Bachelorarbeit

Automatisches Labeln von Objekten in einer Augmented Reality Umgebung

Janelle Pfeifer
Delpstraße 28
53359 Rheinbach
janelle.pfeifer@smail.inf.h-brs.de

Hochschule Bonn-Rhein-Sieg
Institute of Visual Computing
Fachbereich Informatik
Studiengang: Informatik (B.SC.)

Erstprüfer: Prof. Dr. Ernst Kruijff
Zweitprüfer: Prof. Dr. André Hinkenjann

Rheinbach, 1.10.2020

Inhaltsverzeichnis

Bachelorarbeit 1
1 Einleitung 3
2 Grundlagen 3
2.1 Räumliche Zuordnung in Unity 5
2.2 Räumliche Anker 5
2.3 Object Detection 5
2.4 Azure Object Detection 5
2.5 Magic Leap 5
2.6 Webrequests 5
2.7 Camera to World Matrix in Unity 5
2.8 CNN Networks für Object Detection 5
3 Design und Implementierung 5
4 Design 5
5 AzureObject Detection 5
5.1 Pixel to World 6
6 Implementierung 6
6.1 Azure Vision Services 7
6.2 Foto Pixel zu AR Welt 7
6.3 Object Detection mit Hololens Daten 7
6.4 Object in 3D Szene mit Labeln versehen 7
7 Zusammenfassung 7

1 Einleitung

Augmented Reality (AR) ist eine Vermischung der realen Welt mit digitalen Elementen. Es wird durch Anzeigegeräte, wie Handys, Tablets oder Augmented Reality Brillen präsentiert und bietet ein intuitives Benutzerinterface um Informationen über Objekten der realen Welt anzuzeigen. Dafür müssen Informationen über die Umgebung erfasst werden. Es ist wichtig zu bestimmen welche Objekte sich in der Umgebung befinden, die durch AR erweitert werden soll.

Es gibt mehrere Möglichkeiten reale Objekte zu erkennen. Zum einen können Markierungen in der realen Welt verwendet werden. Dabei handelt es sich um statische Bilder, beispielsweise ein Foto, oder ein QR Code, die von einer Kamera eingescannt werden. Der Marker ist einzigartig für jedes Object, das erkannt werden soll, damit sie voneinander unterschieden werden können. Der Nachteil bei diesem Vorgehen ist der Arbeitsaufwand, der damit verbunden ist, jeden Gegenstand einzeln zu Markieren und der AR Applikation die Marker bekannt zu machen.

Wenn man Markierungen in der Realen Welt umgehen möchte, kann man den Nutzer der Applikation bitten, per Geste oder per eye sight auf Objekte der Realen Welt zu weisen, die erkannt werden sollen. Dabei muss von dem Nutzer auch angegeben werden, um welches Object es sich genau handelt, damit die Applikation unterschiedliche Objekte auseinander halten kann und die korrekten Informationen mit ihren assoziiert. Auch hier ist ein hoher Arbeitsaufwand damit verbunden alle Objekte für die Application auszuweisen.

Beide der Verfahren sind nicht auf große AR Umgebungen skalierbar, da sie sehr Arbeitsintensiv sind. Nur eine voll automatische Objekterkennung ist skalierbar.

Um diese Automatisierung zu erreichen kann Image based Object Detection aus dem Bereich der Computer Vision verwendet werden. Dabei werden Objekte in Bildern erkannt, indem nach Charakteristiken gesucht werden, die unterschiedliche Arten an Objekten auszeichnen.[**introToCNN**]

In dieser Thesis wird das Erkennen und Labeln von Objekten in einer AR Umgebung, mithilfe von Image based Objekt Detection, automatisiert. Dabei wird as AR Gerät Magic Leap"verwendet.

2 Grundlagen

Spatial Mapping

Durch Spatial Mapping wird eine 3D Abbildung einer realen Umgebung erschaffen. So können Hologramme mit der echten Welt interagieren, diese Verdecken, oder von ihr verdeckt werden.[**spatialMapping**]

Object Detection

Bei Objekt Detection werden Objekte in einem Bild untersucht. Dabei wird bestimmt um welche Klasse an Objekt es sich handelt, beispielsweise ob es eine Katze oder ein Hund ist, und wo sich das Objekt befindet. Die Ausgabedaten dieser Untersuchung ist eine Liste an Objektarten und eine Liste an Bounding Boxen, die die Positionen angeben.

Artificial Neural Networks

Artificial Neural Networks sind Machine Learning Architekturen. Sie können beispielsweise Musik, Text oder Bilder nach Mustern durchsuchen. Sie sind für keine genaue Aufgabe programmiert, sondern lernen indem sie mit Beispieldaten trainiert werden. Für jedes Beispiel gibt es ein Label, das angibt ob es das gesuchte Muster enthält oder nicht. Die Struktur des Networks verfügt über Gewichte, die Einfluss auf den Output haben. Mit jedem Trainingsbeispiel passt das Network die Gewichte an, sodass der Output dem Label des Beispiels entspricht.[**introToCNN**, **surveyOfDeepLearing**]

Convolutional Neural Networks

Convolutional Neural Networks sind auf das Verarbeiten von Bildern spezialisiert. Sie nutzen aus, das Bilder viele Redundanzen und Informationsarme Bereiche haben, indem sie mit jedem Verarbeitungsschritt Informationen weglassen. So können Rechenzeit und Trainingsdaten verringert werden.[**introToCNN**, **surveyOfDeepLearing**, **cNNforClass**]

Azure maschinelles Sehen

Microsoft Azure bietet einen Computer Vision Service an, der für Object Detection trainiert ist. Der Anwender sendet ein Bild an Microsoft, dort wird es verarbeitet und ein Ergebnis zurückgegeben.[**getAzure**, **whatIsAzure**, **objDetectAzure**, **Azure302Doc**]

Azure Custom Vision

Azure bietet zusätzlich einen Computer Vision Service an, den der Nutzer Trainieren kann um bestimmte Objekte Klassifizieren zu können.[**Azure302bDoc**]

Magic Leap AR Brille

Die Hololens verfügt über 4 Umgebungskameras, eine tiefen Kamera und eine ausrichtbare Kamera. Die Umgebungskameras werden für Spatial Mapping genutzt. Anhand der Topographie kann die Hololens einfache Ebenen, wie die Wände und den Boden eines Raumes erkennen. Die ausrichtbare Kamera kann geschwenkt werden und nimmt Fotos auf. Die Bilder, die dabei entstehen, erhält in Unity eine 'cameraToWorldMatrix', die zum Zeitpunkt der Erfassung für jeden Pixel des Bildes eine Position im Koordinatensystem der AR Umgebung angibt. So kann das Koordinatensystem des Bildes in das Koordinatensystem der Umgebung transformiert werden.[**locatableCamera**]

2.1 Räumliche Zuordnung in Unity

2.2 Räumliche Anker

2.3 Object Detection

2.4 Azure Object Detection

2.5 Magic Leap

2.6 Webrequests

2.7 Camera to World Matrix in Unity

2.8 CNN Networks für Object Detection

3 Design und Implementierung

Das Ziel ist es das Erkennen und Labeln von Objekten in einer AR Umgebung, durch Image based Objekt Detection zu ermöglichen und mit einer Magic Leap Brille umzusetzten.

Wenn der Nutzer den Controller betätigt, beginnt die Detection indem zunächst mit der Kamera der Magic Leap ein Foto aufgenommen wird. Dieses Foto wird dann an Azure Object Detection und Azure Custom Vision geschickt. Als Ergebnis gibt es jeweils eine json Datei. Darin wird angegeben welche Objecte auf den Bilder gefunden wurden und in welchem Diese Werden Verarbeitet um

4 Design

Die Magic leap macht fotos von der Umgebung. die Bilder werden dann mit Azure Analysiert. Das Ergebnis ist eine Json Datei. Es wird für Bildbereicht angegeben welcher Object dort gefunden wurde. Die Lokation ist in Pixeln angegeben und hängt natürlich von dem Foto ab. Die Pixel Position muss umgewandelt werden in eine Position im Raum an der sich das Object befindet.

Die Pixel Position von dem Foto wird in eine Position auf der Clipping Plane der Camera umgerechnet. Ein Raycast wird von dem Urspurng der Camera durc hie Position in die welt geschickt. Der Raycast trifft dann auf das Mesh der Umgebugn dir von dem Magic leap spatial mapping erzeugt wurde. Der Treffpunkt des Raycast ist die Position des Objektes und wird mit einem Label Markiert.

Wenn der Nutzer den Trigger von dem Controller drückt, wird TackePicture getriggered. Ein Thread geöffnet um das Bild zu machen, das dauert nähmlich ein bischen. Zu dem Zeitpunkt wird die aktuelle Position der Camera gespeichert.

Das Foto wird durch die ML Camera aufgenommen. In der Methode OnCaputreRawImageComplete wird aufgerufen, wenn ein Bild aufgebonnen wurde. Dort wird die Analyse der Bilder gestartet.

5 AzureObject Detection

Um Azure Object Detectio zu nutzen muss ein ein Webrequest gemacht werden. Ein Post request an einen web endpoint von Azure. In den header kommt ein authorization Key und das bild in den content.

Azure führt die Analyse durch und gibt eine Json Datei als Response zurück. Beispiel der Json::

Es wird für jedes gefundene Object ein Klassenname und ein Viereck angegeben in des sich das Object auf dem Bild befindet. Die Mitte des Vierecks dient hier als Ankerpunkt für das Object. Die Klasse PixelToWorld bekommt die x und y koordinaten des Mittelpuntke sowie den Objectnamen und die KameraPosition übergeben die Die Kamera hatte, als das Foto aufgenommen wurde.

5.1 Pixel to World

In diesem Abschnitt soll die Pixel Position von einem 2D Foto in eine 3D Position auf auf dem Mesh des Spatial Mappings übertragen werden.

Als erstes wird die Pixel Position auf eine Position auf der Clipping Ebene der Camera umgerechnet. Dafür wird die CameraToWorld Matrix eingesetzt. Um xyz koordinaten zu erhalten die sich auf der clipping ebene befinden. Mit Vector offset von der Camera bestimmen. Mutliplizieren mit cameraToWorld matrix um Koordinate in der welt abhänig von der blickrichtug, dem winkel und der position der kamera zu erhalten. Vekotr mit dem Multimpiuiert wird bekommt 3 dinge: u, v, distance.

Distance ist die distanz vom urspung der kamera. Hier ist die distance zur clipping plane genutzt. U und V geben dann die Kooridnate auf der Bildebene der Kamera an.

Die x und y Kooridnate von dem foto werden in u und v umgewaldelt. Es wurden minimal und maximalwerte für u und v ausprobiert, mit dennen die Ränder des FOTOS AUF DER bILDENENE lokalisiert werden können. Dabei wurde beachtet, das die fotos größer ald die Bildebene sind und ein anderes Seitenverhältnisse haben. Siehe abbildung Bild, wo foto eingeblendet ist, man das view frustum sieht und die ränder markeirt sind mit u und v. Mit den minimal und maximalwerten, sowie der Größe der Fotos werden zwei linearfunktikoionen aufgestellt mit dennen x und y in un und v umgwaldelt werden können. Siehe Abbidlung, die die Funmktioien zeigen. Raycast. Ein Raycast durch den Urpsung der Kamera und den Punkt auf der Clipping Plane. Nutzt ML Raycast Setzte Urpsung und Direction. Der kann Raycast auf Spatial Mesh Siehe Abbildung. Erstelle ein Label an der Stelle. String des Labels wird in raycast start mit lambda gesetztt. Dann muss mlraycast den Inhalt nicht füllen, wenn der die rückgabe methode aufruft. This is sooo vage. Über u und v kann dann der pixel auf der clippingplane bestimmt werden. Foto hat ein anderes seitenverhältnis und ist größer als der view frustum der kamera.

Die Um einen Punkte auf der Clipping Ebene zu erhalten wird ein Abstand von 0.4 angegeben, da die Clipping Plane 0.37 von muss Die Matrix bekommt einen Vector Mit einem Abstand von 0.4 von dem Camera kann eine Position auf der Clipping Plane angegeben werden.

Erste Problem ist, das das Foto ein Anderes Seitenverhältnis und eine andere größe hat als das Display der Magicleap und somit als die Unity Camera.

6 Implementierung

Input von dem Magic Leap controller abwarter. Man muss permission haben um Fotos zu machen ins Internet zu gehen.

6.1 Azure Vision Services

6.2 Foto Pixel zu AR Welt

6.3 Object Detection mit Hololens Daten

6.4 Object in 3D Szene mit Labeln versehen

7 Zusammenfassung