

**Federal University of Bahia
State University of Feira de Santana**

MASTER THESIS

A multi-view environment for markerless augmented reality

Caio Sacramento de Britto Almeida

Master in Computer Science – MMCC

Salvador
December 15th, 2014

MMCC-Msc-0007

CAIO SACRAMENTO DE BRITTO ALMEIDA

**A MULTI-VIEW ENVIRONMENT FOR MARKERLESS
AUGMENTED REALITY**

Thesis submitted to the Master Program in Computer Science from Federal University of Bahia and State University of Feira de Santana, in partial fulfillment of the requirements for the degree of Master in Computer Science.

Advisor: Antônio Lopes Apolinário Júnior

Salvador
December 15th, 2014

Index card.

Sacramento de Britto Almeida, Caio

A multi-view environment for markerless augmented reality/ Caio Sacramento de Britto Almeida– Salvador, December 15th, 2014.

14p.: il.

Advisor: Antônio Lopes Apolinário Júnior.

Thesis (master)– Federal University of Bahia, Institute of Mathematics, December 15th, 2014.

1. Multi-view environment. 2. Augmented reality. 3. Computer graphics.

I. Apolinário Jr., Antônio Lopes. II. Federal University of Bahia. Institute of Mathematics. III A multi-view environment for markerless augmented reality.

CDD 20.ed. 123.45

APPROVAL SHEET**CAIO SACRAMENTO DE BRITTO ALMEIDA****A MULTI-VIEW ENVIRONMENT FOR
MARKERLESS AUGMENTED REALITY**

This thesis was considered worthy of acceptance of the master's degree in Computer Science and approved on its final form by the UFBA-UEFS Master Program in Computer Science.

Salvador, December 15th, 2014

Prof. Dr. Antônio Lopes Apolinário Júnior 1
Federal University of Bahia - Brazil

Prof. Dr. Michelle Ângelo 2
State University of Feira de Santana - Brazil

Prof. Dr. Rodrigo Silva 3
Federal University of Juiz de Fora - Brazil

ACKNOWLEDGEMENTS

I would like to thank my family for all the support during all those years, not only the time spent on the master program, but also during all my graduation. Namely, first I would like to thank my brother, Rodrigo, for being my inspiration on following this career; my twin sister, Thalita, for being my support of all times; my parents, for the great education that was given to me; my grandmother Lourdes; my girlfriend Jéssica; my friends; and specially my advisor Antônio Apolinário for relying on me to make this work happen.

ABSTRACT

Augmented reality is a technology which allows 2D and 3D computer graphics to be aligned or registered with scenes of the real-world in real-time. This projection of virtual images requires a reference in the captured real image, which is often achieved by using one or more markers. But, there are situations where using markers can be unsuitable, like medical applications, for example. In this work, we present a multi-view environment, composed by augmented reality glasses and two Kinect devices, which doesn't use fiducial markers in order to run augmented reality applications. All devices are calibrated according to a common reference system, and then the virtual models are transformed accordingly too. In order to achieve that, two approaches were specified and implemented: one based on one Kinect plus optical flow and accelerometer data from augmented reality glasses, and another one based purely on two Kinect devices. The results regarding quality and performance achieved by these two approaches are presented and discussed, as well as a comparison between them.

Keywords: augmented reality, augmented reality glasses, kinect, transformation, optical flow, markerless

CONTENTS

Chapter 1—Introduction	1
1.1 Motivation	1
1.2 Objectives	2
1.2.1 Features	3
1.3 Thesis overview	4
Chapter 2—Conceptual primer	5
2.1 Augmented reality	5
2.1.1 Fiducial markers	5
2.1.2 Markerless	5
2.1.3 Direct or indirect vision	5
2.2 Cameras	5
2.2.1 Calibration	5
2.2.2 Parameters	5
2.2.3 Multi-view	5
2.3 Sensor-based registration approach	5
2.3.1 Kinect	5
2.3.2 Registration	5
2.3.3 Transformation	5
2.4 Vision-based registration approach	5
2.4.1 Optical flow	5
2.4.2 Lucas-Kanade algorithm	5
Chapter 3—Related work	7
3.1 Markerless augmented reality	7
3.2 Multiple Kinects	7
3.3 Multi-view environment	7
Chapter 4—Solution architecture	9
4.1 Environment	9
4.2 Calibration	9
4.2.1 Augmented reality glasses calibration	9
4.2.2 Initial calibration between Kinects	9
4.2.3 Initial calibration between Kinect and glasses	9

4.3	Communication	9
4.4	Transformations	9
4.5	Method 1: Glasses accelerometer and one Kinect	9
4.6	Method 2: Two Kinects	9
4.7	Hybrid approach	9
Chapter 5—Result		11
5.1	Scope	11
5.2	Analysis	11
5.3	Comparison	11
Chapter 6—Conclusions		13
6.1	Future work	13

LIST OF FIGURES

1.1	Fiducial marker used to represent a tridimensional model over it (??)	2
1.2	Global vision	3

LIST OF TABLES

Chapter

1

In this chapter I present the motivation, objectives and overview of this work.

INTRODUCTION

1.1 MOTIVATION

Augmented reality has taken advantage from the progresses on the fields of multimedia and virtual reality, making feasible new forms of interaction between humans and machines. Differently from virtual reality, that takes the user to a virtual environment, the augmented reality keeps the user on his physical environment and takes the virtual environment to the user's space, allowing the interaction with the virtual world, in a more natural manner and without needing training or adaptation (??). Many times, this interaction means merging virtual images with images captured from a real environment.

One of the greatest challenges on the field of augmented reality is to determine, in real time, which virtual image to be displayed, in which position and how it should be represented. In order to obtain an integration illusion between real objects and virtual objects, the generated object should be aligned with the tridimensional position and orientation of the real objects (??). This can be achieved by estimating the camera position.

On many situations, fiducial markers¹ are used (often this is due to the real time requirements of the augmented reality applications) (??) and are drawn in a way that they can be easily identified. Those markers need to be placed on the target scene and can achieve great results using just a few computational resources. Figure 1.1 shows the usage of a fiducial marker and a tridimensional object being projected over it.

However, besides requiring human interference on the scene, there situations where the usage of fiducial markers is not possible, feasible or comfortable for the target model. That is the case, for example, of medical applications on which this model is a patient. It's also possible to cite other limitations of fiducial markers, like, for example, occlusion (a virtual image could be not projected if the marker is not completely visible) and illumination (the intensity of light reflected by the marker could make it hard to be

¹A fiducial marker or fiducial is an object placed in the field of view of an imaging system which appears in the image produced, for use as a point of reference or a measure.

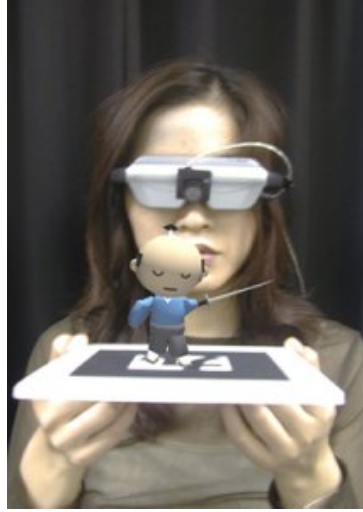


Figure 1.1 Fiducial marker used to represent a tridimensional model over it (??)

identified). Less common, there are approaches that replace fiducial markers (??) (??) by GPS, gyroscopes, accelerometers, cameras, among others (??) (??). These approaches have the advantage of not requiring human interference on the scene (to put a marker or to move it around).

Depending on the way that a user sees the mixed world, augmented reality can be classified on two ways. When the user sees the mixed world pointing his eyes straight to the real positions with optical scene or video, this augmented reality is called *immersive* or of *direct vision*. On the other hand, when the user sees the mixed world by some device, like a monitor screen or projector, not aligned with the real positions, this augmented reality is *non-immersive* or of *indirect vision* (??).

This work proposes a multi-view environment for augmented reality, of direct vision, composed by two Kinects (??) and augmented reality glasses, that allows a watcher visualize, in real time, virtual images merged with real images from the target model. In this approach, it's not intended to use any fiducial marker. Instead, it will be used a geometric approach based on the data captured by each Kinect. This proposed environment can be used, for example, on the medical field (real situations, education and training) or in other situation where a multi-view environment for markerless augmented reality is applicable.

1.2 OBJECTIVES

The environment proposed on this work aims to contribute to augmented reality applications where virtual images need to be merged with real images in real time, without using fiducial markers, and considering the viewing angle of the observer and the position of the target object.

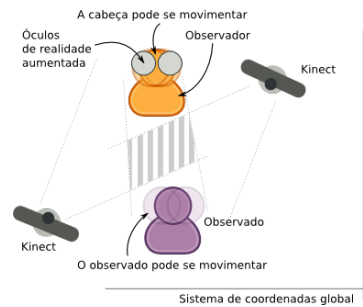


Figure 1.2 Global vision

1.2.1 Features

Based on the study of related works, it was defined the following features that define the scope of the first version of the augmented reality environment, graphically represented on Figure (??):

- There are two main elements on the environment, the observer and the target model;
- Observer and target model are positioned one in front of the other, with some restriction of minimum and maximum distance;
- The observer sees the combination of a real image with a pre-defined virtual image over the target model, through the augmented reality glasses;
- The observer and target model can move their heads, in a way that the virtual image adapts itself in real time to fit the new viewing angles;
- No fiducial marker is used.

The elements presented on this environment are the following:

- Observer - He is the user of the system that wears a pair of augmented reality glasses and is positioned in front of the target model (which can be a human being, for example, or a static object). On a medical context, the observer would be a medical specialist, responsible for observing the patient and responsible for analyzing the combination of the real image (part of the patient body) with a virtual image (from magnetic resonance imaging or computed tomography). The observer is able to move his head.
- Augmented reality glasses - The augmented reality glasses are wore by the observer and have two cameras. The images captured from the target model will be merged with virtual images and displayed on the two lenses of these glasses. On one of the approaches implemented by this work, the observer movimentation is determined from sensors present on the glasses: acceletometer, magnetometer and gyroscopes. Based on sensor data, it's possible to determine the variation on the orientation of the glasses, and so it's possible to know how much the observer has moved his

head. This calculation returns values that define movements on longitudinal axis, vertical axis and lateral axis. The virtual image should be reprojected in real time according to those movements.

- Target model - The target model (a human being, for example), is placed in front of the observer and doesn't use any kind of fiducial marker. The main goal is that the virtual image is placed over the target model. In order to calculate where and how the virtual image should be displayed, it's necessary to identify the position and orientation of the real object relative to the observer. This is done based on two sensors placed on the environment, where the first one captures data from the observer and the second one captures data from the target model.
- Sensors - Two sensors are placed on the environment and capture data from the observer and the target model (one for each). The information captured by the target model's sensor contains its model, that will be merged with the virtual image.

Each device presented on this multi-view (glasses and sensors) environment has its own coordinate system, but all information must be converted to a global coordinate system.

Since there are two sensors and one pair of augmented reality glasses, another objective is to implement two different approaches: one that uses the augmented reality glasses to determine the observer's pose (based on data from accelerometer and magnetometer) and another one that uses a second Kinect device to determine the observer's pose based on a reconstruction of his model.

1.3 THESIS OVERVIEW

The next chapter "Conceptual primer" describes some basic theory needed. It also describes the hardware that are used by this work and how to calibrate them. After that, the "Related work" chapter presents some works related to this one, divided by subject. The third chapter, "Solution architecture", after the theory and related works were presented, explains the steps performed in order to implement the objectives of this work. The results of this implementation are presented on the chapter later, "Results", and finally the conclusions about those results are presented on the last chapter, "Conclusions", where possible future works are also listed.

In this chapter I present the main concepts behind this work.

CONCEPTUAL PRIMER

2.1 AUGMENTED REALITY

2.1.1 Fiducial markers

2.1.2 Markerless

2.1.3 Direct or indirect vision

2.2 CAMERAS

2.2.1 Calibration

2.2.2 Parameters

2.2.3 Multi-view

2.3 SENSOR-BASED REGISTRATION APPROACH

2.3.1 Kinect

2.3.2 Registration

2.3.3 Transformation

2.4 VISION-BASED REGISTRATION APPROACH

2.4.1 Optical flow

2.4.2 Lucas-Kanade algorithm

Chapter

3

In this chapter I present some related works.

RELATED WORK

Cover all the related works, with multiple Kinects, optical flow, markerless augmented reality, medical applications, multi-view environment, reconstruction, etc.

3.1 MARKERLESS AUGMENTED REALITY

3.2 MULTIPLE KINECTS

3.3 MULTI-VIEW ENVIRONMENT

In this chapter I explain in details the steps performed in order to implement the objective of this work.

SOLUTION ARCHITECTURE

4.1 ENVIRONMENT

Technologies, machines, SOs, etc.

4.2 CALIBRATION

4.2.1 Augmented reality glasses calibration

4.2.2 Initial calibration between Kinects

4.2.3 Initial calibration between Kinect and glasses

4.3 COMMUNICATION

Network, sockets, etc.

4.4 TRANSFORMATIONS

4.5 METHOD 1: GLASSES ACCELEROMETER AND ONE KINECT

Cover Lucas-Kanade algorithm, etc.

4.6 METHOD 2: TWO KINECTS

Talk about performance of two Kinects fighting for a single GPU.

4.7 HYBRID APPROACH

When optical flow has just a few feature points, we switch to the second Kinect.

Chapter

5

In this chapter I present the results of the procedure explained in the previous chapter.

RESULT

5.1 SCOPE

Talk about error propagation.

5.2 ANALYSIS

Talk about performance and alignment results.

5.3 COMPARISON

Compare methods 1 and 2 with regards to performance and quality.

Chapter

6

In this chapter I discuss the conclusions of this work and list some possibilities of future works.

CONCLUSIONS

6.1 FUTURE WORK

