012]

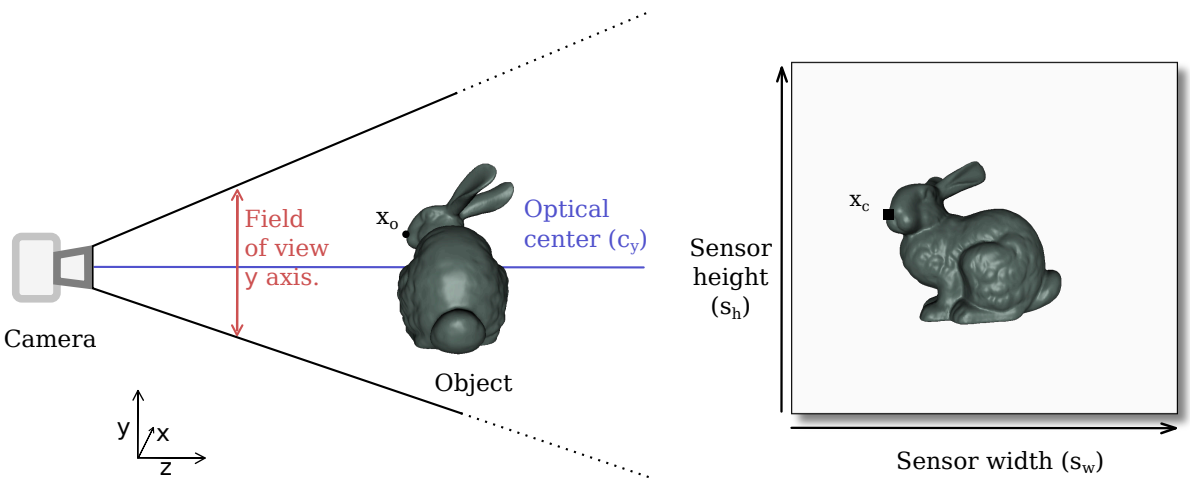
Figure 2.2: Depth cameras enable a wide range of new applications. The real-time capture of shape enables projection on clay for sculpture (top). Embedded on the user, a depth camera and projector makes the hands, and arm interactive, as well as walls or any planar surface (middle). It can also make interactive surfaces that can hardy be instrumented such as water (bottom).

2.2 Spatial Augmented Reality

In this section we introduce briefly how cameras and projection work according to the pinhole camera model. It is followed by a description on how to use the camera parameters to create an OpenGL 3D scene mimicking the camera (or projector). For more information, we invite to reader to refer to "Computer vision: a modern approach" [Ponce et al., 2011], or "Multiple view geometry in computer vision" [Hartley, 2008] for the computer vision part and the OpenGL documentation for the rendering part.

2.2.1 Image formation model

The image acquired by the camera is a projection of the observed elements. Figure 2.3 illustrates how an observed object point (x_o) is projected on one or multiple pixels in the camera image (x_c). This projection is controlled by the internal parameters of the camera (intrinsics) such as the field of view. It is called the camera matrix (K_c) ; note that we will also refer as camera matrix for the projector. A projector is often refereed as an "inverse camera". The image formation model could be simplified to equations (2.1) and (2.2).



(a) The camera observes an object.

(b) Resulting image.

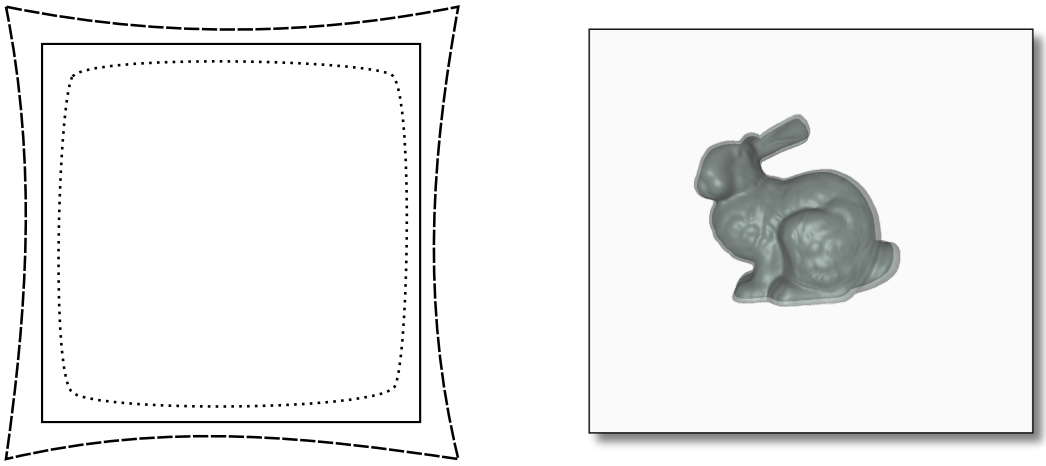
Figure 2.3: Image formation model.

$$x_c = K_c * x_o \tag{2.1}$$

$$K_c = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \tag{2.2}$$

The matrix K_c is the projection matrix, f_x and f_y represent the focal length (and field of view) of the camera and c_x and c_y its optical center. Ideally the optical center is in the center of the image, but it is generally not the case. Moreover, the cameras are not perfect imaging devices: the light goes through lenses before reaching the sensor. These lenses induce small errors on

the projection, known as lens distortions. These distortions are illustrated in Figure 2.4. These distortions can be modeled and compensated, and we do so using OpenCV [Bradski, 2000].



(a) Effets of riadial distorsions:
Solid lines : no distorsion,
Dotted lines: Barrel,
Dashed lines: Pincushion.

(b) Undistorted and distorted images superimposed.

Figure 2.4: Lens distortions effects.

2.2.2 Extrinsic Calibration

In the SAR system, we add a projector. The system is calibrated using ProCamCalib by S. Audet. [Audet and Okutomi, 2009], which enables easy projector-camera calibration. Figure 2.5 illustrates the procam system. The transformation from the projector's point of view to the camera point of view are called *extrinsic* parameters, illustrated in Figure 2.5. We consider that the origin of our coordinate system is the camera, consequently its extrinsics parameters are identity. In equation (2.3) we propose new formulation of (2.1) with identity extrinsic parameters. The "view" of a point (x_p) from the projector's point of view is described by equation (2.4).

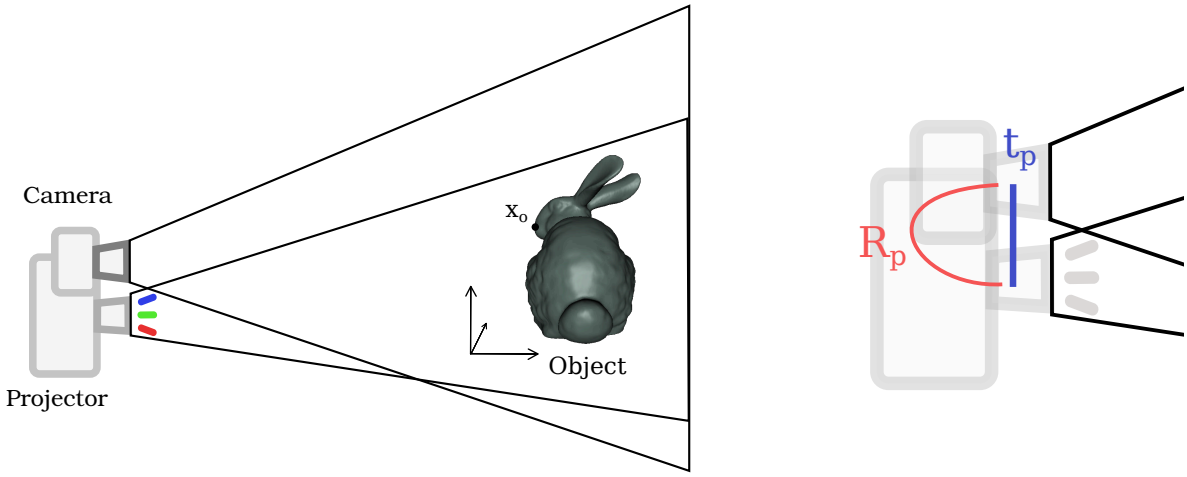
$$\mathrm{x_c} = \mathrm{K_c}\ (\ \mathrm{I}\ \mathrm{x_o} +\ 0) \tag{2.3}$$

$$\mathrm{x_p} = \mathrm{K_p}\ (\ \mathrm{R_p}\ \mathrm{x_o} +\ \mathrm{t_p}) \tag{2.4}$$

From these equations we know how to project a point from the camera's point of view to the projector's image. In order to project a virtual scene on top of the physical world, it is required to create a scene that mimics the projector's capabilities as illustrated in Figure 2.6.

2.2.3 Calibrated Virtual Camera

The OpenGL convention states that the image generated is between $[-1; 1]$ for both x and y axes. Our calibration provides a focal length that transforms distances in millimeters to pixels. We can use it for the creation of the OpenGL scene given some scaling operations. The OpenGL



(a) Procam system: the projector and camera are fixed together.

(b) The relation between projector and camera is a rotation (R_p) and translation (t_p). These are called extrinsics parameters.

Figure 2.5: Procam configuration.

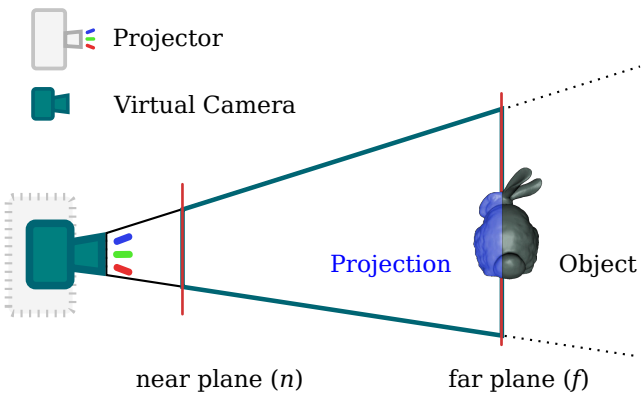


Figure 2.6: The OpenGL scene fits the projector's intrinsics parameters.

camera is normalized, and takes two units in width and height. Consequently the focal of the OpenGL camera can be expressed like this :

$$f'_x = 2\ f_x/s_w \tag{2.5}$$

$$f'_y = 2\ f_y/s_h \tag{2.6}$$

The optical center also needs scaling, and it needs to be negated because a camera looks negative in the z direction. The resulting optical center for OpenGL is computed like this :

$$c'_x = -(2\ c_x/s_w - 1) \tag{2.7}$$

$$c'_y = -(2\ c_y/s_h - 1) \tag{2.8}$$

Now that the focal lengths and optical center are computed, the last parameters for the creation of the OpenGL projection matrix K_{gl} is the near (n) and far (f) plane distances. These distances are application dependent, though it is quite safe to put distances between 10cm (n) and 10 meters (f) for applications with low power projectors. Consequently the resulting projection matrix in OpenGL is this :

$$K_{gl} = \begin{pmatrix} f'_x & 0 & c'_x & 0 \\ 0 & f'_y & c'_y & 0 \\ 0 & 0 & \frac{-(f+n)}{f-n} & \frac{-2fn}{f-n} \\ 0 & 0 & -1 & 0 \end{pmatrix} \tag{2.9}$$

For more information about OpenGL projection matrices we invite the reader to refer to the *gluPerspective* documentation or this website[1]. From now on, it is possible to use the projector to project on any 3D point it can "observe". In order to project geometrically correct images and textures on objects, it is required to know where they are.

[1] http://www.songho.ca/opengl/gl_projectionmatrix.html

2.3 Multi-touch with Kinect

Depth cameras enable the acquisition of the 3D information of a surface. Depth is represented as a 3D point cloud which contains the position of each point and eventually more information such as normal and color. We used a Kinect depth camera, which uses a *light coding* technique developed by PrimeSense [2]. The Kinect camera provides a 640x480 depth image, with 3D position coded on 10bits and color at 30 frames per seconds.

2.3.1 Transformation from Kinect point of view to normalized coordinates

We propose to use this 3D information to make interactive any planar surface, inspired by [Wilson, 2010]. Their technique rely on identifying the differences between the background and the most recent frame to detect touch events. In contrary to them, our technique does not rely on background subtraction. Instead it is based on a threshold operation over a 3D plane.

In order to create a touch interface for 2D and 3D touch we need to define a new coordinate system. We create this coordinate system onto the table and relative to the projection area, as illustrated in Figure 2.7. Our technique requires a calibration phase. This calibration estimates the position and orientation of the table (or any planar surface). Additional, it estimates a transformation from the 3D depth information of the Kinect to a 2D coordinate system corresponding to the trapeze projected on the table (Figure 2.7(b)). The third coordinate in this coordinate system is the distance to the table, as illustrated in Figure 2.7 (c).

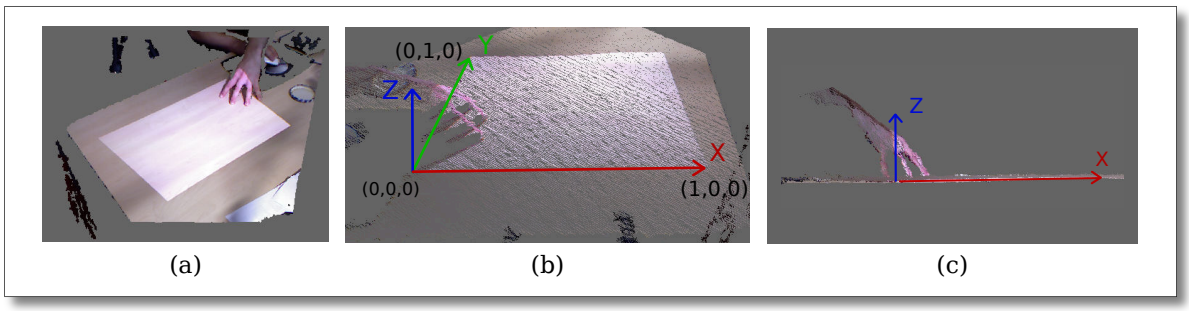


Figure 2.7: The calibration estimates the transformation from the depth camera's point of view (a) to a normalized space corresponding to the projection (b). The two first axis (X and Y) are on the table and are scaled the projection area. The third axis (Z) represents the distance to the table (c).

Calibration

The calibration is done with a dedicated calibration software we created. We use the color information of the Kinect depth map to assist the user in the selection of the color points. We created different calibration methods to best suit to the current context: with or without a projector, and controllable light conditions or not.

[2] www.primesense.com

Calibration without projection

Manual The user sets his or her own visual feedback with physical objects such as paper stickers. Then, the user selects the corresponding points on the Kinect image displayed on the screen.
Pros: Does not require a projector.
Cons: Time-consuming.

Color-detection This technique is based on multiple color detection. The user indicates colors to detect by clicking on them on the Kinect color image. The detection technique is illustrated in Figure 2.8. It is possible to adjust the threshold values for more robust detection. Once the colors are selected, the calibration is automatic.
Pros: Calibration with moving surfaces, fully automatic once the colors are selected.
Cons: Color detection depends on ambient light conditions, requires a calibration board with the colors, may have precision issues.

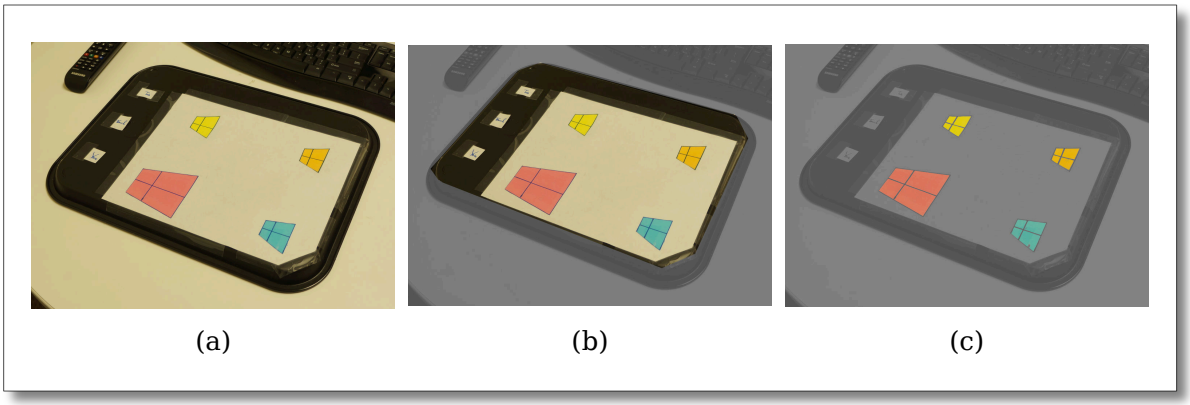


Figure 2.8: Color-detection for calibration. At first, one color is detected in the whole color image of the Kinect: pink in (a). Only the points with the same orientation and connected are selected (b). Finally, the others colors are detected among these points (c). This detection enables to track common colors without background subtraction techniques.

Calibration with projection

Semi automatic Several points are displayed onto the surface one after the other. The user clicks on a point on the Kinect image, then the next one is displayed and so on.
Pros: Quick calibration, human detection of the projected points.
Cons: Projection not always visible.

Automatic The user does not need to perform any action. Firstly, the projection zone is detected by projecting a bright blue color. Secondly each calibration point is projected and the detection is done within this zone only.
Pros: Fully automatic calibration: one button press.
Cons: Requires reasonable light condition (no complete dark nor bright sunlight).

From our experience, the semi-automatic calibration appears as the most quick and robust solution for now. The color detection is also very quick and robust, but it requires a calibration board and it is less precise than a manual technique.

2.3.2 Detection and tracking

Once the calibration has been done we obtain depth points with normalized coordinates on X and Y axes, and distances in millimeters in the Z axis. These points are used to detect and track the user's fingers touching the table as 2D input and the user's hand over the table as 3D input.

Touch input

The touch input is illustrated in Figure 2.9. It uses the 3D plane of the table to detect only the elements above it (a). A thin slice above the table is used to detect the fingers touching the table. However, this touch detection does not rely on contact, but on distance. Consequently, the touch events are fired before the user feels the haptic feedback of the finger touching the table; they are even fired if the finger does not touch the table but is close enough. Each group of points is detected by a connected component search in the image plane. The tracking is simply done by comparing the connected components from one frame to the next.

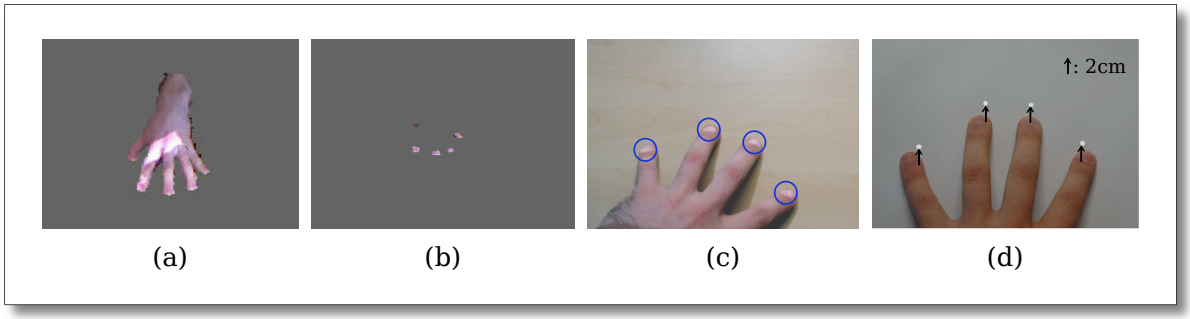


Figure 2.9: Hand detection after threshold operation over the table's plane (a). For the touch detection, only a 1-2cm slice over the table is selected (b). Once the system is calibrated, it is possible to project over the detected elements (c). The touch detection is often shifted (d) so that the user knows precisely where the touch is detected.

Hand input

The same detection is used for the touch and 3D spatial input, with a slight variation on the 3D input. The connected component's position are not computed using the whole information but only the points that are the most forward: the points with higher Y values. Consequently, the 3D pointer is placed on the second phalanx of the middle finger if the palm faces the interactive area. A better solution would be to estimate the principal orientation of the tracked object to select the points along this orientation instead of the Y axis. It would also enable 3D pointing with direction. It was not required by the applications we explored and the finger detection is

tricky with a Kinect. Although, it is possible with specific placement like in [Ren et al., 2011] and clear finger gestures.

Touch on sheets of paper

The hand and touch input are relative to the projected area. In order to make it relative to the sheets of paper we make one assumption: the piece of paper is put on the table. Consequently, we can use the touch 2D touch events in screen space and project them to the piece of paper coordinate system. In order to do this, we transform from the normalized coordinates to 3D coordinates viewed by the camera. The depth of the touch is not known in screen space. Consequently we create a 3D ray and find its intersection with the tracked sheet of paper to obtain where the user is touching in 3D. Using this, we can create paper with touch and spatial input anywhere on the table within the projection area .

2.4 Performance

2.4.1 Paper tracking

The paper tracking is using ARToolKitPlus, its performance in terms of precision depends on the quality of the camera calibration. The calibration we obtain with ProCamCalib (part of JavaCV[3]) provides sub-pixel precision for the camera and projector calibration [Audet and Okutomi, 2009]. Moreover, the tracking with ARToolKit can be done at interactive frame-rate (60fps) on modern machines. Consequently, the tracking rate and raw precision are good enough for our applications.

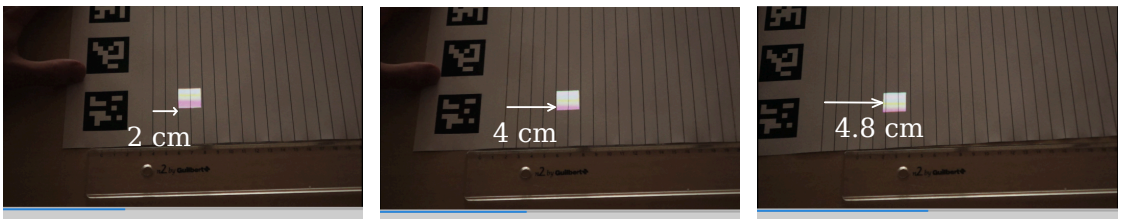
The limitation is the resistance to occlusions. During the drawing, the user occludes many markers resulting in small changes in the normal (orientation) estimation. In order to provide a more comfortable and precise user experience we added a threshold for the update of the position and orientation of the sheet. We suppose that the orientation of the paper changes only when the position of the paper changes. Consequently, the position and orientation is updated only when the paper is moved more than 3 centimeters away from the last recorded location. This distance is empirical: small adjustments are possible without triggering the tracking. The user has to take care of his or her own drifts which is easier to deal with than “shaking” projection.

We conducted a simple experiment to measure the the tracking/projection latency. It is illustrated in Figure 2.10. We project small white square 2 centimeters apart from the left border of the projection and 2 centimeters wide. Then the sheet of paper is moved to the left in a quick motion: 2cm in 200ms = 10cm/second. The camera we used has a frame-rate of about 30 images per second which leads to a resolution of 33 millisecond between two frames. The paper is moved after Frame (a) and the projection follows after Frame (b). As the paper moves rapidly, the error can be important (c). However, some of the users said they liked the fact the projection “follows” the paper. The latency has many sources:

1. Compression of the video by the camera.
2. USB bus is not made for high performance.
3. Linux Kernel used was not a “Real-time” one.

[3] https://code.google.com/p/javacv/

4. ARToolKit detection.
5. Application processing (reduced to minimum in this example).
6. Projection latency.

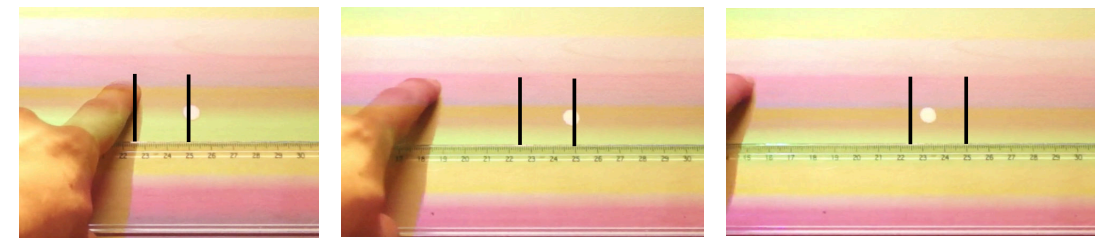


(a) t = 3s 060ms (b) t = 3s 270ms (c) t = 3s 300ms

Figure 2.10: Measurement of the tracking/reprojection delay of ARToolKit tracking. Before the movement, the projection perfectly matches the paper (a). It takes about 200ms to the projection to follow the paper (b). If the paper moves quickly the distance between the paper and the projection can be important (c).

2.4.2 Touch tracking

We conducted the same latency measurement to estimate touch input's latency; it is illustrated in Figure 2.11. We estimate the latency to about 170ms. The explanation of the latency is the same as the paper tracking, except that the ARToolKit detection is replaced by a Touch detection. This latency is higher than touch screen devices and the tracking suffers from the low frame-rate of the Kinect device. Consequently, the tracking of very fast motion is the main limitation of this kind of touch tracking. A better frame-rate and lower latency provide a better user experience as shown by Jota and colleagues [Jota et al., 2013].



(a) t = 1s 510ms (b) t = 1s 687ms (c) t = 1s 780ms

Figure 2.11: Measurement of the tracking/reprojection delay of Touch tracking. At the beginning of the movement the touch has no error (a). The projected elements follows the finger (b) after about 177ms. The error depends on finger's speed and filtering (c).

The touch detection is done on unfiltered data from the Kinect. Depth information from the Kinect is noisy: the depth points are "shaking" from a position to another. Immobile objects have a small error changing at every frame. The precision depends on the distance between the Kinect and the table. We put it at a 60cm distance to the table which leads to more than 10 depth points for each finger touching the table.

In order to compensate for both the low frame-rate and noisy input, we use the 1€ filter [Casiez et al., 2012]. It is composed by a low-pass filter which removes the noise from the touch input resulting in steady touch events. A predictive filter enables a precise motion estimation. Once filtered the touch events are responsive and we did not have any complain nor comments thorough our experiments.

2.4.3 Projection size

Projection enables a wide variety of sizes of creation. We explored very small and large projection areas. In Figure 2.12 (a) we used our system to draw a very small rabbit, the paperclip gives an idea o the size. The drawing in (b) as been done on a whiteboard, the dragon was 80 centimeters plus the flames. Smaller projection and interactive sizes would require special optical elements; while bigger interactive projection and touch interfaces is not problem with the current technologies.

We explored the use of multiple procams to overcome two limitations of the current hardware: the resolution of the projection and the shadows created by the users. The use of multiple procams is straightforward as each procam has its own tracking camera. The first possibility is to create a focus+context system [Baudisch et al., 2001]. One projector projects a low resolution large area, and the second one projects a high resolution small area. The high resolution projector can be moved by the user to the desired zone. In our case it applies well, but the minimum projection distance are generally too high with the current optics to obtain sharp tiny image. The second use of multiple projectors is the reduction of the intensity of the shadows and the augmentation of the projector's resolution. It is illustrated in Figure 2.12 (c).

(a) 3 cm Bunny (b) 80 cm dragon (c) Multiple projectors

Figure 2.12: Advantages of projection over screens: small (a) and large (b) images with the same hardware, and multiple projector at the same time (c).

2.4.4 Digital Screen on paper

We created an Application Programming Interface (API) to create applications within the sheets of paper. Each sheet of paper has its own "screen". The screen has a certain size depending on the piece of paper tracked and the resolution is expressed in pixels per millimeters. In the application the rendering can be 2D, 3D, and even 3D using head tracking described in chapter 4. All the sizes are expressed in millimeters instead of pixels like in traditional user interfaces.

In addition to the screen, each piece of paper have access to the touch events in its own coordinate system. A schematic view of the system with the coordinates system is illustrated in Figure 2.13, and a photography of it in Figure 2.14. Additionally , we created a button API to add buttons on a paper screen. Buttons are placed and sized in millimeters, and events can be handled by the paper application. It enables the creation of simple user interfaces with interactive elements.

In chapter 6 we propose a simplified version of the API, closer to the Processing language to explore creative coding on paper.

It is also possible to obtain a picture of the digital screen, either with the tracking camera or a specific high resolution camera. The possibilities offered by this image acquisition are also explored in chapter 6.

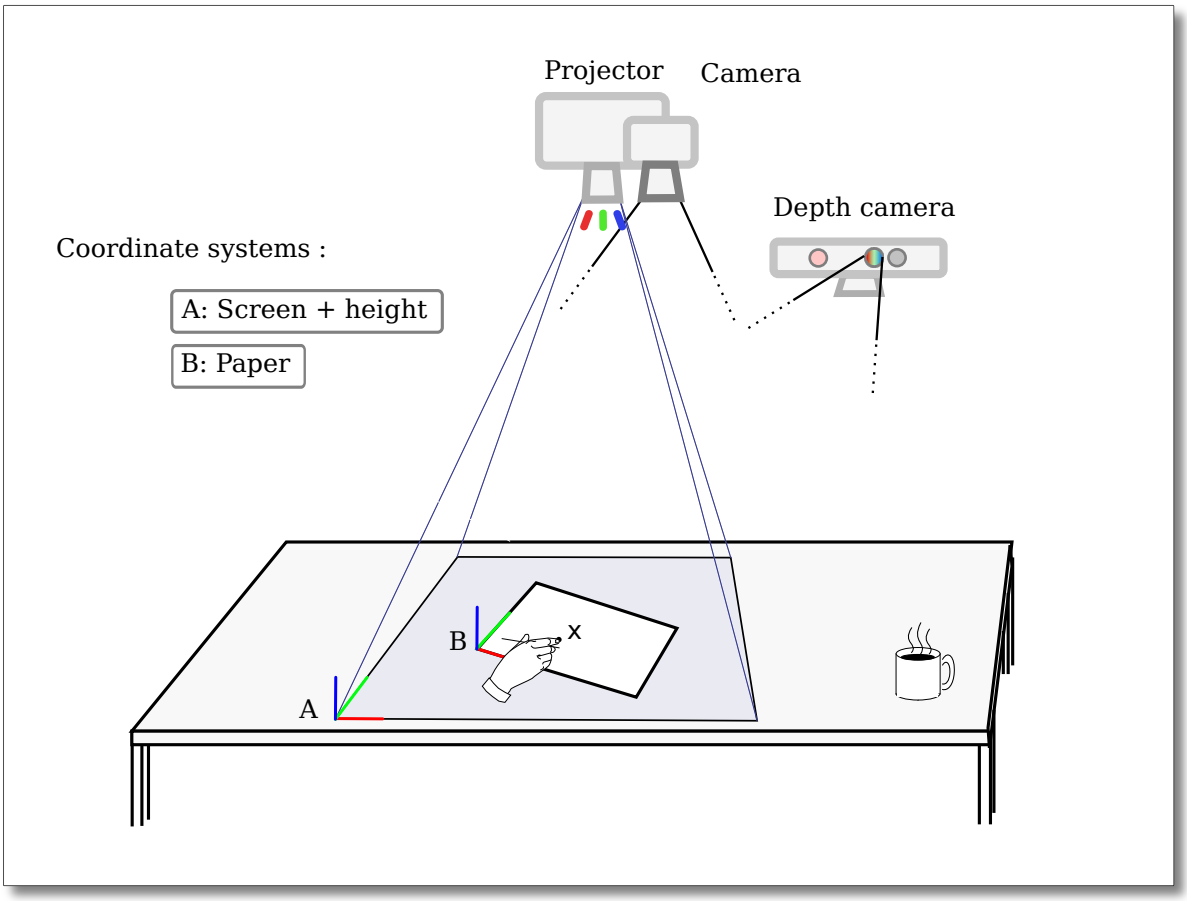


Figure 2.13: Schematic view of the SAR system.

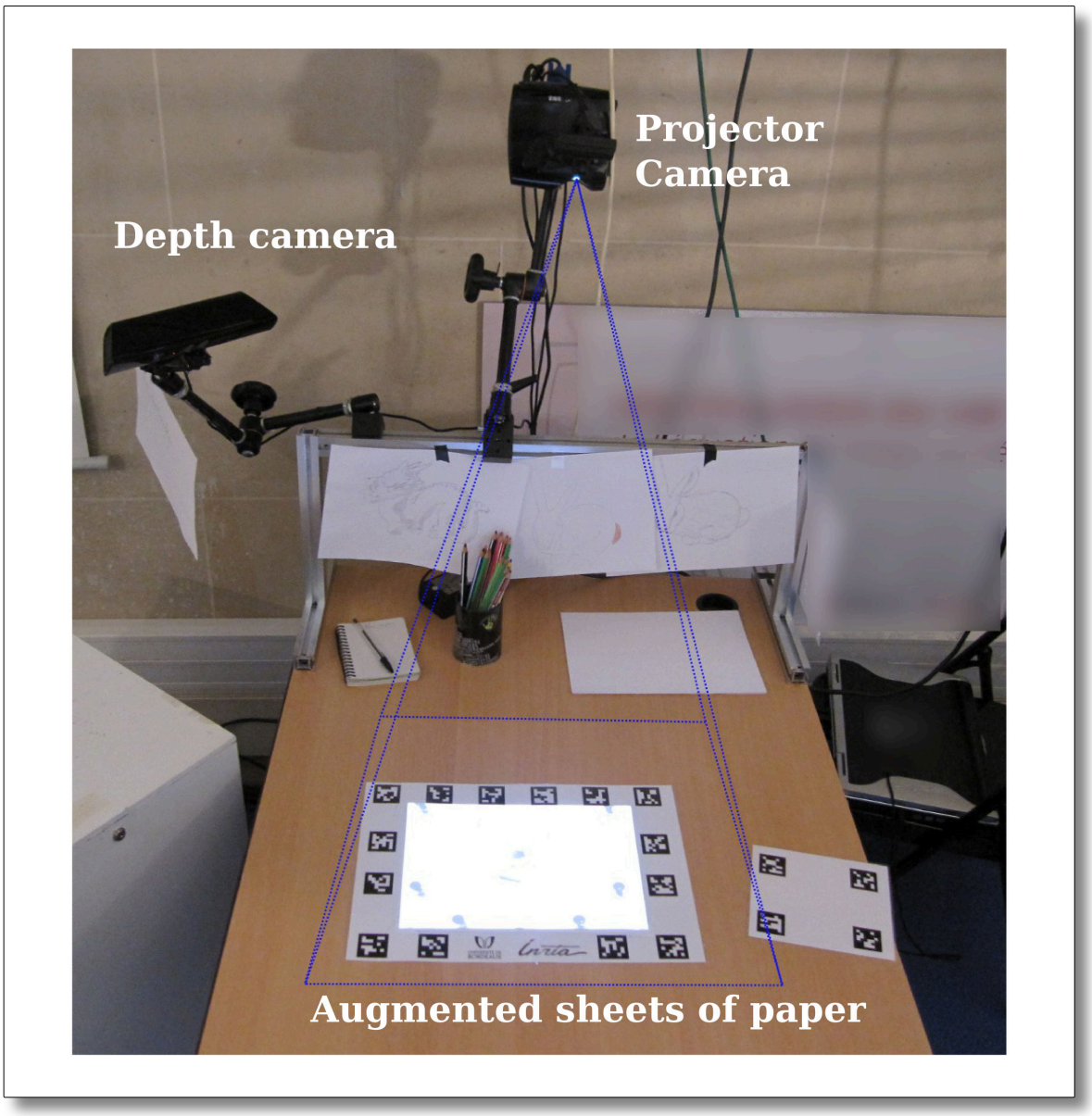


Figure 2.14: Photo of the SAR system during a public exhibition (Palais de la Découverte, November 2011). A larger view is illustrated in Figure 5.1.

Part II

Augmented Drawing

Chapter 3

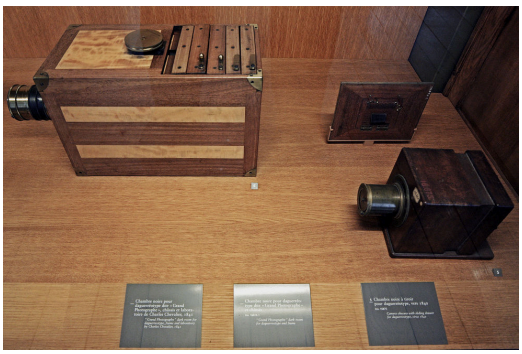
Guiding the drawings

This chapter is part of the work and experiments we conducted on physical drawing. The goal of these tools and applications is the creation of physical drawings and paintings using Spatial Augmented Reality (SAR). In this chapter we propose tools inspired by teaching books. In the next we propose tools using 3D rendering.

3.1 Motivations

3.1.1 Projection for drawing

The camera obscura model has been used for centuries [Gernsheim and Gernsheim, 1969] as the first optical device to project a planar image onto the real world (Figure 3.1). It is as simple as a hole as thin as possible letting light pass through a dark room. Nowadays, equivalent systems are the art projectors, which enables the projection of still images. They come with a wide variety of resolutions and intensities; video projectors are also sold as *art projectors*.



Camera obscura for Daguerreotype called "Grand Photographe" produced by Charles Chevalier, 1841 (Musée des Arts et Métiers)

Camera obscura, from a manuscript of military designs. 17th century, possibly Italian.

Figure 3.1: Camera obscura images, from Wikipedia.

The exact origins of projection for drawing are not really known, and even today the use of projection to create drawings or paintings is somehow considered as cheating for some people. However, the use of projection is like any other structuring technique, such as using a grid to copy, enlarge or reduce an image. It is a common technique in drawing and painting, and as this is not used to identically copy copyrighted material it is widely accepted by artists.

3.1.2 Preliminary work on projection for drawing

We conducted experiments on drawing by projecting an image on the piece of paper. These preliminary experiments were performed to evaluate the ease of use of our system for drawing, the precision limitations and its scalability.

We proposed the users to bring their own image or select one from a few images we provided. Then, they took the necessary time to draw it on a A4 sheet of paper with a pencil or colored pencils. The resulting drawings are illustrated in Figure 3.2. Each user took about 40 minutes to finish their drawing and their first comments were valuable and taken into account for the next applications. Here are the two main comments:

- The precision of the projection is crucial: a small error in the tracking-projection loop leads noticeable errors, even a few millimeters error. The details in the projection are smaller than one millimeter, and the pencil strokes can be extremely thin.
- The intensity of the projection has to be controllable by the user. The projection is very intrusive during the drawing and consequently the lighter strokes are hard to see. It is even more important during the shading and coloring phases.

Additionally, we observed that the tracking speed was not important at all. While the user is drawing he or she blocks the piece of paper with the non-dominant hand and draws with the dominant hand. Consequently, the tracking is needed only when the user changes his or her drawing position. These movements depend on the user's drawing skills: a novice artist will move the paper a lot, while a confirmed one will be more agile and will not require to move the drawing support as much. This preliminary work encouraged our system and approach to assist the drawing.

3.2 Image filters

We explored the use of simple image filters to ease the drawing. It could be used to to show hidden details from an image, or get rid of zones where the details are less important. The insight on which details to draw or ignore requires training and experience. We propose to aid the artist on these decisions by the use of image analysis techniques.

We propose to use three image enhancement methods, as illustrated in Figure 3.3. The enhancements are applied directly on the projection using the touch interface. They are displayed as circles, the image (f) in Figure 3.3 shows a view during drawing.

The first filter (b) brightens the projection on a given zone. It is used to show more details from low-contrast zones. The second filter (c) blurs the image and makes it less bright. It is

Figure 3.2: Photo during the creation of a drawing, and resulting drawing from photos. Each drawing took about 40 minutes.

Figure 3.3: From a photo (a), the artist can directly apply image filters: brighten (b), blur(c), edge detection (d)(e), or a combination of them (f).

used to set regions which are less interesting at a given time. The last filter (d)(e) is a derivates of gaussians filter. This filter displays the edges (high gradient zones). The size of the analysis can be adjusted to display different gradient sizes. Moreover, the sensibility of the filter can also be adjusted to display more or less informations. These filters are complemented control of the projection's intensity. This set of tools provide different point of view on the same image; it is possible to use them to reveal details that are hard or impossible to see without filters.

3.3 Construction lines

The idea to create *Digital Construction Lines* (DCL) comes from the analysis of multiple books for beginner artists : [Sanmiguel, 2009][DuBosque, 2009][Edigo, 2010] [Ronin, 2011][Hachette, 2009]. All of these books advise the beginners to create construction lines as much as they need. Consequently, the first advises are to draw the construction lines with a "hard" pen (H) and light strokes to make them easy to erase. Some construction lines are sketches that will form the drawing, others are structuring elements that will not appear on the final drawing. They are either made by trial and error to place the different shapes to draw, or following strict rules such as perspective construction lines.

Our approach is to create tools that add new possibilities to existing tools. Consequently the DCL are designed to complement the Physical Construction Lines (PCL). We propose to create the DCL directly on the drawing paper, like physical ones. Here is a list of the main features of digital construction lines:

- Instant and clean deletion.
- Projection over drawn or painted areas.

- Automatic generation.
- Interactive edition, with undo/redo, copy, paste etc...

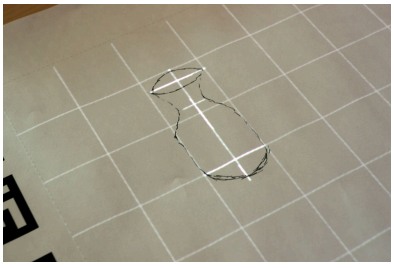
We explored three types of construction lines. The first type is general purpose construction lines for structure: lines and a grid. The other types are drawing specific: perspective construction lines, and character construction lines. All three techniques have been tested through either formal or informal user studies.

3.3.1 General purpose construction lines

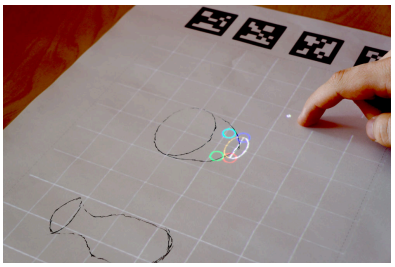
General purpose construction lines are DCL that could be useful in most drawings. An example of the use of these construction lines is illustrated in Figure 3.4. It is possible to create lines, curves, axis-aligned lines and a grid. These tools are enough to create the basic structural elements of a drawing. They are designed to be used with a minimum training time, so that we could compare them with physical construction lines. We first describe the user interface, and interaction techniques; then we compare the physical and digital construction lines in a controlled user study.

In our approach we propose tools to create digital construction lines, but it is to the artist to decide where to draw them. Another approach is explored by Iarussi and colleagues in [Iarussi et al.,]: they automatically generate construction lines from image analysis. Their approach is complimentary to ours.

Examples of construction lines for symmetry. From Cours complet de Dessin [sanmiguel, 2009]



The grid enables a simple structure, the lines provide an idea of the sizes.



The digital curves are more complex to create...

Figure 3.4: General purpose construction lines are simple lines that structure the drawing. From the example on the left we propose two examples of the use of DCL.

3.3.2 Augmented keypad

The user interface is created for bi-manual use. The non-dominant hand is used for application control and the dominant hand for pointing. This dominant/non-dominant combination is the same as a keyboard - mouse configuration: the same hand is used for using the mouse and pointing with the touch interface.

The controls are embedded on an Augmented Keypad. The keypad is a small keyboard with keys from zero to nine and math signs. We modified a keypad to use any key as a screen like the Optimus Maximus Keyboard. Instead of small screens, we use projection like in [Gellersen and Block, 2012] which is much cheaper and already available in our system. We also put a status text on the top part of the keypad to remind the user in which mode they are.

We did not modify all the keys, as the "plus", "minus", "enter", "insert" and "delete" keys could be used as is for most applications. In fact, "minus" and "plus" are used to switch between lines, "delete" is used to delete a line, and "enter" to unselect (validate) the current line. The keypad is also tracked and can be moved at any time. Each key had a white sticker on it to enable clear projection, although a matte white keypad would have been better.

We propose to use a consistent pattern for the user interface. The menu button is accessible at all times. The most used buttons are placed on the second row (from bottom), then on the third row and so on. Additionally, most of the buttons are "ON/OFF" buttons: one press to activate it and one press to deactivate it. At each time only the actionable buttons are displayed and each button press leads to an indication in the status area. The user interface is illustrated in Figure 3.5.

3.3.3 Pointing

In the line edition tool, it is possible to use direct or indirect pointing (see Figure 3.6). The usual pointing pattern is at first direct pointing, then indirect refinement of the position. The indirect pointing does not suffer any occlusion and the movement speed is scaled at 1/4th of the finger's speed. If the destination is another point, or on the grid the point has the possibility to snap to this other point. When snapped the points are not "glued" to together: each point is always independent during edition.

For the grid (Figure 3.6 (a)), the touch interface is used to move and scale it. Its size is displayed in the status area when resizing: a specific size can be required for technical drawings and copies at a given scale. The grid movements are always relative and it is often faster to move the grid a little bit to mark a point than creating or tracing a construction line. The users whom performed the best during the user study are the one who used the more the grid. Consequently, an evolution of the user interface for performance would integrate a "grid movement" button available at all time.

3.4 User evaluation

We conducted a user study to evaluate the usability of the system. The first goal was to validate the perceived precision and quality of the projection mapping system. The second goal was to check if our digital construction line tools could effectively complement or even replace physical construction lines. We did an independent measure evaluation on 28 subjects on a perspective drawing.

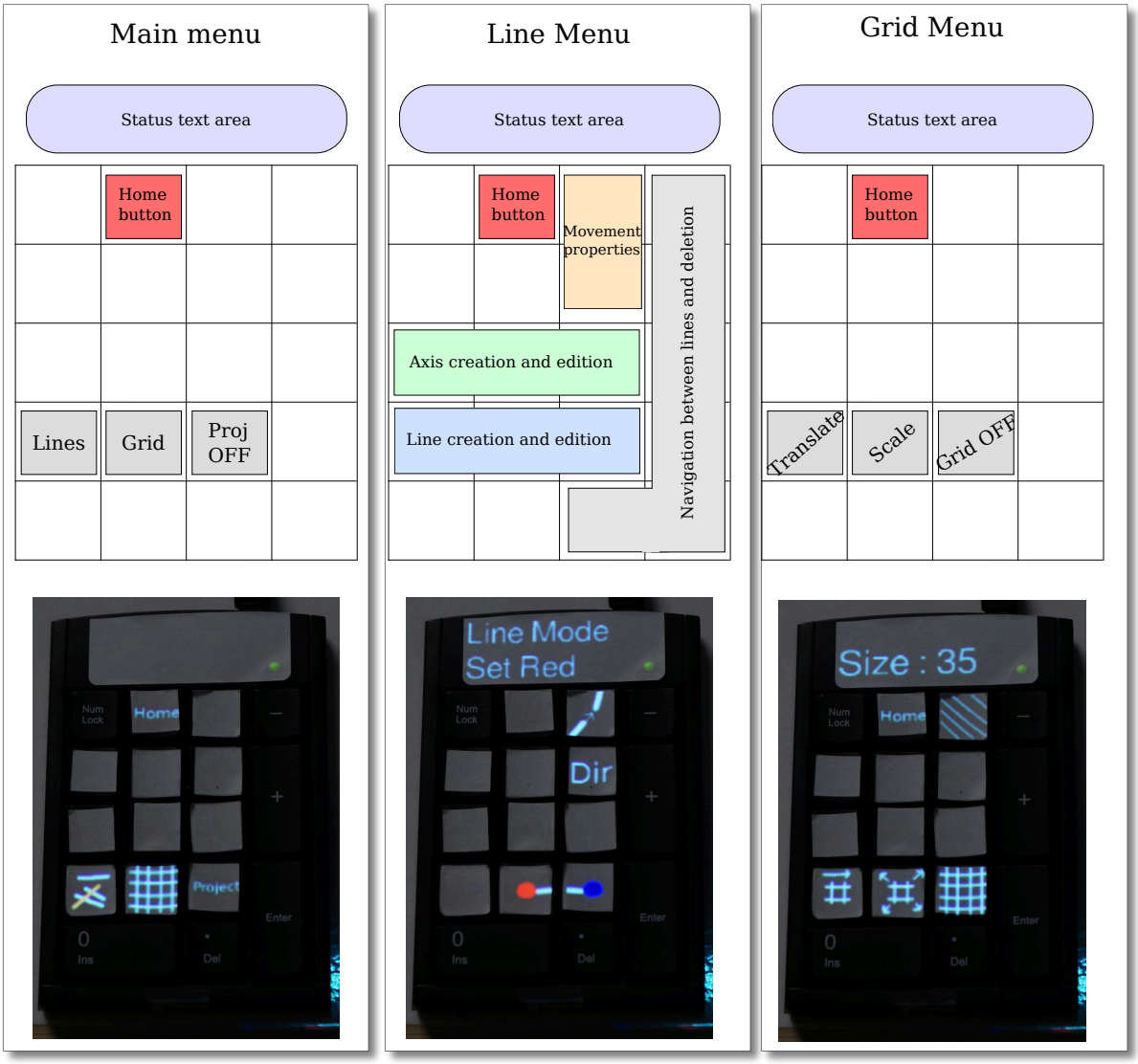
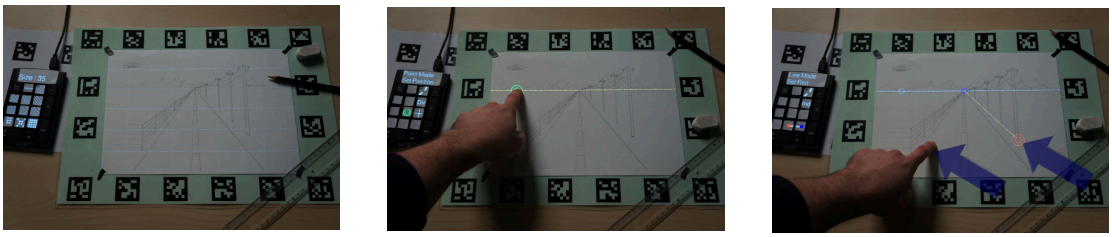


Figure 3.5: The user interface is projected on a keypad. Top: views of the three different modes. Bottom: photos of the projections.

3.4.1 Task

We chose to use an exercise from a pedagogical book on drawing entitled "Dessiner pas a pas" [Ronin, 2011] (p.370-372). The task consists in a step-by-step perspective drawing exercise of 17 construction lines which leads to a 11 lines house. It is made to be easy and it is extremely guided: each step contains a picture of the desired result and a short text describing how to achieve it. We chose this guided task so that the drawings would be comparable to each other. Moreover it did not require any expertise in drawing, consequently the user's level in drawing did not interfere much during the experiment.

We divided the users in two groups, one group (BO: Book Only) uses the instructions, a H pencil, an eraser and a ruler. They had to follow the instructions, and we instructed them that all the lines that will be erased should be as light as possible to leave as less traces as



(a) Grid (b) Direct horizontal line (c) Indirect line

Figure 3.6: The users could manipulate a grid (a), draw horizontal (b), vertical, or lines with any orientation and length (c). The positioning of a line is either direct (b) or indirect (c).

possible. The BO group was composed of 6 male and 7 female participants. The second group (P: Projection) had the same tools, and our system with the application described above. They were instructed to avoid drawing construction lines as much as possible, in order to only draw the final strokes. The P group was composed of 7 female and 8 male participants.

3.4.2 Participants

The participants were volunteers aged from 22 to 59 ($\mu = 29.9$). Most of them were novices in drawing, a few of them had a few drawing courses or draw at home, but none of them were trained professionals. We discarded two participants: one did not follow the instructions, and the second did not understood the concept of construction lines.

3.4.3 Apparatus

Both groups had to complete questionnaires before and after the drawing task. They were instructed to do the exercise at their own pace, although they knew the exercise was timed.

The BO volunteers were alone or in small groups in well-lit rooms with a large space to draw. Each of them had their own A4 paper, pencil, eraser, ruler and instructions. Once they were all ready, we asked for silence and gave them the written and oral instructions.

The P volunteers were in a virtual reality room with controlled light conditions. These constant light conditions were required to provide the same projection experience for every user. The projection zone was 54cm by 34 cm with a 720p projector (5 pixels per mm^2). The projected image averaged a resolution of 650*460 pixels on a A4 paper sheet.

Each participant had to begin with a 10 minutes maximum training exercise. One user asked for more time to learn how to use the interface.

3.4.4 Results

Drawing experience

The three first questions of the questionnaire were common to both groups. They were on a 1-7 Likert scale (1: Strongly agree, 7: Strongly disagree) with some space for comments. The

questions and their results are displayed in Figure 3.7. The task was found easy (μ=1.6) and amusing (μ=2.07) by both groups. Furthermore they felt as confident with digital lines as with physical ones: BO μ=2.07, P μ=2.2. These results indicates that the experience of drawing may not be altered, or at least that our system does not negatively impact the drawing experience.

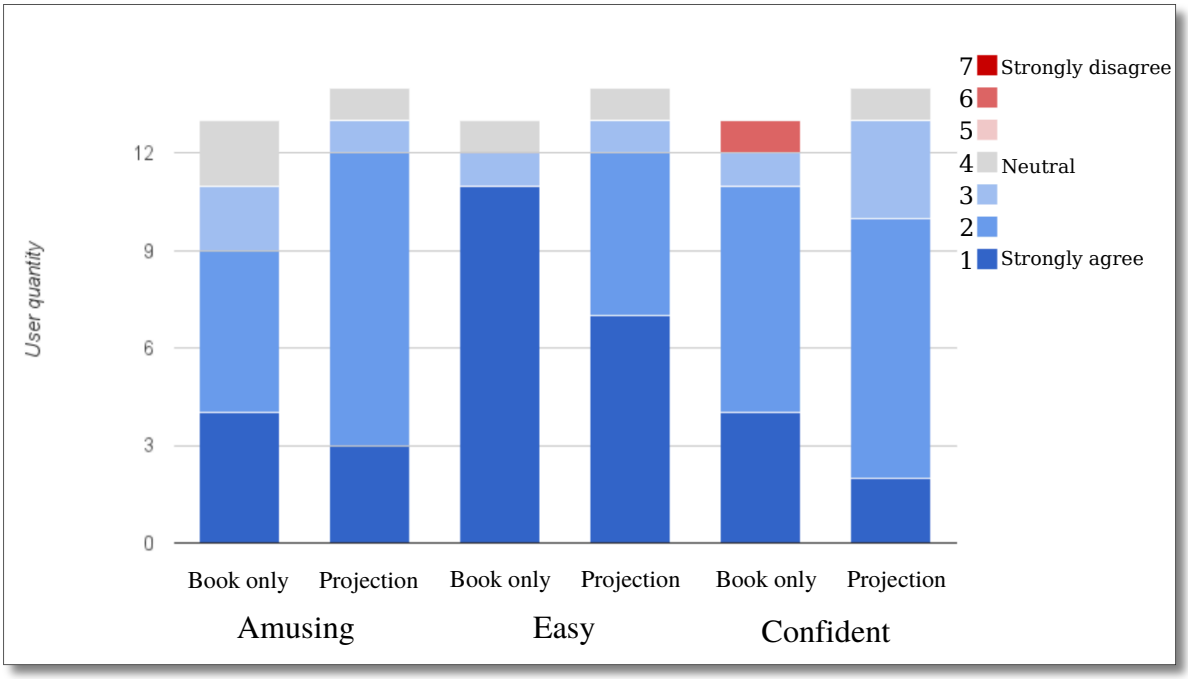


Figure 3.7: Subjective evaluation about the fun the users had, the ease of the task and their confidence before tracing.

Observations on the projection

The Projection group had five more questions about the user interface and the projection quality. They were stated as illustrated in Figure 3.8. The interface was designed to be simple and minimalist, so that the users could focus on the drawing task without a long training time. It was found easy to use by most users (μ=2.3).

We also collected objective results on the interface use. They showed that the users did not make many mistakes. The average creation number of horizontal or vertical lines is 4.9 ($\sigma = 2.98$) and standard line is 17.9 ($\sigma = 4.32$). In the task they had to create 4 vertical lines and one horizontal, and instructions from the book indicated the creation of 17 lines. The average number of keys pressed on the key pad was 185 ($\sigma = 61$), for 13.5 ($\sigma = 4.4$) actions per minutes.

The question about the size of the projection zone divided the users in two groups. One group did not move the piece of paper, consequently they never had any trouble with the size (8 users). The other group changed their position more often and also the piece of paper. Consequently, they encountered the projection's limits (7 users).

The accuracy of the projection was found good enough (μ=2.5). However, some users

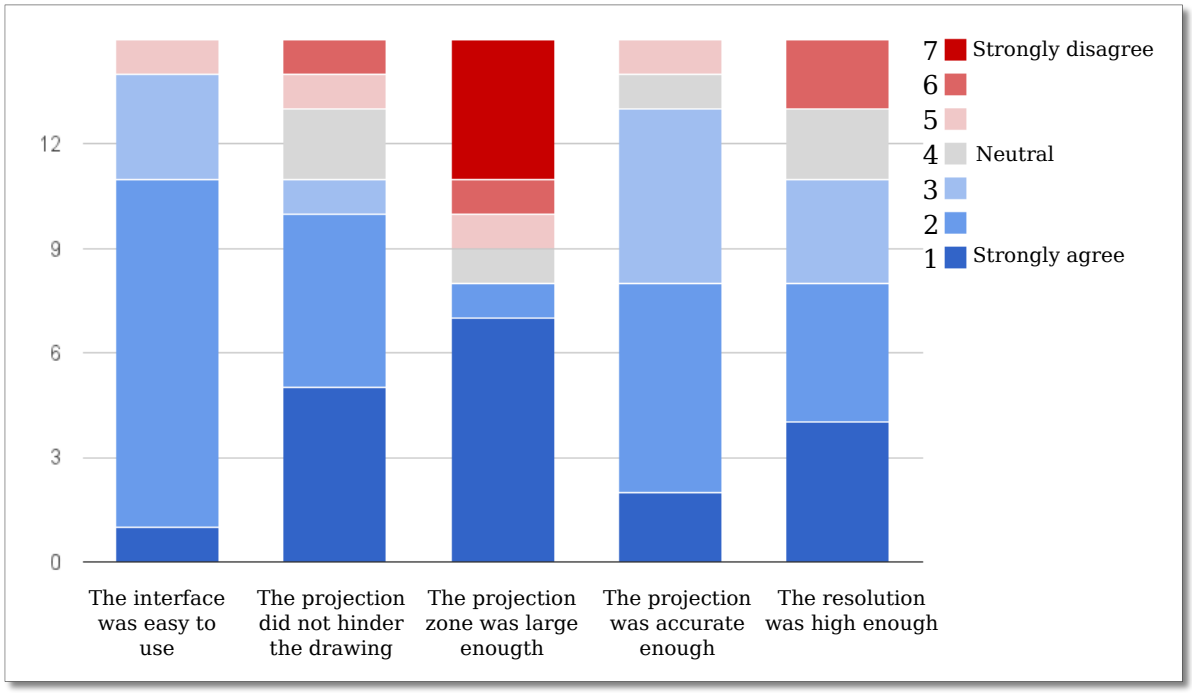


Figure 3.8: Subjective evaluation of the augmented reality system.

reported that the definition of accuracy was not completely clear to them. It was mixed with the displayed size of construction lines. It is a positive point for us, it means that they did not encounter any projection error. The perception of errors seemed dependent of the user's drawing abilities (linear correlation coefficient : -0.395). The practice of drawing trains the visual system to see better and better understand the shapes of objects. It can be an explanation of the different perception of the imprecision of the system. Trained users were also more sensible to the filtering we implemented.

The projection changes the perception of drawing, its intensity has to be adjusted to the current light conditions with dedicated controls. With our controlled light conditions we managed to set projection intensities that enabled a good visibility for both projection and drawing. Most users found the projection more useful than harmful : μ=2.5, $\sigma = 1.6$.

Drawing results

We monitored the drawing duration, as we mentioned before the users were instructed to draw at their own pace. The users without the projection were faster: 10 minutes in average ($\sigma = 2.5$) compared to 14 minutes ($\sigma = 2.9$) with the projection. The difference is statistically significant (t-test: $p<0.05$); it was expected because the user interface was not speed oriented but made to be easy to use. Moreover, the user had to take decisions on what kind of line to project or use use the grid. An example of the variety of the uses of the system is illustrated in Figure 3.9. Another important observation is that most users did more digital construction lines than instructed, *e.g.* in Figure 3.9 (c) a hidden part of the roof is projected.

Each drawing have been evaluated by at least five persons. The evaluation was on the quality

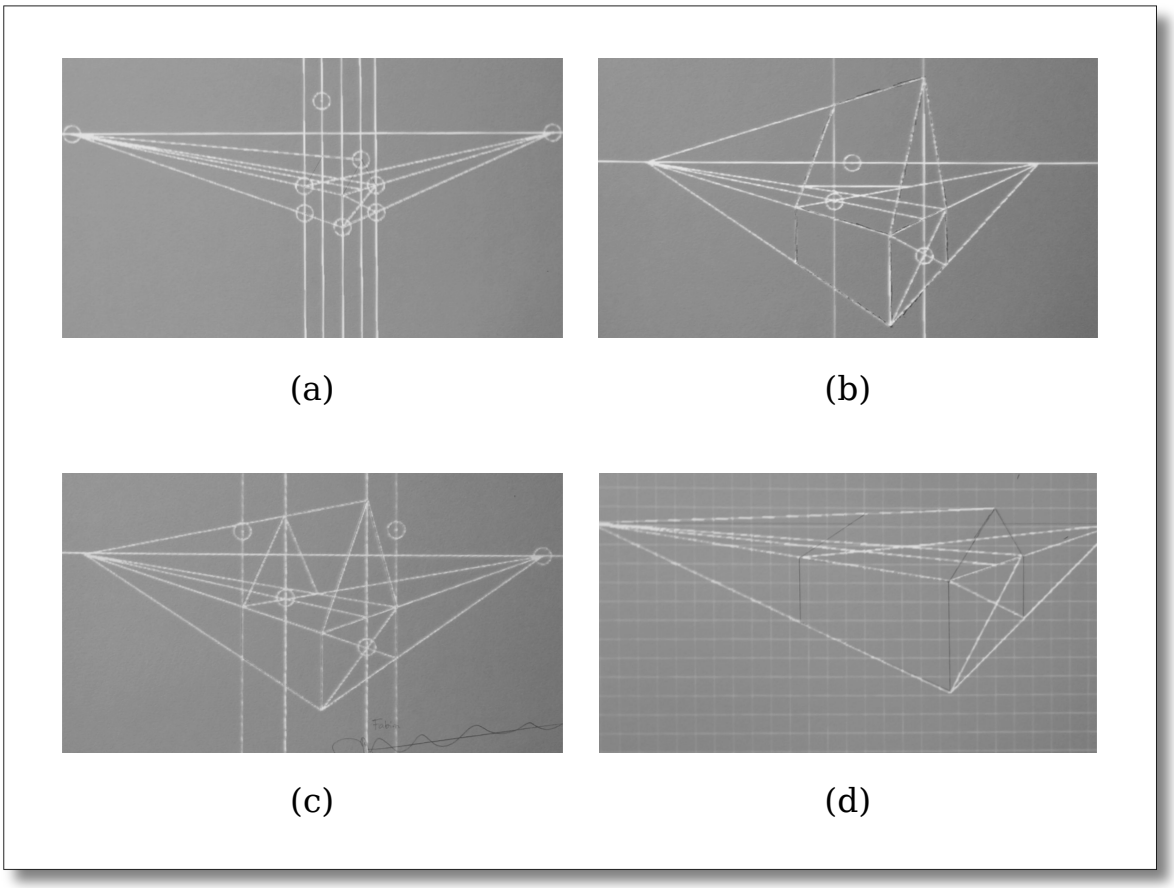


Figure 3.9: Projection at the end of four drawings. User (a) created fewer digital construction lines than instructed, but enough for the drawing. User (b) did not use vertical lines for each walls of the house. User (c) created construction lines in occluded areas (back of the roof). User (d) extensively used the grid instead of horizontal and vertical lines.

of the resulting drawing, the qualitative elements are illustrated in Figure 3.10. Here are the results :

- Stars: Intersection with too many pencil stroke, BO: μ=6.4 ; P: μ=1.5.
- Light Line: Trace of an erased line, BO: μ=8.6 ; P: μ=0.2.
- Heavy Line: Trace of an erased heavy line (error) BO: μ=1.0 ; P: μ=0.2.

We expected the difference in number of *stars* and *light lines* to be high and distinctive between the two groups. However, the heavy lines number is also different and the difference is statistically significant. Both values are very low and the P users had to think more before tracing, and they spent more time in drawing. Consequently, we are not sure if the error rate decreased because of our system or the drawing time. It would be interesting to evaluate more complex tasks to get a clear observation on the error ratio.

The overall results are good and they show that our system may successfully replace the physical construction lines in this task. It means that it is a valid alternative, and it complements

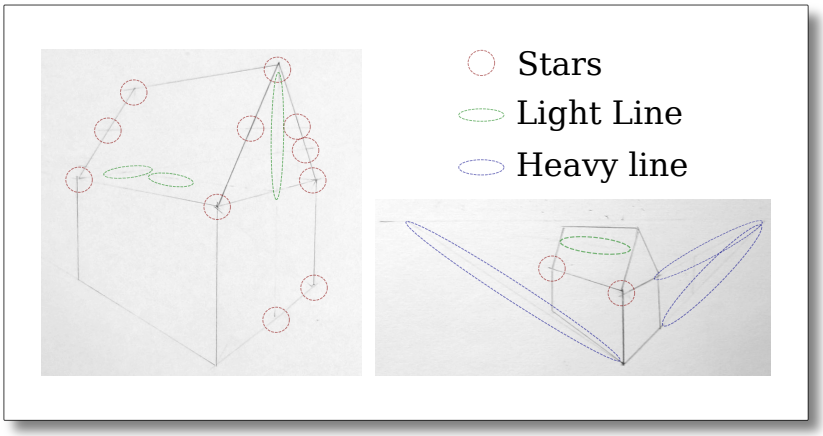


Figure 3.10: Errors in the drawing without projection. The stars represent intersections errors, the light lines represent traces of construction lines and the heavy lines represent errors impossible to fully erase.

the physical construction lines.

The time performance was not explored in this user study, it is quite difficult to compare existing physical to new digital tools in terms of training time because of the past experience of every user. The next tool is dedicated to perspective construction lines. With the perspective construction line tools the drawing time is diminished drastically. The drawing from the user experiment is done in less than 4 minutes with the specific tool with little training and could be improved with interface evolution (such as the grid-movement button). As a comparison, the best time was 6 minutes without projection by the most experienced subject (50% more).

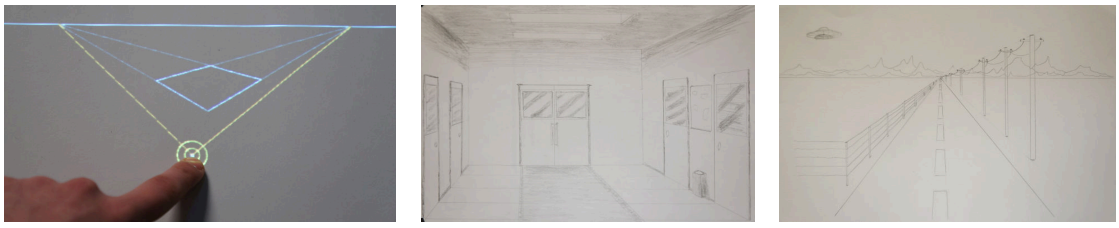
3.5 Dedicated construction lines

3.5.1 Perspective construction lines

Perspective drawing requires specific construction lines, usually they originate from an object (such as a house) and goes towards the horizon at a vanishing point. There are different types of perspective: one point perspective is when the viewpoint is facing an object. Two point when objects are rotated from the user's point of view. Three point perspective is for extreme cases where the viewpoint is very close to the ground or far above it, *e.g.* such as skyscraper drawings. The perspective drawing tool we created allows the manipulation of a "horizon" and up to two vanishing points.

When a new point is created a line going towards each vanishing point is also created (see Figure 3.11). We extended this tool to perspective quad drawing for two-point perspective. The perspective quad enables very fast creation of four lines just by setting the position of the first point and two distances along the first lines. The quad is highlighted and the rest of the lines is thinner (thus darker) in order to keep the drawing readable.

The fast generation of many lines leads rapidly to a visual clutter of the drawing area. Consequently, we propose to the user to "hide" the lines: the lines gets darker and the points are still able to be snapped to. This perspective construction line tool is inspired by the "Analytic



(a) (b) (c)

Figure 3.11: Perspective construction line tool (a), and two conic (one point) perspective drawings (b)(c).

Drawing of 3D scaffolds" [Schmidt et al., 2009] and it is a functional proof-of-concept. The creation of multiple views of an object using this kind of perspective tool could lead to a solution for modeling 3D objects on paper. It may enable to bridge the gap between the "paper design" and the 3D model to complement the 3D scaffolds.

3.5.2 Character drawing

Character drawing is difficult to achieve because our perception of humans, and more specifically faces, is very good [Puce et al., 1996]. In this work, we focused on body proportions and pose.

Bounding cage

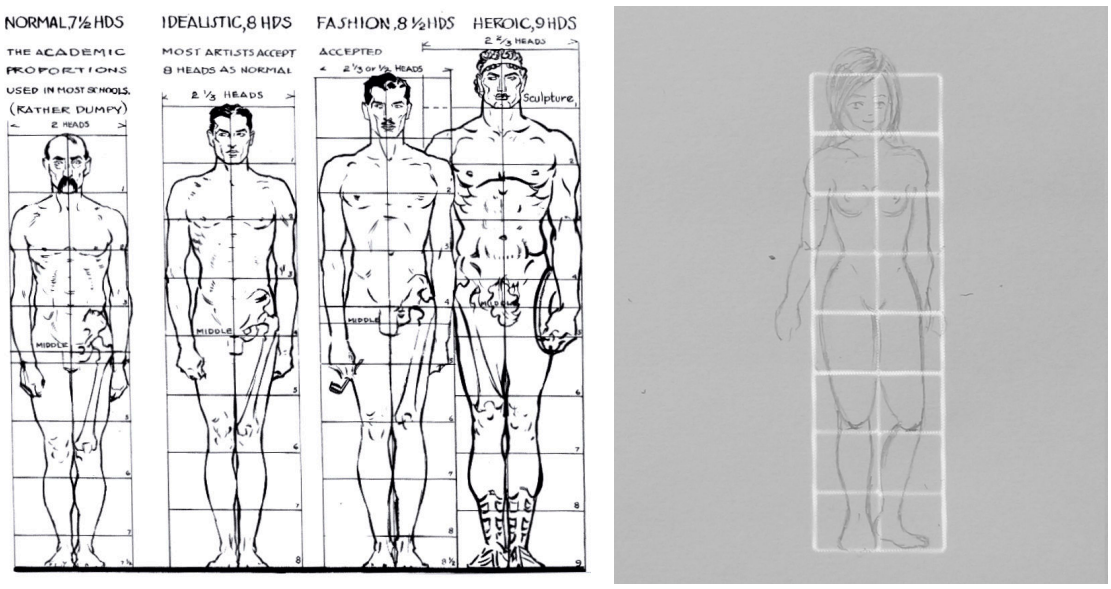
We propose to project proportions for the whole body. These proportions changes with sex, age and strength. They also vary between teaching books; in this work we followed the proportions rules from "Figure Drawing for All It's Worth" by Andrew Loomis [Loomis, 1949] (Figure 3.12 (a)). The user can set the size of the construction lines, then he or she can choose to change strength or the age. The strength varies from 6.5 heads to 9 heads in height, with corresponding width. The age varies from 1 year to 20 years according to Loomis' descriptions.

This tool constraints the pose of the character to a standing position. It is a good exercise to learn how to draw limbs and memorize the proportions, as illustrated in Figure 3.12 (b). This cage constrains the drawing which can be frustrating for the artist, but is a necessary step for beginners. For advanced artist, Loomis adapted the cages to different body poses and perspective conditions.

We could extend this "cage" rendering to 3D rendering enabling perspective and complex poses. However, we focused on a more standard way of creating character poses: drawing from a model.

Drawing from a model

The usual way to visualize a given human pose is to use a mannequin, human or wooden.



(a) Character proporitions by Andrew Loomis [Loomis, 1949].

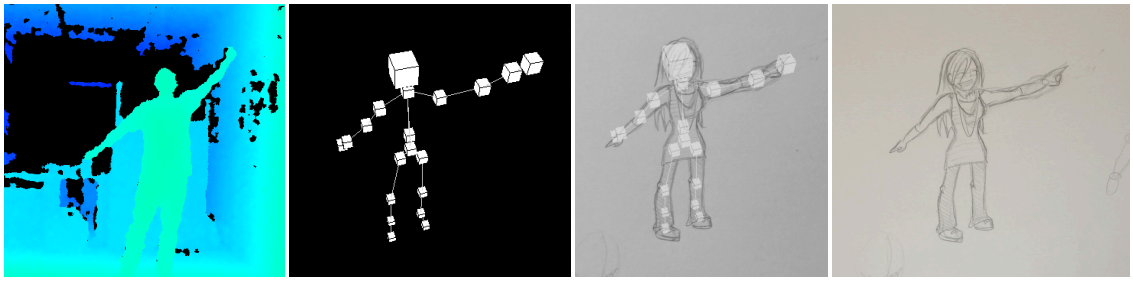
(b) Drawing done using a projected grid for proportions.

Figure 3.12: Character drawing applications. Left: a grid guides the character's proportions. Right: the 3D skeleton from a Kinect guides the drawing.

Here we propose to use a human model, and extract his or her pose using the skeleton tracking from Kinect. Another possibility would be to take a photo of someone and project it at a given scale. In one hand a photo provides rich visual information, on the other hand our solution enables the simulation of any point of view. In Figure 3.13(b) the point of view is very high, and it would require some equipment or climbing to get a picture with this angle. Moreover, we work with a simple representation of the body, consequently it is possible to apply deformations on the skeleton, *e.g.*: for cartoon characters.

The skeleton tracking is quite robust and allows a wide variety of poses. The 3D skeleton is projected on the paper and the point of view can be adjusted using the touch interface. The user can then draw the character using the projected skeleton, one of our examples is illustrated in Figure 3.13.

A natural extension of this work is the projection of other difficult parts of the body such as the face and hands. Once again, using 3D models instead of photos could enable more poses and deformations that suit the drawing. Then it would be possible to use the important body of work on face tracking and hand tracking for artistic creation.



(a) Depth view (b) skeleton view (c) projection (d) result

Figure 3.13: Character drawing from a depth camera. The pose is captured from the depth stream (a). The point of view of the skeleton is adjusted for the drawing (b). The resulting image is projected on paper (c) to achieve a drawing (d).

Chapter 4

Drawing of a 3D digital model

The idea behind this chapter is that many drawings depict observed objects. The artist draws what he or she sees. In this chapter, we focus on still life creations: where the scene is created or at least arranged for the drawing. Instead of drawing a physical scene, we propose to the artist to manipulate and edit a 3D scene directly on paper. Consequently, it is possible to create an infinite number of scenes and to use different rendering styles to guide the drawing.

4.1 Motivations

Still life drawing is an important exercise on visual perception, the artist draws what he or she "sees" not what he or she "understands". This capability requires time and training. In this chapter, we work one step before: on the creation and edition of the scene to visualize. Moreover, we also work one step after: once the scene is created it can be rendered using specialized 3D graphics.

The creation of synthetic 3D scenes enables infinite possibilities for object composition, lighting conditions, and point of view generation. Additionally, a rendering does not depend only on the scene but also from the rendering technique. There are many *non photo-realistic rendering* (NPR) techniques elaborated to mimic human drawing. We propose to use them for the creation of physical drawing instead of digital drawing.

From all these possibilities, we created a solution to manipulate a 3D scene directly on paper and change the light conditions. We propose a solution to visualize 3D objects as if they were put on paper and interaction metaphors to interact with them. We also explored how a cartoon shading could make drawings faster and easier.

4.2 3D scene rendering on paper

4.2.1 Problem

The standard way of visualizing 3D scenes is to suppose that the user faces the rendered image. We illustrate this in Figure 4.1 where the user looks in the blue axis direction on the monitor and it is perpendicular to the red and green axes. When we use the paper as a screen, this

assumption is no longer valid. It is important to have a correct perspective to understand the 3D scene. For monoscopic or still images the small perspective problems are not a problem, but for stereoscopic or moving images it is very important. Errors in rendering of a moving scene or stereo scene confuses the user and can lead to headaches or motion sickness because the brain tries to make sense to something that does not.

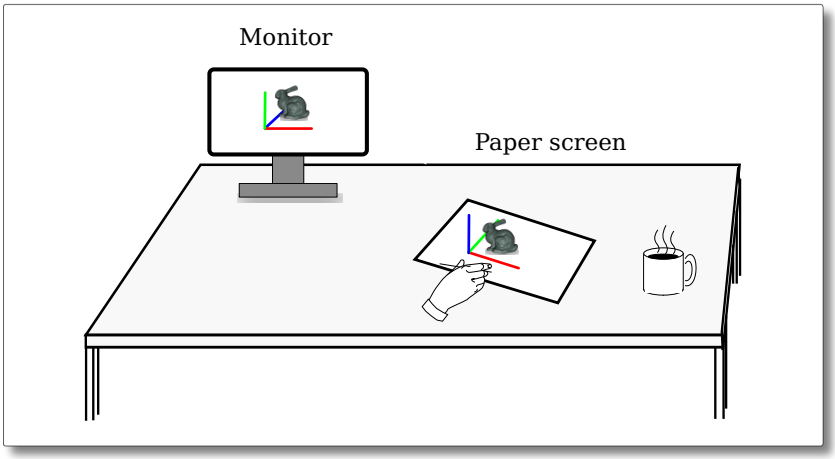


Figure 4.1: The rendering on a monitor and a paper screen is different: on a monitor the user is facing the screen. With a paper screen the screen is moving: the user's point of view is required to obtain correct rendering.

In order to better understand this perspective problem, we illustrate it with examples of photos in Figure 4.2 [1] [2]. Usually, this kind of images where the perspective is not facing the screen/canvas/paper are called anamorphosis.

4.2.2 Solution

In order to render 3D objects in the correct perspective it is required to take into account the user's point of view. In this case, we also take into account the screen location as it is also subject to movement. This kind of rendering is common in virtual reality. It is either called "virtual window" when the elements visualized are far inside the screen, or "Fishtank VR" when the elements seem contained within a small space inside the screen.

In our solution, the visualized elements stay onto the sheet of paper, like if they were glued on or inside the piece of paper. It creates the same effect as the "pop-up anaglyph" of Figure 4.2. Spindler and colleagues propose a wide range of uses of such paper windows in [Spindler et al., 2012]. With good head tracking and stereoscopic rendering, it creates a compelling holographic rendering for one user only, illustrated in Figure 4.3.

4.2.3 Limitations

We chose to fix the user position in a sitting position in front of the table. The fixed head position is both a technical and practical limitation. The most simple solution is to add a camera

[1] Images from :http://www.stereomaker.net/sample/popupgallery/gallery.htm
[2] Pop-up anaglyph images from: http://www.stereomaker.net/galleries/katori/main.htm

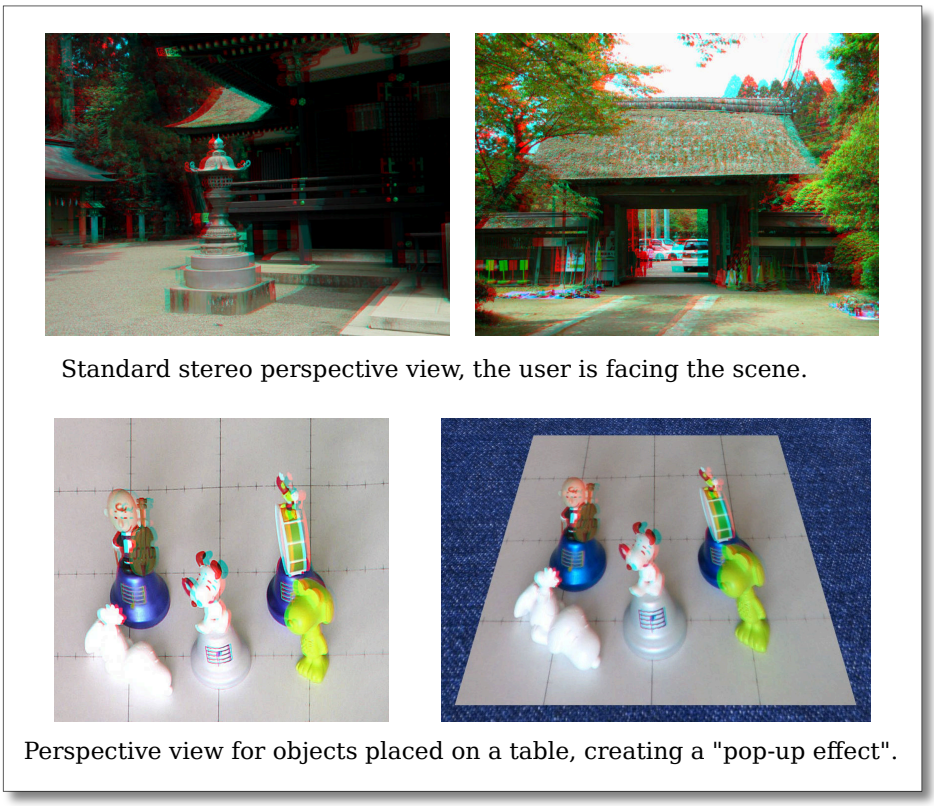


Figure 4.2: The top images have to be seen facing the document. The bottom images can either be seen with the document put on a table (left) or facing it (right).

Figure 4.3: Stereoscopic view of the 3D scene using red/cyan filters.

to track the user's position. However, one more a camera includes another video stream to the PC in addition to the paper tracking flux and the Kinect flux. Three camera flux in one PC can usually not achieve high frame-rate (30~60fps each) due to hardware bus limitations.

Consequently the head tracking has to be done using an additional computer. A motion capture system using IR markers would take some of the "naturalness" off: with the current system, the user can sit and start using it without wearing anything more. The solution we tried is to use an additional Kinect on a dedicated tracking PC. The two kinect should not have the same views: overlapping areas are interfering with each other. Some solution exists to remove these interference such as using motors to blur the view of the other Kinect patterns [Maimone and Fuchs, 2012]. With one Kinect oriented towards the table, and the other one to the user, it works properly. This head tracking solution was working, but the skeleton tracking rate (using Microsoft SDK) was too low and its latency too high to provide a compelling experience.

Another limitation is created by our implementation of the rendering: we use the paper like a screen. Consequently, the elements that pops-out of the screen are not rendered even though it would be possible. In order to achieve this, we would have to update the rendering using projective texturing [Segal et al., 1992] as done in "The office of the future" [Raskar et al., 1998] and MirageTable [Benko et al., 2012].

4.3 Interaction

In this section, we describe how the input modalities described above are exploited for interaction with projected 2D and 3D content. With our system, users can interact through tangible and gesture input. Moreover, both tangible and gesture-based interaction metaphors try to mimic real-life behaviors in order to favor ease of learning.

4.3.1 Tangible interface

Point of view

The first interaction technique we propose is to change object's point of view by simply moving the sheet of paper. The 3D object is placed as if it was put on top or inside the piece of paper. Consequently, when the piece of paper moves, its rendering is updated accordingly. This is illustrated in Figure 4.4. These movements provide a good perception of the object's location. It is easy to distinguish between the elements placed on paper level such as the carpet in Figure 4.4 and the elements that expands over the paper level such as the rabbit and columns. The elements on paper level always stick to the same loaction whichever the paper and user's location.

Menu

The switch between the different modes is done by the selection of projected buttons on a small piece of paper (on right in 4.6). The menu selection and the rendering area are separated. Consequently the user can adjust the menu location to his or her handedness. Moreover, it prevents undesired selections: when the menu is placed outside the projection area it is disabled. Consequently, a mode can be locked just by the placement of the menu.

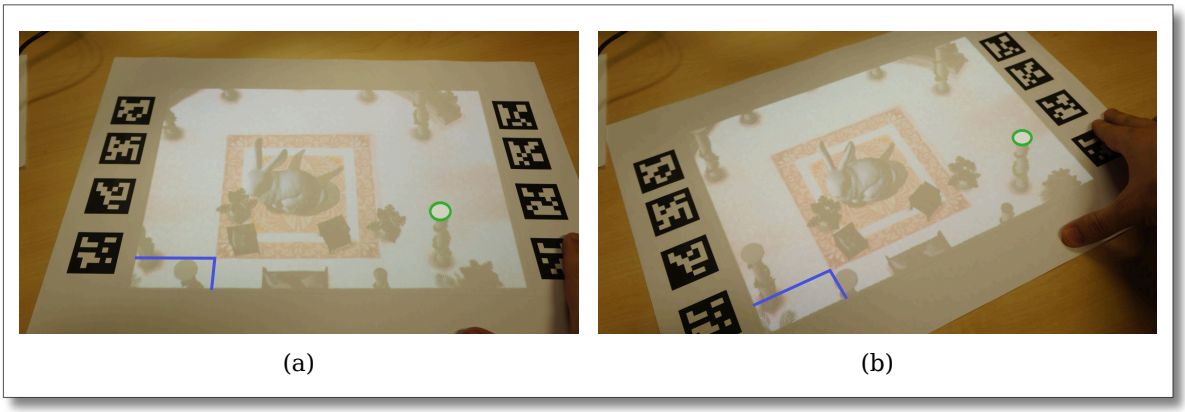


Figure 4.4: (a) Before the movement, (b) after the movement. The distance to the carpet at paper-level (in blue) stays the same after the movement whereas the position of the top of the column (in green) above the paper-level has moved.

This paper menu is an alternative to the projected keypad introduced in the previous chapter. On the one hand, the menu projected on a piece of paper enables easier physical modifications: the piece of paper is lighter and wireless. On the other hand the keypad offers haptic feedback: passive feedback for finger placement and active feedback when the button is pressed.

The two interfaces have different advantages, the keypad offers the performance and reliability of a keyboard. Whereas the projected menus offers the flexibility of paper and the possibility to have interactive areas with customizable sizes and organization.

Limitations

"GUIs are inconsistent with our interactions with the rest of the physical world."

Radical Atoms, by Ishii *et. al.* [Ishii et al., 2012]

For AR applications, the Graphical User Interface (GUI) needs to be minimalist or inexistant. However, the creation applications which take place both in the physical and digital world do require GUIs. We explored the creation of rich interfaces with a GUI on multiple sheets of paper, illustrated in Figure 4.5. In this interface, the selected elements project sub-menus onto the table. However, for the elaboration of good GUI on paper many questions rises yet to answer :

- When do an interface element should be embodied in a piece of paper instead of being a sub-menu?
- Is it preferable to have projection of a menu contained within the piece of paper or around it?
- Are the size and shape of the paper important for projected menus?
- Is it interesting to enable to lock the menu at one location, independent from the tangible?
- Does the "electronics" such as the keypad impede the augmented reality feeling?

The usual research problems and solutions on graphical user interface also apply to projected interfaces. It is extremely difficult to create rich SAR user interfaces only with tangibles. Consequently, the creation of menus and gestures that are adapted to most people and still provide an "augmented reality feeling" rather than a "video game feeling" or "software feeling" is an interesting research direction for future work.

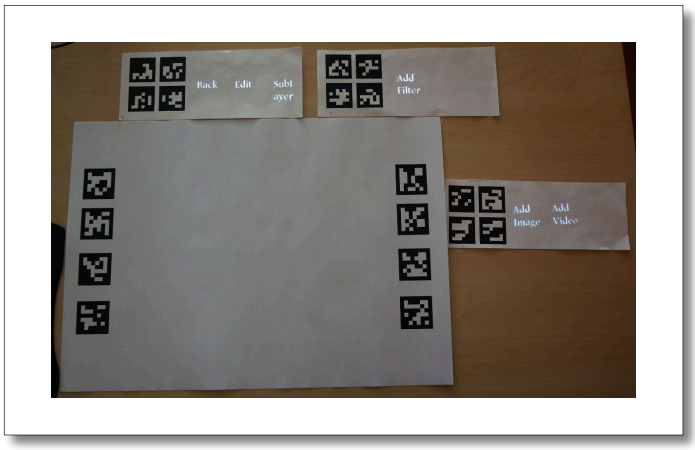


Figure 4.5: Exploration on the use of multiple tangible menus for more complex applications.

4.3.2 Touch interface

2D manipulation

The second interaction technique is based on touch and 3D pointing described in chapter 2. The manipulation of the scene is done using a multi-touch Rotate-Scale-Translate (RST) similar to [Knoedel and Hachet, 2011]. The scene can be translated along the paper sheet plane, rotated along its normal and uniformly scaled as illustrated in Figure 4.6 (a).

We propose three modes of exploration, the first one is by hand movement over the piece of paper. This exploration was created to enable simple interaction with the projected elements. The second exploration mode uses two finger to rotate, move and zoom. This interaction requires some training because the touch interface using Kinect is different to touch screens. Consequently, the users must learn when the fingers are detected and how to use this new kind of touch detection.

With the third interaction technique, the user can move the scene with one finger. Furthemore, it could also be rotated and uniformly scaled by putting one hand paper and touching the paper, as illustrated in Figure 4.7. It is a similar approchach to "Rock & Rails" by Wigdor and colleagues [Wigdor et al., 2011]. This is an exploration to overcome the limitations of the Kinect touch interface by using its strength such as the possiblity to estimate the hand's pose.

3D Pointing

The 3D pointing is used to place a light in the scene. The light positioning is direct: the light is placed on the user's fingertip. Consequently, it stays over the 3D scene and cannot be too far

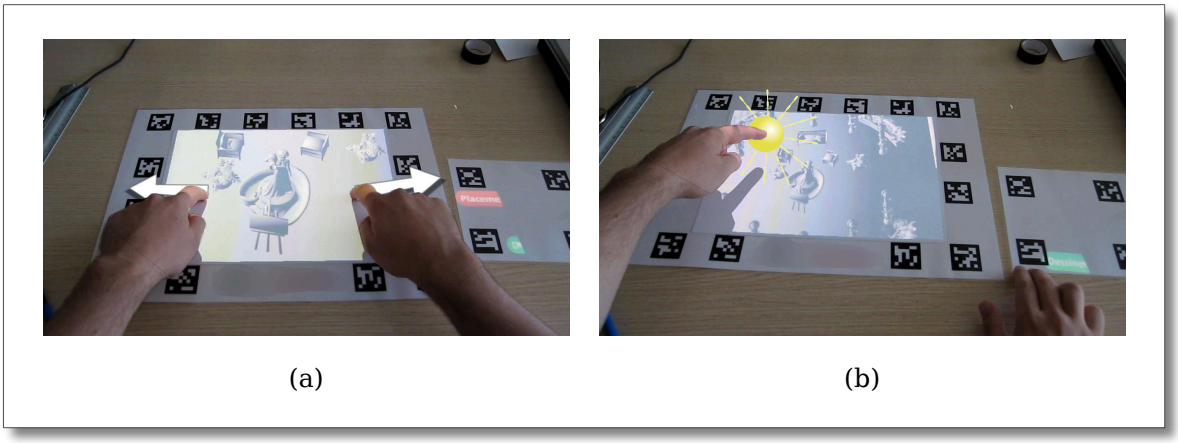


Figure 4.6: It is possible to zoom and rotate using the touch interface (a). 3D pointing can be used to place the light source (b).

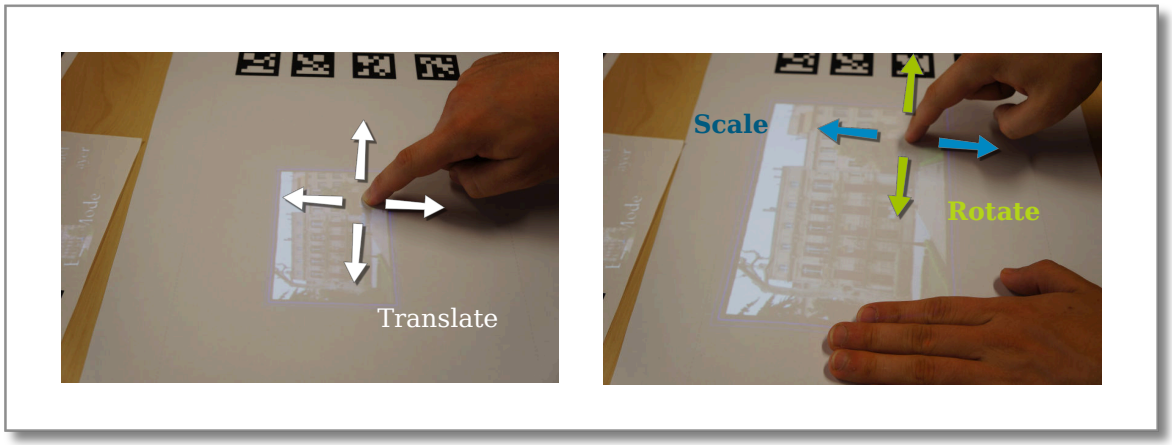


Figure 4.7: Translation can be done using one finger (a). Rotation and uniforme scale are performed with one hand on paper, and a finger from the other hand interacting (b).

from the scene due to the limited detection area. In order to precisely set the light location, the user needs to use one hand for positioning and the other to validate the position.

These interaction techniques constrain the possible point of views. This limitation enables a general public use without training nor instructions. We discuss the interaction metaphors and rendering techniques explored for a general public use in a small time duration in the next section.

4.3.3 General public use

We demonstrated this tool in two different public settings. The first one is a demonstration of it with explanations for 10 to 15 minutes, followed by a drawing demonstration (at Palais de la Découverte). The users had seen how to use it before trying the touch interface. In the second

public setting, the users do not have any instructions (at Cap Sciences). The demonstrations are detailed in Chapter 6.

The first feeling for most users is that the system is "magical" because they did not see the projector and Kinect. Then it was eventually frustrating for them because they were used to touch screens. The touch interface is not as responsive as a touch screen, but is was not the main problem. On touch screens, a touch event is sent when the user touches the screen. However, in our implementation hovering just a few millimeters over the table creates a touch event. Additionally, the touch interface is not limited to a small space: it is spread on the paper sheet and around it. Consequently, when a user is sitting in front of the table, he or she rests their hands on the table creating involuntary touch events. This limitation appeared to be the main issue of the touch interface. It could be addressed by the adoption of a widget-based approach instead of global controls.

4.4 Shading for Drawing

4.4.1 Simplification

Drawing from a projection involves decision making on which elements to draw and how to draw them. These decisions have an important impact on the resulting drawing. With *non photo-realistic rendering* (NPR) it is possible to display only the most relevant details for drawing. However, the suppression of useless details from a 3D scene to obtain a simpler rendering is a difficult problem. It has been addressed in "Where do people draw Lines" by Cole *et. al.* [Cole et al., 2008]. Here we do not have the usual goals of NPR such as fully automatic and temporally coherent rendering. Our approach here is to adjust the rendering to achieve simpler and faster drawings.

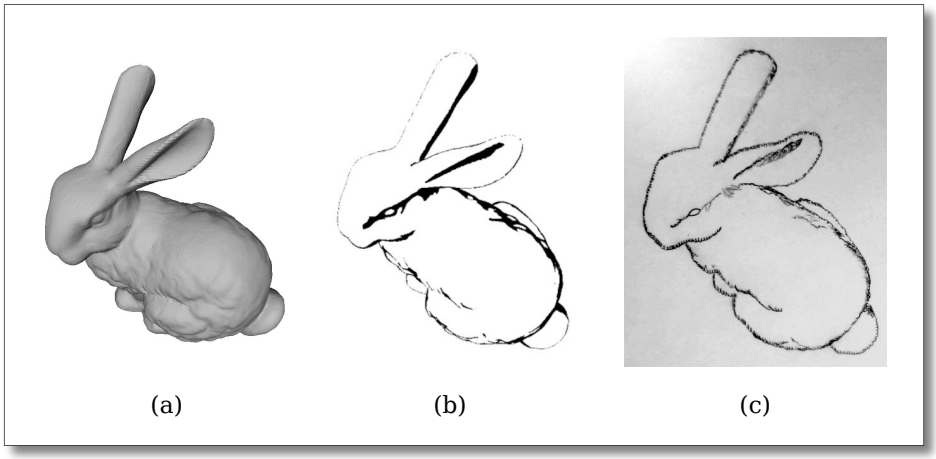


Figure 4.8: Standard Phong shading (a) involves more decision making than cartoon shading (b) for drawing. With simple shading the artist can focus on style (c).

At the beginning of this part, we mentioned that drawing a A4 image took about 40 minutes for each user. With NPR and smaller drawing, we managed to enable the creation of simple yet good looking drawings in 5 minutes. This is illustrated in Figure 4.9. This simplification was

done to limit the demonstration sessions within a short time-lapse. Yet the exploration of NPR for drawing is promising, in the next section we propose an example use of NPR for drawing.

4.4.2 Composition

It is possible to create a drawing or painting from a unique scene by creation a composition of different rendering and/or lighting conditions. The information to display can be extracted from the geometry of the object, as done in [Mellado et al., 2012]. It can also be revealed by specialized rendering such as *Radiance scaling* [Vergne et al., 2010] which enhances the details of the rendering.

We created a simple example of composition: a drawing is created from three images with different lighting conditions. Indeed, a correct lighting is not always required for drawing. The shadows and reflections are usually created to suit the final drawing rather than real life light conditions. In Figure 4.9, we propose to mix three lighting conditions, each light condition favors each a part to draw: the ears (a), the body (b), and shadow (c) of the rabbit.

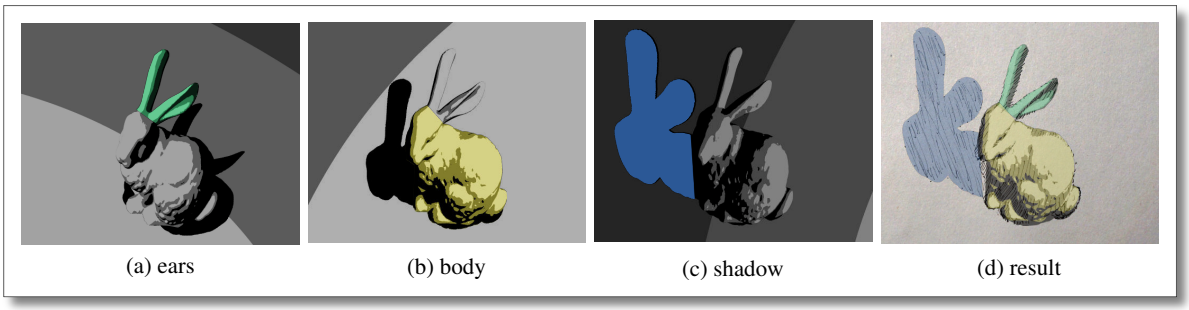


Figure 4.9: The artist can choose different lighting conditions by hovering his or her hand over the scene. One for the ears (a), one for the head (b) and one for the body (c) to achieve one drawing (d).

4.5 Stereographic drawing

4.5.1 Introduction

Stereoscopic visualization has existed since the 16th century. Thanks to recent advances of dedicated technologies such as shutter glasses, polarized light, and lenticular lenses, stereoscopic displays are now quite common. People can benefit from rich visualizations of 3D content that have been prepared for them, but they are generally excluded from the stereoscopic creation process.

Our motivation in this work is to provide tools to allow anyone to create a stereographic pair of images. Contrary to approaches where the full creation process is done digitally (*e.g.* [DeepSketch,]), we explore an approach where user creates two drawings directly on two pieces of paper.

4.5.2 Existing techniques

According to Phillipe Coudray [Coudray,], the easiest way to create 3D drawings is to make the 3D effect as billboards. Using this technique, each part of the drawing will be on a different plane. The plane will be deeper inside the drawing when increasing the left/right disparity, and closer by decreasing it. When these techniques are mastered, the artist can create some objects going through multiple planes. Consequently the different parts will bump instead of being flat. Still, the drawings needs to be viewed using free-view methods as illustrated in Figure 4.10. These methods require the drawing to be side by side and it also requiresome training for the user. Furthermore, they can be visually uncomfortable.

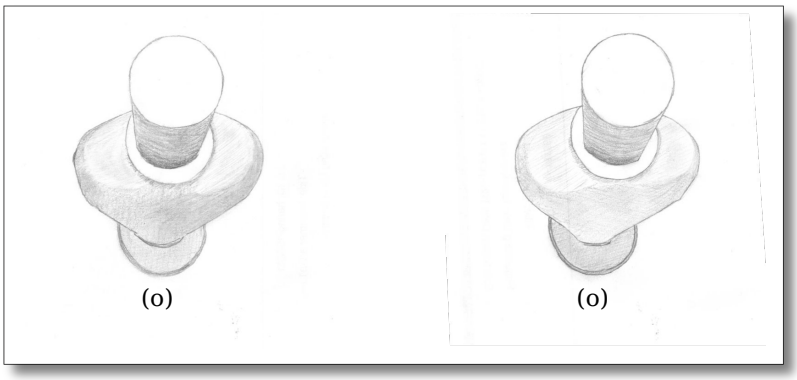


Figure 4.10: Free view stereoscopy. The reconstruction can be done by looking at the right image with the left eye and the left image with the right eye.

A stereo drawing prototype has been built by Vladimir Tamari[3DD,]. This stereo drawing device constraints two pencils with a fixed length. It allows precise creation of two drawings at the same time. A stereoscope is included to visualize the drawing without moving the creations. Although it favors precision, the comfort of drawing can be altered by such a machine. Furthermore, the size of the drawing is constrained to the size of the device, and the form-factor of the pen can also be limited. Consequently, the range of drawings is limited in size and style.

Conversely, we use a projection and vision system where the scaling factor can be easily tweaked to the desired creation. This system guides the artist, without limiting him or her to a given drawing support and tool.

4.5.3 Proposed system.

We extended the previous approach to enable 3D stereoscopic drawing. The 3D scene edition and visualization is now performed using a stereoscopic rendering, in our case using an anaglyph rendering. This is illustrated in Figure 4.11 on the left. Using this, the artist has a preview of the two drawings to create, one for the left eye and one for the right eye. Then, the left image is projected on a paper sheet, and the right image on another one. An overview of the system is illustrated in Figure 4.12.

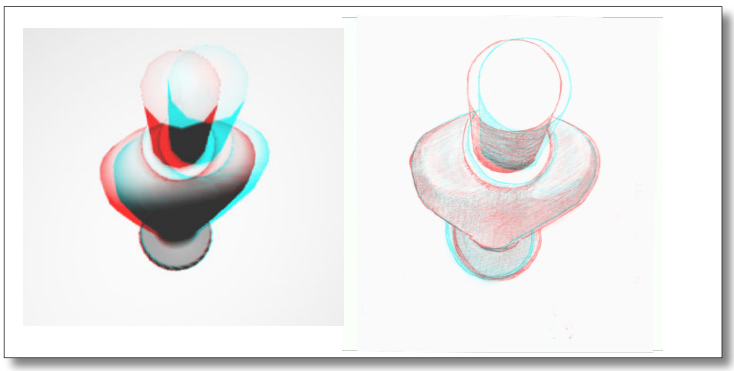


Figure 4.11: Left: Anaglyph view of the two projected images. Right: resulting drawings captured and colored and shifted for anaglyph visualization.

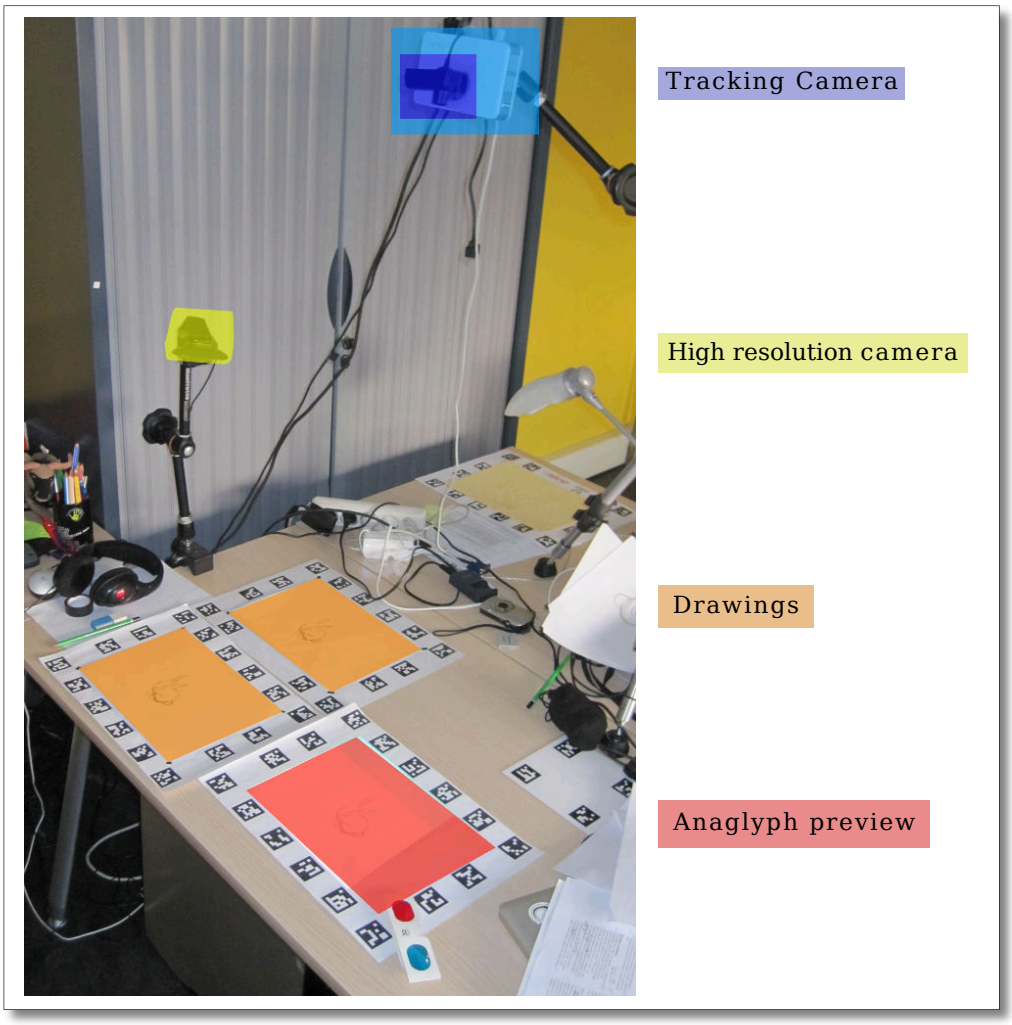


Figure 4.12: Free view stereoscopy. The reconstruction can be done by looking at the right image with the left eye and the left image with the right eye.

Realtime reprojection and adjustment

The difference between the left and right image have to be consistent for depth perception [Lambooij et al., 2007]. Furthermore, the drawings should be as similar as possible for the sharpest parts. Consequently, we created a tool to reproject the left rendering on the right drawing area, and vice versa. By shifting the projection on the left or on the right, the artist can examine different parts of the piece of art and correct parts of the drawings if necessary.

Consequently, the artist can draw the same line for both the right and left views, and compensate the copying errors, as illustrated in Figure 4.13. To achieve this, we use a high resolution camera in order to get a clear view of the drawing.

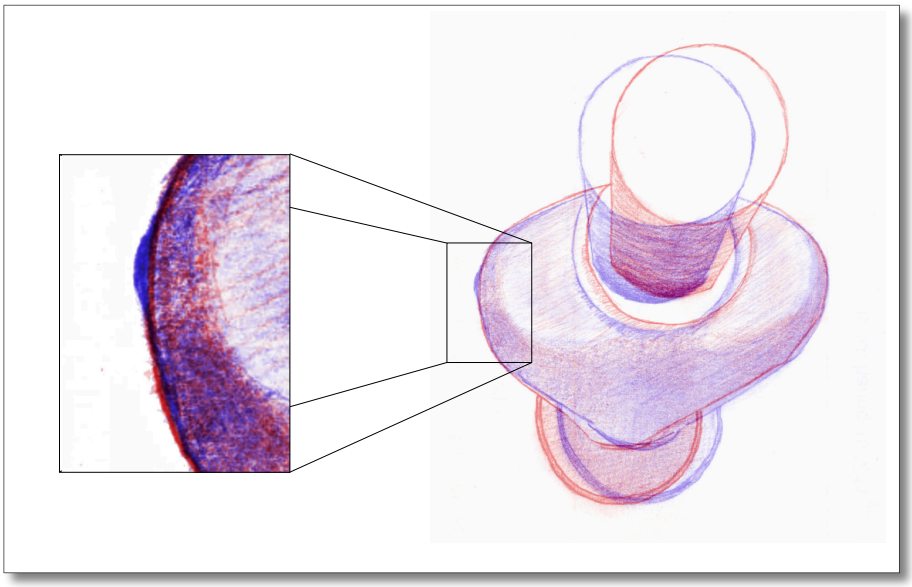


Figure 4.13: Details on an undesired left/right disparity. The left image, in blue, is shifted on the right, in order to match at a good depth.

Realtime stereoscopic visualization

The two drawings are captured by the high resolution camera. They can be combined to create a stereoscopic view, in our case, an anaglyph view (Figure 4.11 right). Alternatively, using a dedicated screen or a stereo projector, the artist could see the stereo drawing with true colors. This realtime visualization capabilities enables the artist to spot the differences between the drawings. He or she can correct them directly, or by using the technique described above.

In summary, we use projected images as the basis for the creation of drawings. Then in order to improve these drawings, our system offers the user to visualize the resulting stereo drawing as well as tools to ease the correction. In the next section, we propose some more advanced tools to ease the creation of stereoscopic drawings.

4.5.4 Perspectives

Automatic detection of disparities

In the current prototype, the drawings are based on two projections, one for each eye, of a 3D object. As each image has depth information, it can be used to track the depth disparities. The disparity error can be displayed onto each drawing, for example using color codes. The main advantage of this automated detection is that it can be used to correct the drawings, thus decreasing the visualization sickness such as headache or nausea that came from a bad stereoscopic scene.

Shape perception with stereoscopic drawings

We conducted some preliminary tests with different amateur artists, and different skill levels. Novice artists used the projected shading without adding shape information. The curvature of the object is badly represented. Conversely, a user who had taken drawing courses did the shading following the object's curvature. From one drawing to another, the stereoscopic experience is very different. The monoscopic shape perception created by stylized shading seems to have an important influence for the depth perception of the stereoscopic scene. Shape representation tends to lead to a better depth perception, which is not really surprising. Thus some studies could be conducted on shape and depth perception with stereoscopic drawings. It would provide guidelines for the artist, on how to create stereoscopic depth with other means than left/right disparities.

Creation from an existing 2D drawing

The creation of a stereoscopic drawing may rely on an existing drawing instead of a 3D object. In order to take full advantage of our tools, the user could use automatic [Gould et al., 2009] or semi-automatic depth retrieval method. He or she could set the depth of different part of the images, and a depth map would be generated. As a consequence, from a left image, a right image can be generated automatically. As the generated image will contain some occlusions errors and deformations, the tool can indicate the regions where the projected image may be incorrect.

Part III

Beyond Drawing

Chapter 5

Usage and Applications

During the three years of the PhD, I did many demonstrations to a wide range of people. From drawing creation with 6 years old children and with their grand parents, we went to demonstration to local government officials and research peers. In this section, we describe the applications developed for demo purposes, some of them were for a short period of time, and others were made to last for a few years. These demo applications show the variety of contexts where our SAR interface may benefit to end users.

5.1 Palais de la Découverte

The demonstration at the Palais de la Découverte lasted 3 months in Paris. It was part of an exhibition called "un chercheur, une manip"[1] where a researcher explains a part of a research problem and proposes to the general public to experiment with his or her research problem. In this demonstration, we explained for 10 to 15 minutes the problem with the public listening. It was followed by a 10 minutes questions and answers session. At the end of the demo, one drawing was demonstrated and a visitor could also make a drawing and go home with it. Some of the resulting drawings are displayed in Figure 5.1.

The demonstration was done with the application described in Chapter 4. It was the first general public exhibition we did, and most of the observations on drawing in Chapter 4 came from this time period. The projection setup was quite successful, as most visitors said it was "magical" because they did not really know what a projector nor a Kinect could do. The acceptability of this kind of system seems very high, because we do not force any new device and try to keep a strong link to the physical world.

We observed original behavior from kids at the Palais de la Découverte: one kid did not draw over the projection (of a dragon), he just added flames coming out of the mouth. His answer on why he did that : "The dragon was already there".

[1] http://www.palais-decouverte.fr/index.php?id=357

Figure 5.1: Left: Photo of the setup at the Palais de la Découverte. Right: Child drawing and some of the drawings made during the exhibit.

5.2 Rencontres Inria Industries

The "Rencontres Inria Industries" (RII) were held in Strasbourg in November, 21st 2012. The RII are a dedicated time for Inria to demonstrate and exchange with industries about its research. The theme in 2012 was Health. We created two demonstrations for the event, one enables the visualization of 3D skull and a simple manipulation of it using the touch interface, as illustrated in Figure 5.2.

The second one was about X-ray visualization. The idea of the application is to manipulate the image, and analyze it in real time with image filtering. We gathered a few images from Internet and we enabled the possibility to navigate in a 2D X-ray, to apply image filters on the image to display more information, to annotate physically on paper and take an image of the annotation and current view for remote collaboration. This application was the first to propose a use for the capture of drawing, moreover the menu selections were embedded in a numeric keypad. The ease of programming and selecting using the keypad lead to the projected keypad presented in Chapter 3.3.2. The RII lead to the beginning of a collaboration with a french company. This collaboration is not part of this thesis due to a non-disclosure agreement (NDA).

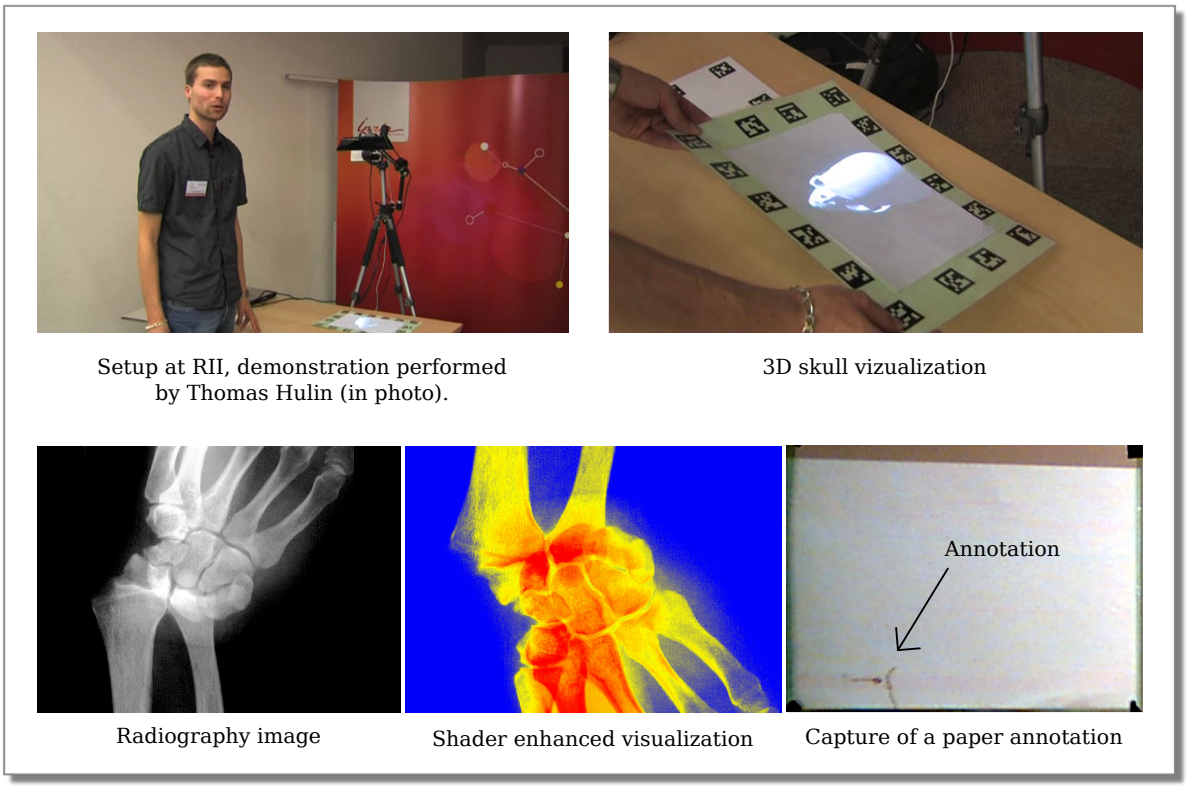


Figure 5.2: Demonstration at RII.

5.3 Semaine digitale

The "semaine digitale"[2] is an annual event in Bordeaux where local and national entities working in new technologies gather together. The 2013 theme was "la vi(ll)e de demain", about the future digital life and city. We created a 3D visualization of Bordeaux from an existing 3D model. The city was projected on a piece of paper, using the "pop-up anaglyph" technique described in 4.2. We proposed the visitors to see the 3D objects using anaglyph glasses and moving the piece of paper in front of them as illustrated in Figure 5.3. This demo was done to get more feedback on visualization using the pop-up anaglyph. Most of the users had small "whoa" exclamation of awe once they started to move the piece of paper. Indeed the model was quite detailed and provided a good impression holding a small model put on the piece of paper. Although the feedback were very positive, the lack of head tracking was a bigger limitation than for the previous applications. Indeed, the users were moving or sitting and the paper was also moving. However, we explained to the user how to find the correct location to view the model, and within a few seconds they found it.

[2] http://citedigitale.bordeaux.fr/

Figure 5.3: Demonstration at Bordeaux city hall for the "Semaine digitale", with Alain Juppé former prime minister and current mayor of Bordeaux.

5.4 Cap Sciences

Cap Sciences is a science museum and mediation center in Bordeaux. Our work is demonstrated in their "living lab"[3]: it is a dedicated space for research lab and companies to show and experiment with new technologies (Figure 5.4). We worked with Cap Sciences to replicate our system and install it for a general public use. Cap Sciences is one of the most visited cultural location in Bordeaux. They have important exhibitions such as the international Lascaux exhibition.

In prehistoric times, animals were painted on rock walls (for example, in the Lascaux caves) or engraved on various objects. In both cases, the drawings appear on surfaces with many concavities and convexities. Consequently, the perception of these drawings depends strongly on the viewer's and light source's positions. So that the exhibit visitors can understand these parameters' impact, we mounted an activity based on a 3D scan of an engraved reindeer antler. Visitors can change the viewpoint and lighting on the antler and reproduce the engraved image as illustrated in Figure 5.5.

Visitors have been enthusiastic about this piece of paper that magically responds to their touch. They generally don't pay attention to the technical setup hidden behind the table. With PapARt, users don't have to touch the technological hardware. Hitting the table or drawing on the paper doesn't affect the system's integrity. Another positive point is that visitors can keep the piece of paper on which they draw. Also, several people can collaborate around PapARt, which is fruitful in public exhibitions. The visitors have had no difficulties changing the view by moving the piece of paper or applying multitouch gestures. Similarly, they manage to change the light source's position with direct movement above the paper. With this interface, they can find a good viewpoint and light position for the engraving

[3] http://www.cap-sciences.net/pageseditos,423,left_77FBAA2B.html

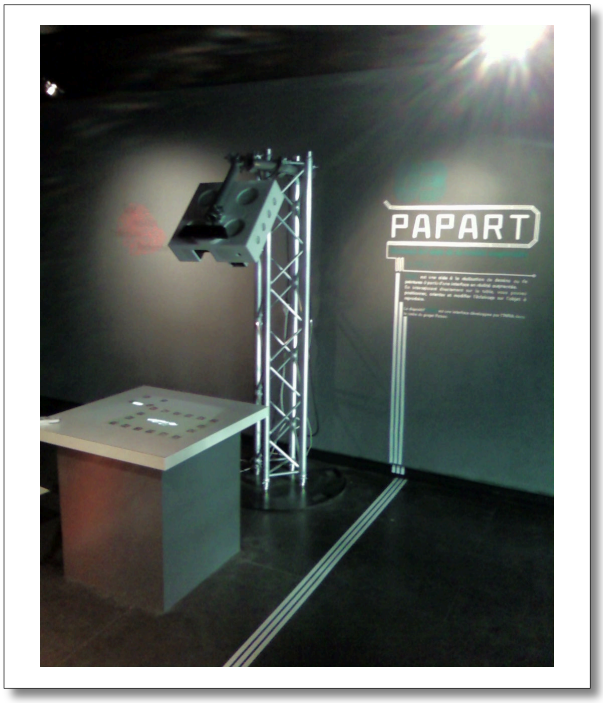


Figure 5.4: Living lab at Cap Sciences, the setup is called PapARt.

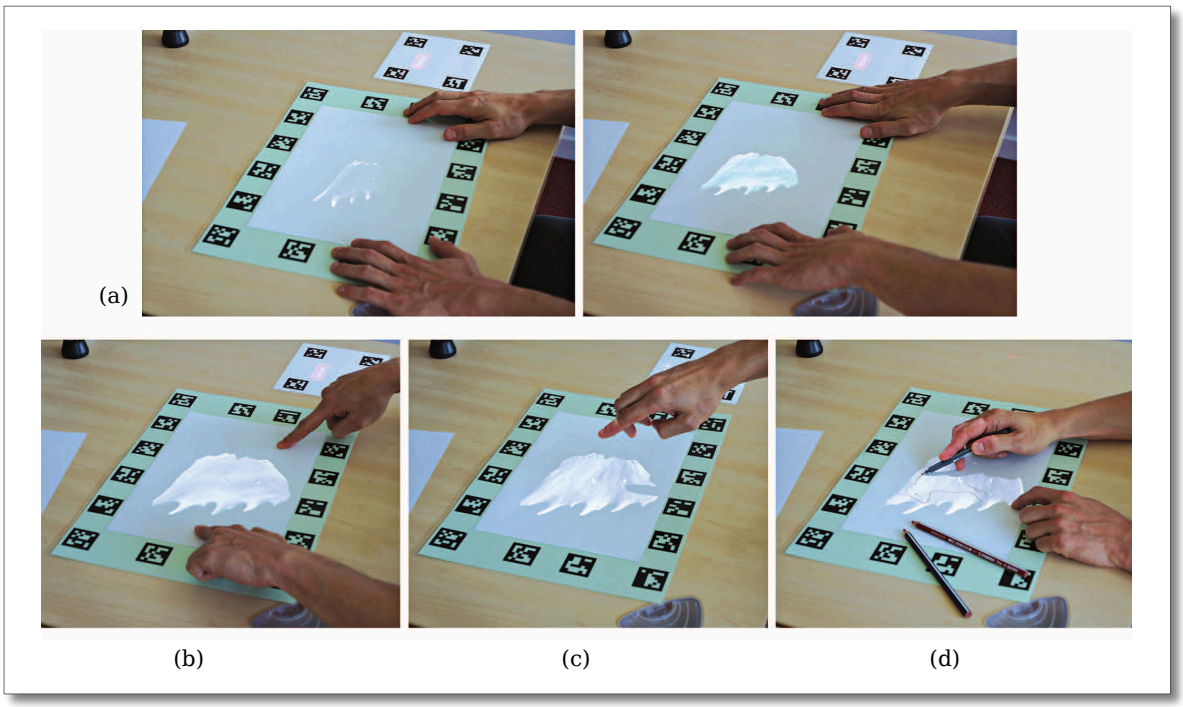


Figure 5.5: Application during the Lascaux exhibit. (a) Modifying the viewpoint by sliding the paper on the table. (b) A scaling operation using multi-touch gestures. (c) Modifying the light source through 3D spatial input. (d) Reproducing the image engraved on a deer antler, using freehand drawing during the exhibit.

The collaboration with Cap Sciences is very stimulating because of these events and we also obtain practical feedback from the employees performing the demonstrations of our system. It pushes us to create better user interfaces for the enjoyment of the visitors.

5.5 Marbotic

We explored the use of interactive projection to teach the alphabet letters to toddlers. It is a collaboration with Marbotic[4]: a company that creates tablet applications to teach the alphabet letters. The concept we explored is to enable the kids to trace the letters using projection on a piece of paper. We illustrate the kind of exercise we want to create in Figure 5.6. It may prove more effective than tracing on a screen because it also trains the hand-pen-eye coordination.

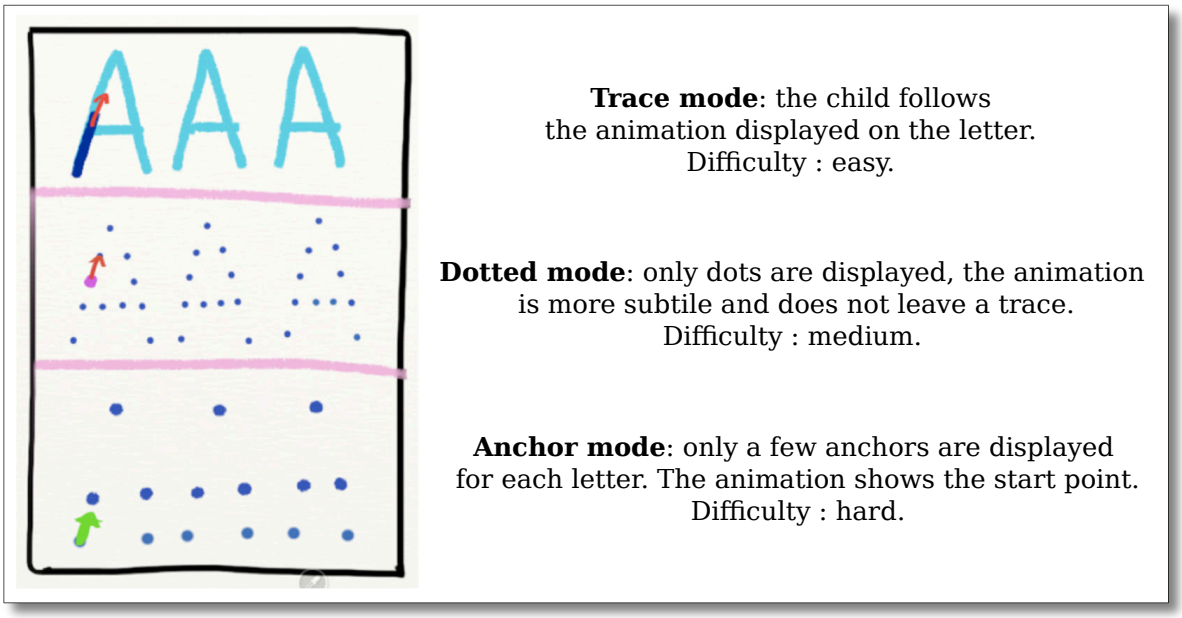


Figure 5.6: Example of application to teach how to draw letters with three different level of difficulty.

We developed a prototype that enables the projection of large letters and words on paper and we conducted an experiment to obtain feedback from children. The users were aged from 4 to 6 years old and they tried the system for an average of 20 minutes. The kids did not have any trouble interacting with the system, and they enjoyed copying the projection. The result showed that most of the kids had drawing training and they did what they usually do: trace the contours first, then fill the inside. It is obviously not what is desired when one learns to write, as the letters are only contours and never filled. This shows an important difference between the digital medium and projection: with projection, the task learned on paper also apply for projection on paper.

With the ubiquity of paper, the creation of paper interfaces has to overcome important biases taking into account the user's experience. We showed that tracing between thick projected

[4] http://www.marbotic.fr/

letters is not “natural”, and if we want it to work we need to project thin letters, according the the pen size. Another limitation is the lack of pen tracking, pen tracking is important for user guidance: where to start a letter and check if the user is tracing in the right direction. This tracking could be implemented using the camera to track the pen’s position on paper using image features. These two limitations are the main ones encountered for now. We believe they could be teared down with a better setup and more work on our computer vision algorithm.

Chapter 6

Towards Mixed Media Creation

In the previous chapters we focused on augmented reality applications. In this chapter, we describe our mixed media explorations. Instead of starting from an existing task, we propose new tasks such as coding applications for pieces of paper. Consequently, we do not augment reality with digital elements, we mix physical and digital possibilities.

6.1 Creative coding

The result of creating coding is art. The term usually encompass images, videos, sounds, and interactive creations that are generated by computer, or more specifically by code. Creative coding is coding for entertainment: entertainment of the programmer and eventually for the people seeing, hearing and/or interacting with the result.

At first it was available to the people who learned to code in complex programming languages. However, nowadays coding is available to everyone with simple programming languages such as Python for general purpose programming or Processing[pro, 2013] for creative coding. A complete history of creative coding is available in the book "Processing: Creative Coding and Computational Art" by Ira Greenberg [Greenberg, 2007] available for free online.

6.1.1 Processing

Processing was a project from two MIT students Casey Reas and Ben Fry, today they are still involved in the project and the second version of Processing is out since August 2013. It is made to be as simple as possible and it is powerful enough to rapidly code interactive software. It is built on top of Java and benefits from the many java libraries available. The choice of Processing to code all the applications at the beginning this PhD was with the hope that the result of this work could be easily used by a large number of people.

In this section we propose to use the power of Processing to create programs that will be displayed and interactive on sheets of paper. We extended the previous tool to aggregate images, videos and, programs together.

6.1.2 Code for the paper

Code for paper is different from code for screens. The main difference is about sizes. On a screen all the sizes are in pixels, on paper the sizes are both in meters and pixels. We propose some solutions to deal with this size problem. We keep the Processing variables for pixel sizes such as "width" and "height", and we add variables for millimeter sizes: "displayWidth" and "displayHeight" along with the variable "screenResolution" that links the two together. The difference between Processing and our API is illustrated in Figure 6.2. This mechanism provides a valid solution for simple applications. However, in our application composition tool the application can be moved and resized. It is necessary to elaborate a way to use fixed sizes in millimeters for some elements that require a valid physical size and a dynamic size for digital elements.

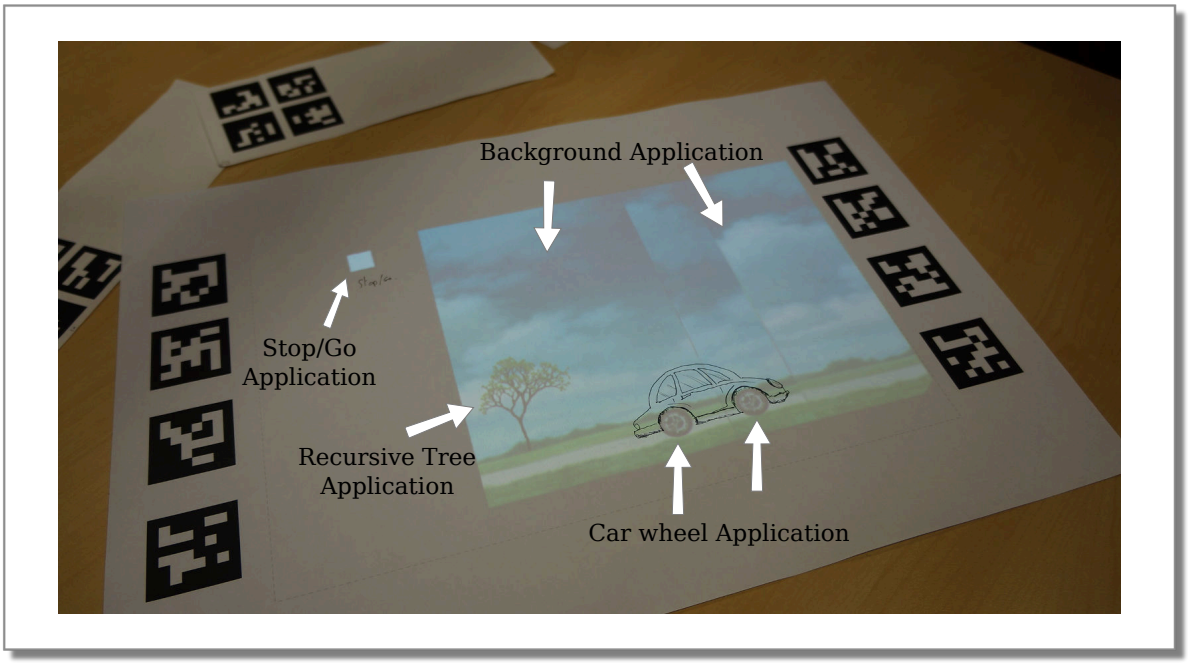


Figure 6.1: Example of application composition. The car is drawn on paper. The applications are either custom made or downloaded online (Recursive Tree comes from OpenProcessing.org, by Poersch).

We also propose to reuse the variables "mouseX" and "mouseY" to be used with the touch interface. If a touch event is detected inside the application, it is be send to these variables. The other touch events are ignored by the sketch. Moreover, it is possible to use the touch interface with tangible elements. Here is an example: we have two sketches, one for rotating wheel and the second for moving background that are part of a moving car design (as illustrated in Figure 6.1). We add a third one which is just composed of a "stop / go" button. When the button is pressed (or an object is put on it) the car goes, when it is released the car stops.

```
// Rotation car wheel, Paper version
PImage wheelImg = null;
class CarWheel extends SubSketch{
    public void setup(PApplet parent){
        // display size in millimeters
        this.displayWidth = 30;
        this.displayHeight = 30;
        // size in Pixels
        this.width = (int) (displayWidth * screenResolution);
        this.height = (int) (displayHeight * screenResolution);
        // initialize the sketch
        this.initSketch(parent);

        if(wheelImg == null)
            wheelImg = loadImage(sketchPath + "/wheel.png");
    }
    public void draw(){
        clear(0, 0); // instead of background to have a transparent background.
        scale(screenResolution); // screenResolution already set: nb of pixels per mm
        translate(displayWidth / 2, displayHeight /2, 0);
        rotate((float)millis() / 500f);
        translate(-displayWidth / 2, -displayHeight /2, 0);
        image(wheelImg, 0, 0, (int) displayWidth, (int) displayHeight);
    }
}
```

```
// Rotation car wheel, Processing version
PImage wheelImg = null;
void setup(PApplet parent){
    // display size in pixels
    size(400, 400, P2D);

    if(wheelImg == null)
        wheelImg = loadImage(sketchPath + "/wheel.png");
}
void draw(){
    background(0, 0);
    // scale(screenResolution);
    translate(width / 2, height /2, 0);
    rotate((float)millis() / 500f);
    translate(-width / 2, -height /2, 0);
    image(wheelImg, 0, 0, (int) width, (int) height);
}
```

Figure 6.2: Processing for paper code (top) and standard Processing code (bottom) for a rotating wheel application.

6.1.3 Ideas of applications

The animated car is a simple mockup of what is possible to do. With the current system the possibilities are already important for scrapbooking (like in [West et al., 2007]), interactive exploration and storytelling. In order to achieve a large set of applications we need to add the possibility for each program to communicate and know where the other programs are displayed and start new instances of programs. Using this, it is possible to imagine a program for "house", composed of sub-program "windows" and it can instantiate "fire" and "firefighter" programs for storytelling. This kind of programming enable simple prototyping of single applications with Processing and collaborative programming for rich composed applications.

The next step is the multi-medium integration: the possibility to use multiple pieces of paper for collaboration and large scale stories and games. The creation of applications on a dedicated piece of paper opens the creation of applications dedicated to a given object or set of objects. Ultimately the goal is to code for physical objects instead of just planar sheets of paper. However, nowadays the 3D scanning and tracking techniques are not mature enough for implementation without heavy instrumentation.

6.1.4 Implications on creativity

An interesting exploration of this work is its impact on one's creativity. The creation of Processing sketches for a screen, web browser or tablet is very different than from sketches for paper. The creativity on paper seems multiplied because of the direct link between the mind and the pen. From an idea such as: "I wonder if a circle there look good" the realization is instantaneous. Once it is done, it can be coded, and enhanced "many concentric circles is better !". A study on the impact of creative coding on paper instead of screen would be interesting to conduct to obtain clear answers on this.

6.2 Image acquisition for drawing

6.2.1 Animation

Traditional animation consists in the creation of many drawings that when viewed sequentially create an illusion of movement. In addition to the artistic skills, it requires to create many drawings and it is extremely time-consuming. Nowadays, the animation studios still have either hand-drawn animations or use computer-generated animations. The research focus on computer animation to improve interpolation between key-frames [Whited et al., 2010] and simulate plausible deformations [Sýkora et al., 2009]. In this work, we propose some ideas and a proof-of-concept implementation of tools to assist traditional animation.

6.2.2 Proof-of-concept implementation

We created a tool to create animations with four drawings. The first drawing "background" (see Figure 6.3) is a non-animated part that can be used to aid the animation. The "background" is inverted and projected in Frames 1 to 3. The user can then draw elements in the three frames. After a key press the animation is ready and rendered in the "Animation" square.

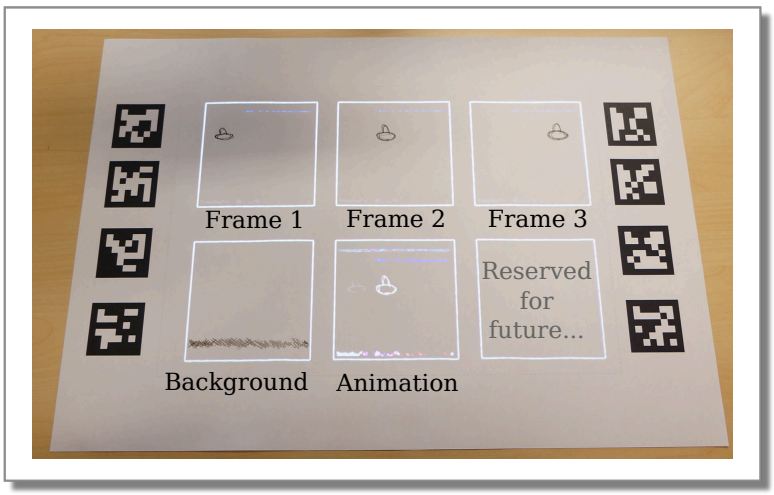


Figure 6.3: Photo of the proof-of-concept implementation.

This simple example is enough to grasp the challenges for the creation of a complete animation tools. The first difficulty comes from computer vision analysis: the drawings are subtractive colors and the projection is additive. Consequently, the reprojection of captured elements is a challenge in itself. Moreover the capture of the drawn elements require the detection of them. Detection of light lines is a complex task in an uncontrolled environment: small light changes will impact the quality of the capture. Once again, there is no work in the capture of drawings in partially controlled light environment (low intensity changes).

From the interaction point of view, the specification of the elements to reproject and the possibility to partially copy frames is a challenge. Moreover, the question of the space management is also compelling: the management of large quantities of animation frames withing a reasonably small amount of space can be problematic. Finally, these animations needs to be combined and edited and some adjustment may be required off-line with a dedicated software.

6.2.3 Physical/Digital image composition

In the previous chapters we explored projection to create purely physical drawing, here we propose to use projection to texture a drawing. With the possibility to capture the drawing, it is possible to use it with image edition software such as The Gimp [1]. Consequently, one can create elements to project that will exactly fit the drawing. We demonstrate this with a example illustrated in Figure 6.4: a photo is projected (a) and a drawing is made from it (b). Then, some elements of the photo are reused to texture the drawing and synthetic elements such as the blue wheels are added (c). Finally, the mixed drawing is composed of drawn and projected elements (d). We pushed further the integration of physical and digital drawing by adding a digital pen tablet that enable the projection of precise digital strokes.

The pen and tablet integration seemed promising for the creation of mixed visual art. However, it will require an integration with digital creation software to explore further the possibilities. The users liked the possibility to create both physical and digital strokes with a pen.

[1] The gimp. http://www.gimp.org.

More explorations about this will point out the subtle differences between the physical and digital strokes. Some are obvious, such as the friction of the pen and the possibilities offered either by physical and digital pen. However, there are still many questions about the use of both for construction lines such as: “Does the physical and digital strokes have different roles, do they complement or supplement each other ?”. The possibilities of mixed creation are just emerging, and hopefully it could be a new media which combines the flexibility of projection mapping and the legacy of thousand of years of visual art.

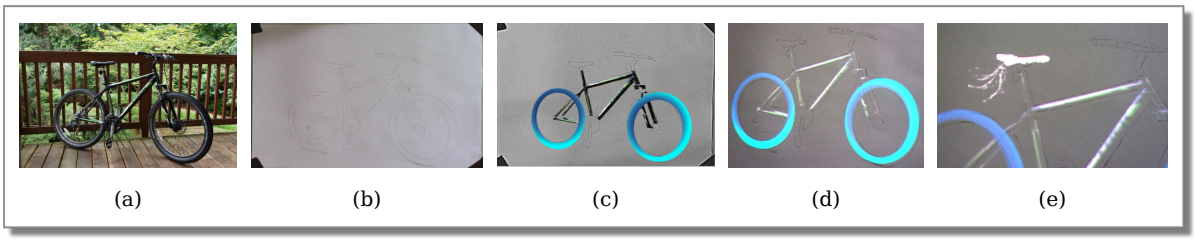


Figure 6.4: From a photo (a), a drawing is created (b), then edited using The Gimp (c). The new layers are projected on paper (d). The projection can be precisely edited using a pen tablet (e).

Conclusion

"Judge your success by what you had to give up in order to get it."

Tenzin Gyatso, 14th Dalai Lama.

The goal of this thesis is the creation of physical and mixed reality visual art. In order to assist this creation, we created a spatial augmented reality (SAR) system which enables precise projection onto sheets of paper. In addition to its precision, our SAR system enables touch and hovering detection using a depth camera. Therefore, we can create touch screens out of pieces of paper.

Touch from a depth camera is different from usual touch screens. Consequently, we explored interaction metaphors that take advantage of the new possibilities offered by depth cameras such as 3D pointing and the detection of hand pose. Moreover, we proposed two input methods for menu selections: on a piece of paper and an augmented keypad. Furthermore, we discussed the problems with the creation of rich tabletop paper interfaces.

The first experiments on image projection for drawing were encouraging. We enabled the possibility to analyze a projected image directly on the paper to assist the artist into his decision on which part of the projection to copy. Our main contribution for drawing is the possibility to create digital construction lines. These construction lines have advantages over the physical ones: they can be easily editable, erased, and they can be projected over wet paint or any other fragile medium. We demonstrated through a user study that they could replace physical construction lines in our case, and they lead to results with less traces of erased lines. We extended the generic construction line tool to two dedicated construction line tools for perspective and character drawing.

In the fourth chapter, we proposed a way to visualize and interact with a 3D model projected on a sheet of paper. It is possible to manipulate a 3D scene directly on paper: to move, rotate, scale, and light it. The rendering can be done using non-photorealistic techniques to ease and fasten the drawing. We extended this work to enable stereographic drawings. Our solution ease the stereo creation, it enables an anaglyph visualization while drawing, and techniques to find left/right disparities.

Finally, we explored the creation of interactive mixed media such as tools to create traditional animation and to create digital textures for drawings. In order to explore all the possibilities

of this media, we also propose to code applications for the paper using a simple programming language. Thorough this thesis, we described SAR interfaces and we proposed some solutions specific to our applications. The use of touch and tangible has been explored on screens, but there are still many questions for applications that displayed only on pieces of paper on a table and the table itself. The next steps of this work is to make any surface interactive and displayable; and to offer the possibly to easily create applications and textures for any existing object.

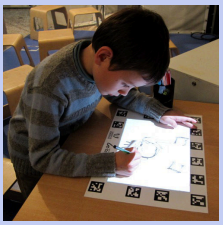
The creation of applications that are only on sheets of paper felt "magical" for many of the users who tried our system. We conducted a few user studies and many demonstrations and the overall appreciation was very good. SAR technologies well accepted by the general public because they are not intrusive and provide a real "augmented reality feeling".

In conclusion, in this thesis we proposed an exploration on SAR. Starting from an existing task, we added digital elements firstly to ease the task, then to create mixed output that require both the physical and digital part. We encourage all augmented reality students and researchers to continue to push the boundaries of the physical world with augmented reality, so that we will spend more time working and communicating with humans instead of watching screens.

Perspectives

Chapters 2 & 5

Spatial Augmented Reality System

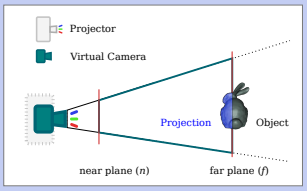


User-friendly Multi-touch calibration
Documentation on the calibration - Packaging
Public release of the Multi-touch API
Public release of the AR and SAR rendering API

Multi-touch on moving surfaces
Complete Depth camera - projector calibration

Demonstrations to show the possibilities of SAR

Pen tracking to guide the strokes
Study on performance for teaching: SAR vs Tablets



Chapters 3 & 4

Augmented Drawing Applications

User study on Digital vs Physical construction lines
User study on the use of image filters for drawing

Explorations on drawing from a 3D model vs Photo
Complete user-friendly construction line tools
Automatic generation of construction lines

Creation of advanced touch metaphors for depth cameras
Studies about the size and number of paper interfaces

Perception study for stereographic drawings
Creation and evaluation of partial drawings for stereo

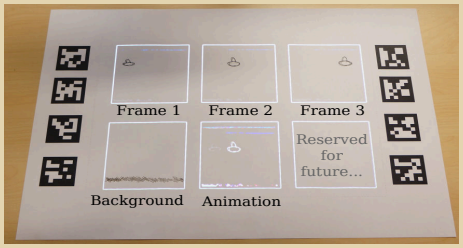
Elaboration of shaders dedicated to drawing

Chapter 6

Mixed Media Creation

Integration via plug-in for The Gimp
Image analysis to detect the pen strokes.
Animation software to elaborate

Integration with Processing for Creative Coding
Study on creativity with paper applications
API to create interactive planar objects
Extension to make interactive any object



Perspective on the missing parts of this thesis : Creation of a prototype for Healthcare project
User study on new interaction device
(from Microsoft Research internship)

Bibliography

[3DD,] 3dd by vladimir tamari. http://www.ne.jp/asahi/tamari/vladimir/3dd.html.

[pro, 2013] (2013). Processing. http://processing.org/about/.

[Audet and Okutomi, 2009] Audet, S. and Okutomi, M. (2009). A user-friendly method to geometrically calibrate projector-camera systems. In *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, pages 47–54. IEEE.

[Audet et al., 2010] Audet, S., Okutomi, M., and Tanaka, M. (2010). Direct image alignment of projector-camera systems with planar surfaces. In *Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco*, pages 303–310. IEEE.

[Audet et al., 2012] Audet, S., Okutomi, M., and Tanaka, M. (2012). Augmenting moving planar surfaces interactively with video projection and a color camera. In *Virtual Reality Workshops (VR), 2012 IEEE*, pages 111–112. IEEE.

[Baudisch et al., 2001] Baudisch, P., Good, N., and Stewart, P. (2001). Focus plus context screens: combining display technology with visualization techniques. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 31–40. ACM.

[Benko et al., 2012] Benko, H., Jota, R., and Wilson, A. (2012). Miragetable: freehand interaction on a projected augmented reality tabletop. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 199–208. ACM.

[Bimber et al., 2005] Bimber, O., Raskar, R., and Inami, M. (2005). *Spatial augmented reality*. AK Peters Wellesley.

[Bradski, 2000] Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

[Casiez et al., 2012] Casiez, G., Roussel, N., and Vogel, D. (2012). 1€ filter: a simple speed-based low-pass filter for noisy input in interactive systems. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, pages 2527–2530. ACM.

[Cohn et al., 2012] Cohn, G., Morris, D., Patel, S., and Tan, D. (2012). Humantenna: using the body as an antenna for real-time whole-body interaction. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, pages 1901–1910. ACM.

[Cole et al., 2008] Cole, F., Golovinskiy, A., Limpaecher, A., Barros, H., Finkelstein, A., Funkhouser, T., and Rusinkiewicz, S. (2008). Where do people draw lines? In *ACM SIGGRAPH 2008 papers*, pages 1–11. ACM.

[Coudray,] Coudray, P. Phillipe coudray's website. http://www.philippe-coudray.com/Pages/Dessins%20stereo.html.

[DeepSketch,] DeepSketch. Deepsketch 3d sketchbook for ipad. http://machineboy.com/blog/apps/deepsketch/.

[DuBosque, 2009] DuBosque, D. (2009). *Dessiner pas a pas*. Evergreen, Koln.

[Edigo, 2010] Edigo, C. (2010). *Dessin Fiches pratiques (French)*. Edigo.

[Fitzmaurice et al., 1995] Fitzmaurice, G. W., Ishii, H., and Buxton, W. A. (1995). Bricks: laying the foundations for graspable user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 442–449. ACM Press/Addison-Wesley Publishing Co.

[Flagg and Rehg, 2006] Flagg, M. and Rehg, J. (2006). Projector-guided painting. In *Proceedings of the 19th annual ACM symposium on User interface software and technology*, pages 235–244. ACM.

[Gellersen and Block, 2012] Gellersen, H. and Block, F. (2012). Novel interactions on the keyboard. *Computer*, 45(4):36–40.

[Gernsheim and Gernsheim, 1969] Gernsheim, H. and Gernsheim, A. (1969). *The History of Photography: From the Camera Obscura to the Beginning of the Modern Era*. Thames & Hudson.

[Gould et al., 2009] Gould, S., Fulton, R., and Koller, D. (2009). Decomposing a scene into geometric and semantically consistent regions. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 1–8. IEEE.

[Greenberg, 2007] Greenberg, I. (2007). *Processing: Creative Coding and Computational Art*. friendsofED.

[Guimbretière, 2003] Guimbretière, F. (2003). Paper augmented digital documents. In *Proceedings of the 16th annual ACM symposium on User interface software and technology*, pages 51–60. ACM.

[Hachette, 2009] Hachette (2009). *Cours de dessin faciles (French Edition)*. HACHETTE COLLECTION.

[Hagbi et al., 2009] Hagbi, N., Bergig, O., El-Sana, J., and Billinghurst, M. (2009). Shape recognition and pose estimation for mobile augmented reality. In *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*, pages 65–71. IEEE.

[Harrison et al., 2011] Harrison, C., Benko, H., and Wilson, A. (2011). Omnitouch: wearable multitouch interaction everywhere. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 441–450. ACM.

[Hartley, 2008] Hartley, R. (2008). *Multiple view geometry in computer vision*. Cambridge university press.

[Hilliges et al., 2012] Hilliges, O., Kim, D., Izadi, S., Weiss, M., and Wilson, A. (2012). Holodesk: direct 3d interactions with a situated see-through display. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, pages 2421–2430. ACM.

[Holman and Vertegaal, 2008] Holman, D. and Vertegaal, R. (2008). Organic user interfaces: designing computers in any way, shape, or form. *Communications of the ACM*, 51(6):48–55.

[Holman et al., 2005] Holman, D., Vertegaal, R., Altosaar, M., Troje, N., and Johns, D. (2005). Paper windows: interaction techniques for digital paper. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 591–599. ACM.

[Iarussi et al.,] Iarussi, E., Bousseau, A., and Tsandilas, T. The Drawing Assistant: Automated Drawing Guidance and Feedback from Photographs. To be presented at ACM Symposium on User Interface Software and Technology (UIST) - 2013 This work was funded in part by Autodesk (research and software donations) and by ANR-12-JS02-003-01 DRAO.

[Ishii et al., 2012] Ishii, H., Lakatos, D., Bonanni, L., and Labrune, J.-B. (2012). Radical atoms: beyond tangible bits, toward transformable materials. *interactions*, 19(1):38–51.

[Izadi et al., 2011] Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., et al. (2011). Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 559–568. ACM.

[Jota et al., 2013] Jota, R., Ng, A., Dietz, P., and Wigdor, D. (2013). How fast is fast enough?: a study of the effects of latency in direct-touch pointing tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 2291–2300, New York, NY, USA. ACM.

[Kato and Billinghurst, 1999] Kato, H. and Billinghurst, M. (1999). Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Augmented Reality, 1999.(IWAR'99) Proceedings. 2nd IEEE and ACM International Workshop on*, pages 85–94. IEEE.

[Kim et al., 2012] Kim, D., Hilliges, O., Izadi, S., Butler, A. D., Chen, J., Oikonomidis, I., and Olivier, P. (2012). Digits: freehand 3d interactions anywhere using a wrist-worn gloveless sensor. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, pages 167–176. ACM.

[Kim et al., 2013] Kim, H.-J., Kim, H., Chae, S., Seo, J., and Han, T.-D. (2013). Ar pen and hand gestures: a new tool for pen drawings. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 943–948. ACM.

[Knoedel and Hachet, 2011] Knoedel, S. and Hachet, M. (2011). Multi-touch RST in 2D and 3D Spaces: Studying the Impact of Directness on User Performance. In *IEEE 3DUI*.

[Lambooij et al., 2007] Lambooij, M., IJsselsteijn, W., and Heynderickx, I. (2007). Visual discomfort in stereoscopic displays: a review. *Stereoscopic Displays and Virtual Reality Systems XIV*, 6490(1).

[Lee et al., 2013] Lee, J., Olwal, A., Ishii, H., and Boulanger, C. (2013). Spacetop: integrating 2d and spatial 3d interactions in a see-through desktop environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 189–192. ACM.

[Lepetit and Fua, 2005] Lepetit, V. and Fua, P. (2005). *Monocular model-based 3d tracking of rigid objects: A survey*. Now Publishers Inc.

[Loomis, 1949] Loomis, A. (1949). *Figure Drawing for All It's Worth*. The Viking Press; 1st edition.

[Mackay, 1998] Mackay, W. E. (1998). Augmented reality: linking real and virtual worlds: a new paradigm for interacting with computers. In *Proceedings of the working conference on Advanced visual interfaces*, pages 13–21. ACM.

[Maimone and Fuchs, 2012] Maimone, A. and Fuchs, H. (2012). Reducing interference between multiple structured light depth sensors using motion. In *Virtual Reality Workshops (VR), 2012 IEEE*, pages 51–54. IEEE.

[Mellado et al., 2012] Mellado, N., Barla, P., Guennebaud, G., Reuter, P., and Schlick, C. (2012). Growing Least Squares for the Continuous Analysis of Manifolds in Scale-Space. *Computer Graphics Forum*.

[Milgram and Kishino, 1994] Milgram, P. and Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329.

[Mistry and Maes, 2009] Mistry, P. and Maes, P. (2009). Sixthsense: a wearable gestural interface. In *ACM SIGGRAPH ASIA 2009 Sketches*, pages 1–1. ACM.

[Molyneaux et al., 2012] Molyneaux, D., Izadi, S., Kim, D., Hilliges, O., Hodges, S., Cao, X., Butler, A., and Gellersen, H. (2012). Interactive environment-aware handheld projectors for pervasive computing spaces. In *Pervasive Computing*, pages 197–215. Springer.

[Ponce et al., 2011] Ponce, J., Forsyth, D., Willow, E.-p., Antipolis-Méditerranée, S., d'activité RAweb, R., Inria, L., and Alumni, I. (2011). Computer vision: a modern approach. *Computer*, 16:11.

[Puce et al., 1996] Puce, A., Allison, T., Asgari, M., Gore, J. C., and McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *The Journal of Neuroscience*, 16(16):5205–5215.

[Raskar et al., 1998] Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L., and Fuchs, H. (1998). The office of the future: A unified approach to image-based modeling and spatially immersive displays. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 179–188. ACM.

[Ren et al., 2011] Ren, Z., Yuan, J., and Zhang, Z. (2011). Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera. In *Proceedings of the 19th ACM international conference on Multimedia*, pages 1093–1096. ACM.

[Rivers et al., 2012] Rivers, A., Adams, A., and Durand, F. (2012). Sculpting by numbers. *ACM Transactions on Graphics (TOG)*, 31(6):157.

[Ronin, 2011] Ronin, G. (2011). *Dessiner - pas a pas.* Larousse.

[Sanmiguel, 2009] Sanmiguel, D. (2009). *Cours complet de Dessin.* Dessain et Tolra.

[Sato et al., 2012] Sato, M., Poupyrev, I., and Harrison, C. (2012). Touché: enhancing touch interaction on humans, screens, liquids, and everyday objects. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, pages 483–492. ACM.

[Schmidt et al., 2009] Schmidt, R., Khan, A., Singh, K., and Kurtenbach, G. (2009). Analytic drawing of 3d scaffolds. In *ACM Transactions on Graphics (TOG)*, volume 28, page 149. ACM.

[Segal et al., 1992] Segal, M., Korobkin, C., Van Widenfelt, R., Foran, J., and Haeberli, P. (1992). Fast shadows and lighting effects using texture mapping. In *ACM SIGGRAPH Computer Graphics*, volume 26, pages 249–252. ACM.

[Sellen and Harper, 2003] Sellen, A. and Harper, R. (2003). *The myth of the paperless office.* The MIT Press.

[Sodhi et al., 2012] Sodhi, R., Benko, H., and Wilson, A. (2012). Lightguide: projected visualizations for hand movement guidance. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, pages 179–188. ACM.

[Spindler et al., 2012] Spindler, M., Büschel, W., and Dachselt, R. (2012). Use your head: tangible windows for 3d information spaces in a tabletop environment. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*, pages 245–254. ACM.

[Steimle et al., 2013] Steimle, J., Jordt, A., and Maes, P. (2013). Flexpad: a highly flexible handheld display. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 2873–2874. ACM.

[Sutherland, 1965] Sutherland, I. E. (1965). The ultimate display. *Multimedia: From Wagner to virtual reality.*

[Sutherland, 1968] Sutherland, I. E. (1968). A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, pages 757–764. ACM.

[Sýkora et al., 2009] Sýkora, D., Dingliana, J., and Collins, S. (2009). As-rigid-as-possible image registration for hand-drawn cartoon animations. In *Proceedings of International Symposium on Non-photorealistic Animation and Rendering*, pages 25–33.

[Takahashi et al., 2012] Takahashi, Y., Matoba, Y., and Koike, H. (2012). Fluid surface: interactive water surface display for viewing information in a bathroom. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*, pages 311–314. ACM.

[Vergne et al., 2010] Vergne, R., Pacanowski, R., Barla, P., Granier, X., and Schlick, C. (2010). Radiance scaling for versatile surface enhancement. In *Proceedings of the 2010 ACM SIGGRAPH symposium on Interactive 3D Graphics and Games*, pages 143–150. ACM.

[Wagner and Schmalstieg, 2007] Wagner, D. and Schmalstieg, D. (2007). Artoolkitplus for pose tracking on mobile devices. In *Computer Vision Winter Workshop*, pages 6–8. Citeseer.

[Wang and Popović, 2009] Wang, R. Y. and Popović, J. (2009). Real-time hand-tracking with a color glove. In *ACM Transactions on Graphics (TOG)*, volume 28, page 63. ACM.

[Wellner, 1993] Wellner, P. (1993). Interacting with paper on the digitaldesk. *Communications of the ACM*, 36(7):87–96.

[West et al., 2007] West, D., Quigley, A., and Kay, J. (2007). Memento: a digital-physical scrapbook for memory sharing. *Personal and Ubiquitous Computing*, 11(4):313–328.

[Whited et al., 2010] Whited, B., Noris, G., Simmons, M., Sumner, R. W., Gross, M., and Rossignac, J. (2010). Betweenit: An interactive tool for tight inbetweening. In *Computer Graphics Forum*, volume 29, pages 605–614. Wiley Online Library.

[Wigdor et al., 2011] Wigdor, D., Benko, H., Pella, J., Lombardo, J., and Williams, S. (2011). Rock & rails: extending multi-touch interactions with shape gestures to enable precise spatial manipulations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1581–1590. ACM.

[Wilson, 2007] Wilson, A. (2007). Depth-sensing video cameras for 3d tangible tabletop interaction. In *Horizontal Interactive Human-Computer Systems, 2007. TABLETOP'07. Second Annual IEEE International Workshop on*, pages 201–204. IEEE.

[Wilson and Benko, 2010] Wilson, A. and Benko, H. (2010). Combining multiple depth cameras and projectors for interactions on, above and between surfaces. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, pages 273–282. ACM.

[Wilson et al., 2012] Wilson, A., Benko, H., Izadi, S., and Hilliges, O. (2012). Steerable augmented reality with the beamatron. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, pages 413–422. ACM.

[Wilson, 2010] Wilson, A. D. (2010). Using a depth camera as a touch sensor. In *ACM International Conference on Interactive Tabletops and Surfaces*, ITS '10, pages 69–72, New York, NY, USA. ACM.

[Yilmaz et al., 2006] Yilmaz, A., Javed, O., and Shah, M. (2006). Object tracking: A survey. *Acm Computing Surveys (CSUR)*, 38(4):13.