

Alma Mater Studiorum · Università di Bologna

Scuola di Scienze

Corso di Laurea in Informatica per il management

**Realizzazione di un'applicazione in Mixed Reality di
“LandMark Recognition” con Microsoft HoloLens**

Relatore:

Chiar.mo Prof.

Lorenzo Donatiello

Presentata da:

Carlo Fanciulli

I Sessione

Anno Accademico 2016 / 2017

Indice Contenuti

Indice Contenuti	3
Indice Immagini	5
Introduzione	7
1. Hololens	8
Cos'è Hololens	8
Tipi di Realtà	9
Mixed Reality	11
2. Input-Output Hololens	12
Input	12
Hardware	14
Output	16
Prestazioni	17
Limitazioni	20
Altri Device (Competitors)	23
3. Sviluppo con Hololens	25
Unity	25
Applicazioni Olografiche per Hololens	26
Primi Sviluppi	28
Usi	28
4. Progetto: LandMark Recognition	32
Introduzione	32
API per il riconoscimento di immagini	32
API: Google Cloud Vision	34
API: Wikipedia	38
Descrizione dell'applicazione	39
User Experience	49
Conclusioni	51
Bibliografia	52

Indice Immagini

Fig. 1.1 Hololens.....	7
Fig. 1.2 Reality–Virtuality Continuum.....	8
Fig. 1.3 Augmented/Virtual Reality Revenue Share 2020F.....	9
Fig. 2.1 Sistema Input/Output.....	11
Fig. 2.2 Input principali di Hololens.....	12
Fig. 2.3 Componenti Hardware Hololens.....	13
Fig. 2.4 Specifiche complete hardware di Hololens.....	14
Fig. 2.5 Funzionamento lenti di Hololens.....	15
Fig. 2.6 Sistema di Performance nel portale web di Hololens.....	17
Fig. 2.7 FOV Hololens.....	19
Fig. 2.8 Distanza ottimale per l’utilizzo.....	20
Fig. 2.9 Meta 2 Gesture.....	22
Fig. 2.10 Confronto dispositivi di Mixed Reality.....	23
Fig. 3.1 Proprietà della Main Camera in Unity.....	25
Fig. 3.2 Ologramma di una macchina.....	27
Fig. 3.3 Ologramma della tavola periodica degli elementi.....	28
Fig. 3.4 Utilizzo Hololens in sala operatoria.....	29
Fig. 3.5 Utilizzo Hololens nel settore del design.....	30
Fig. 3.6 Casco con Hololens integrato per utilizzo militare.....	30
Fig. 4.1 Utilizzo delle API Microsoft.....	32
Fig. 4.2 Utilizzo delle API Google.....	32
Fig. 4.3 Logo di Google Cloud Platform.....	33
Fig. 4.4 Sistema per il riconoscimento di luoghi storici.....	34
Fig. 4.5 Logo Wikipedia.....	37
Fig. 4.6: Diagramma di sequenza dell’applicazione.....	40
Fig. 4.7 Interfaccia del progetto Unity.....	40
Fig. 4.8 Gerarchia del progetto Unity.....	42
Fig. 4.9 Interfaccia primaria per lo scatto della foto.....	48
Fig. 4.10 Visualizzazione in primo piano del nome del monumento identificato.....	48
Fig. 4.11 Pannello contenente tutte le informazioni prese da Wikipedia.....	49

Introduzione

“When you change the way you see the world, you can change the world you see”.

Con questo slogan Microsoft ha dato inizio a una nuova era tecnologica che potrebbe cambiare interamente la percezione della realtà e dell’interazione con le tecnologie che abbiamo avuto fino ad adesso. Letteralmente.

Annunciato durante la conferenza di lancio di W10 nel 2015 dopo 5 anni di lavorazione top secret¹, HoloLens è il primo prototipo di visore a realtà mista che funziona in totale autonomia. Questa prima versione sta già riscontrando una notevole popolarità a causa della rivoluzionaria tecnologia proposta e della portabilità del dispositivo, benché sia ancora disponibile solo per gli sviluppatori.

In questo lavoro di tesi viene presentata un’applicazione per Hololens per il riconoscimento di edifici storici e conseguente visualizzazione di informazioni che lo caratterizzano.

Prima di descrivere l’implementazione realizzata, si introducono i concetti fondamentali relativi alla Mixed Reality e ai suoi campi applicativi, si discutono i limiti e le potenzialità della tecnologia Hololens confrontandola anche con altre proposte disponibili sul mercato.

L’applicazione realizzata viene discussa nel capitolo 4.

Ovviamente, trattandosi di un prototipo molto recente, rilasciato a Marzo 2016, la fase di sviluppo ha richiesto un’approfondita analisi delle caratteristiche operative di questa tecnologia.

¹ <https://www.wired.com/2015/01/microsoft-hands-on/>

1. Hololens

1.1. Cos'è Hololens²

E' difficile descrivere cos'è Hololens in quanto è un dispositivo che ha portato un'innovazione incredibile in un momento inaspettato e che risulta impossibile da comparare.



Microsoft descrive questa sua creazione come “computer olografico totalmente autonomo”.

Questo significa che è possibile, attraverso di esso e in particolare alle lenti, vedere oggetti virtuali tridimensionali in tempo reale da ogni angolazione e distanza senza bisogno di essere collegato, con cavi o wireless, ad un computer.

Hololens, infatti, ha tutti i componenti integrati che permettono un'elaborazione pienamente funzionale.

Il sistema operativo utilizzato è Windows 10 grazie alla quale è possibile sviluppare applicazioni universali che girano su qualunque dispositivo Microsoft.

In più è stato implementato l'assistente vocale “Cortana³” che risulta essere molto efficiente e pratico per Hololens.

Le lenti utilizzano sensori avanzati, un display ottico 3D ad alta definizione ed un sistema di scansione spaziale che consentono a chi li indossa di poter fruire di applicativi di realtà aumentata attraverso un'inedita interfaccia olografica con la quale poter interagire mediante lo sguardo (Gaze), i gesti delle mani (Gestures) o la voce (Voice).

Hololens viene spesso considerato erroneamente come un dispositivo di realtà aumentata o virtuale. Per questo motivo vediamo nel capitolo successivo quali sono le distinzioni principali tra le varie realtà, nate e cresciute negli ultimi anni, per capire appieno le potenzialità della “Mixed Reality” creata da Microsoft.

² Fig. 1.1: Hololens

³ <https://support.microsoft.com/en-us/help/12630/hololens-cortana-on-hololens>

1.2. Tipi di Realtà

Per cogliere le caratteristiche e le potenzialità di questo visore di Microsoft è importante rendere chiaro in quale modi è stata arricchita la percezione sensoriale umana negli anni passati.⁴

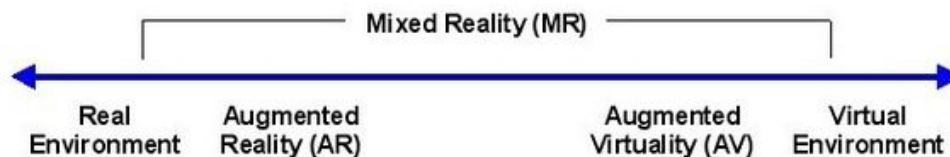


Fig. 1.2: Reality–Virtuality Continuum⁵

La vista è parte fondamentale della nostra vita. Ogni giorno arricchiamo la nostra vita e conoscenza grazie a cosa vediamo e nel modo in cui ci interagiamo.

Ricaviamo ben l'80% delle nostre impressioni grazie alla vista.

Riuscire a portare la fantasia e l'immateriale nella nostra realtà è stata parte dell'immaginazione umana per migliaia di anni.

Stiamo superando questo ostacolo e realizzando i nostri sogni proprio grazie alla tecnologia, all'attuale capacità computazionale che abbiamo raggiunto e all'avvento dei dispositivi mobile.

Come possiamo notare dall'immagine soprastante diverse realtà hanno preso piede soprattutto in questi ultimi anni.

La più influente, specialmente in ambito mobile, è la **Augmented Reality (AR)** in cui vengono addizionati oggetti virtuali alla realtà comune.

In questo tipo di realtà è comunque predominante la percezione del mondo reale in quanto gli oggetti virtuali servono solamente come informazioni aggiuntive.

I Google Glass sono un chiaro esempio di dispositivo per l'AR ma attualmente viene particolarmente implementato per i videogiochi.

⁴ <https://www.foundry.com/industries/virtual-reality/vr-mr-ar-confused>

⁵ https://en.wikipedia.org/wiki/Reality–virtuality_continuum

Diversamente l'Augmented Virtuality (AV) consiste nell'aggiunta di informazioni della realtà conosciuta in un ambiente virtuale.

Solitamente richiede una discreta potenza di calcolo come ad esempio per processare degli input sensoriali del mondo reale da collegare nel mondo virtuale.

Un caso è l'uso del giroscopio per giocare con lo smartphone.

La Virtual reality (VR), che sta riscontrando successo particolarmente in ambito videoludico, è tutto ciò che non è reale. Consiste nella creazione di un mondo interamente virtuale che permette di vivere esperienze che non hanno forma fisica ma che possono simulare una situazione realistica (come la fisica).

Dispositivi famosi sono l'Oculus Rift e il Samsung Gear VR.

La Mixed Reality (MR) rappresenta un ibrido che unisce le proprietà della realtà aumentata, quella più vicina a quella vissuta e che permette di aumentare la nostra visione reale mettendo in sovrapposizione oggetti virtuali e di quella virtuale, ovvero in cui il soggetto è immerso completamente in un ambiente digitale.

Questo tipo di realtà è sicuramente quella più promettente e, come vedremo, con tante promesse.

I grafici sottostanti dimostrano l'interesse economico di questo tipo di realtà.

La mixed reality viene comunque racchiusa tipicamente nella Augmented Reality.

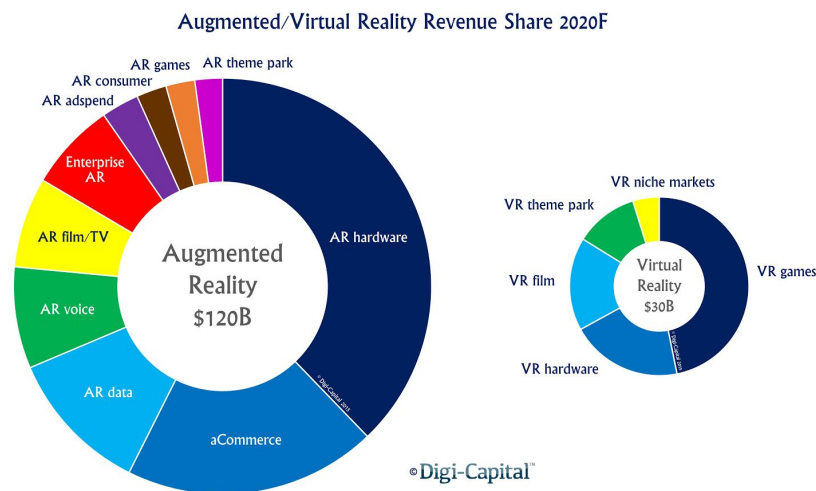


Fig. 1.3: Augmented/Virtual Reality Revenue Share 2020F

1.3. Mixed Reality

Come introdotto nel capitolo precedente la Mixed Reality cerca di combinare il meglio sia della AR sia della VR.

Questa realtà permette di vivere l'esperienza estremamente coinvolgente della realtà virtuale e insieme di mantenere un'interazione tra realtà esperita e oggetti virtuali propria di quella aumentata, con lo scopo di amplificare e migliorare l'esperienza complessiva dell'utente.

Con la Mixed Reality, in particolare utilizzata con Hololens, è possibile un riconoscimento dell'ambiente fisico circostante così da creare una mappa 3d della zona in modo da permettere al device di sapere esattamente dove e come posizionare gli oggetti virtuali, realisticamente, nello spazio mentre sarà possibile interagire con essi attraverso dei movimenti della mano chiamate Gestures.

Vedremo nel capitolo successivo come sarà possibile interagire con gli ologrammi attraverso lo sguardo (Gaze), i gesti (Gestures) e la voce (Voice)⁶.

Questo modo di far interagire realtà e oggetti virtuali permette ad essi di comportarsi a tutti gli effetti come oggetti reali che sanno interagire con l'ambiente e tra di loro.

La Mixed Reality, benché sia presente sul mercato relativamente da poco tempo, ha l'incredibile potenziale di cambiare l'approccio che abbiamo verso il nostro mondo, come lavoriamo, interagiamo e ci comportiamo.

Più la tecnologia avanza più avremo dei dispositivi piccoli e portatili che ci permetteranno di adottarli con maggiore frequenza e quindi maggior impatto.

⁶ <https://developer.microsoft.com/en-us/windows/mixed-reality/gestures>

2. Input-Output HoloLens

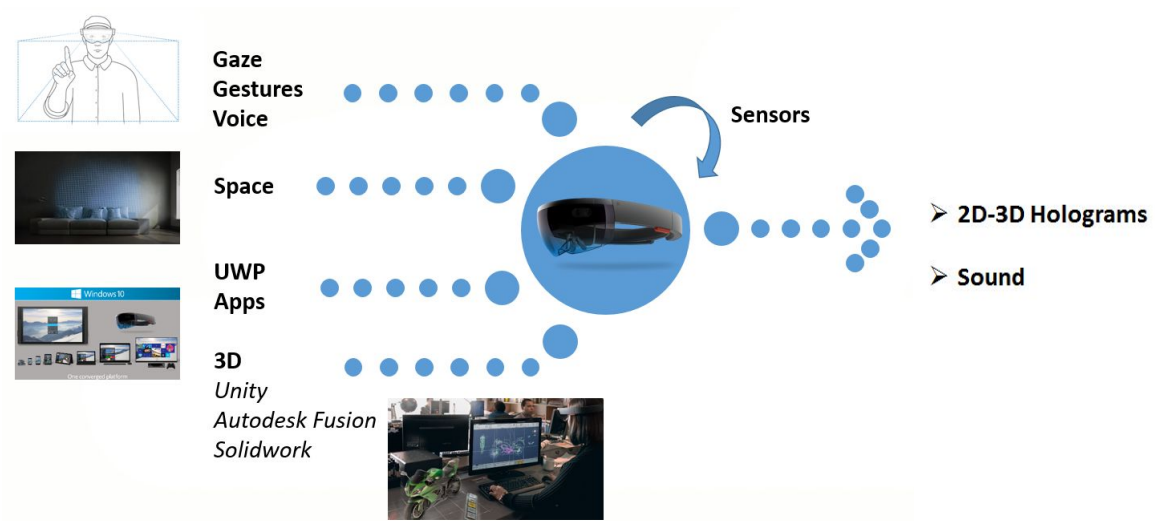


Fig. 2.1: Sistema Input/Output

2.1. Input

Gli input supportati da questo dispositivo sono i seguenti:

- **Gaze:** E' necessario per capire dove l'utente sta guardando così da determinare la sua intenzione.
Non vengono usati gli occhi ma la posizione e l'orientamento della testa dell'utente per poter poi interagire con la realtà e gli oggetti olografici.
- **Gestures:** Permettono l'interazione e la manipolazione degli ologrammi.
Sono pochi i gesti utilizzabili per interagire con i vari oggetti virtuali ma soddisfano gran parte delle necessità richieste per l'utilizzo ottimale delle applicazioni. Come vedremo, altri competitors, che utilizzano la stessa realtà, hanno un approccio differente con gli oggetti.
- **Voice:** Permette di avere un'interazione più immediata e diretta rispetto alle Gestures.

E' possibile utilizzare delle keyword da indicare in input (possono essere anche frasi) che vengono riconosciute dal dispositivo, attraverso l'integrazione nel

sistema di quattro microfoni, e con cui è possibile associare un'azione.

Compensa in modo soddisfacente la scarsità delle Gestures in quanto stravolge i metodi usuali per influenzare degli oggetti virtuali.

- **Spazio circostante:** un input alquanto inusuale ma di importanza rilevante.

Lo spazio circostante viene mappato in tempo reale tramite i vari sensori e l'utilizzo della HPU per garantire una migliore interazione tra gli ologrammi e la realtà.

- **Qualunque applicazione supportata dalla UWP** (Universal Windows Platform) può girare sotto forma di ologramma in Hololens.⁷

In particolare è possibile lanciare le applicazioni disponibili nel Windows Store proiettandole in superfici 2D su Hololens.

Un esempio è Skype che permette, con un uso anche migliorato per il device, un modo immediato e intuitivo per comunicare con altre persone in tutto il mondo.

- **Modelli 3D** creati attraverso Unity o importati da software per la progettazione cad 3D come SolidWorks o AutoDesk.

Questo input è fondamentale per creare applicazioni indipendenti che utilizzano oggetti virtuali scelti dallo sviluppatore.

Ovviamente, come da prassi nel mondo videoludico, fa da struttura portante l'utilizzo di modelli 3d semplificati e soprattutto ottimizzati per il tipo di dispositivo che poi lo monterà.



Fig. 2.2: Input principali di Hololens

⁷ <https://abhijitjana.net/2016/04/08/bringing-your-universal-windows-apps-to-microsoft-hololens/>

2.2. Hardware⁸

La grande innovazione riguarda anche il sofisticato lato hardware di questo dispositivo, che possiede tutte le componenti⁹ per funzionare in modo indipendente, ovvero un computer a tutti gli effetti: la capacità di funzionare senza cavi è un tratto distintivo rispetto ad altri prototipi di competitors che apre a possibilità di utilizzo molto più ampie.



A livello di componenti, presenta un IMU (Inertial Measurement Unit) che include un giroscopio, un accelerometro e un magnetometro, 2 GB di RAM, una telecamera di profondità, 4 sensori per mappare l'ambiente circostante, una fotocamera da 2.4 megapixel, 4 microfoni e un sensore di luce ambientale.

In aggiunta alla CPU e alla GPU, HoloLens utilizza una HPU (Holographic Processing Unit¹⁰), un co-processore creato da Microsoft con specifiche molto interessanti che serve a processare in tempo reale miliardi di calcoli al secondo, nello specifico le informazioni rilevate dalle videocamere e dai sensori presenti per gestire la mappatura dello spazio, il riconoscimento dei gesti e della voce.

La durata della batteria è abbastanza limitata, circa 2.5 ore in utilizzo elevato. Soprattutto in ambito di sviluppo rappresenta un ostacolo perché si ha poca autonomia, ma probabilmente questo aspetto verrà superato in futuro.

A ricordarci che questo dispositivo è attualmente un prototipo c'è inoltre la non ottimale risoluzione supportata dalle lenti (1268×720 per occhio) e un campo di visione ridotto (30 x 17.5 gradi) che non permette di vedere molto di ciò che si trova lateralmente nell'ambiente virtuale.

⁸ Fig. 2.3: Componenti Hardware HoloLens

⁹ https://developer.microsoft.com/en-us/windows/mixed-reality/hololens_hardware_details

¹⁰ <https://m.windowscentral.com/here-are-specs-custom-hpu-chip-inside-microsoft-hololens-headset>

Il Field Of View (FOV) limitato¹¹ amplifica il problema “distanza ottimale” tra il soggetto e l’oggetto, che se troppo grande o troppo vicino viene visualizzato parzialmente. L’effetto di straniamento si supera con l’abitudine, ma attendiamo di vedere come interverranno nella prossima versione per migliorare il FOV (campo visivo).

Microsoft ha comunque già risposto alle lamentele rimarcando che il progetto ha dato importanza alla “densità olografica¹²”, cioè la particolare caratteristica di luminosità e definizione dell’ologramma che per l’azienda rappresenta il vero punto focale.

Infatti la “risoluzione olografica” risulta composta da più di 2 milioni di “punti luce” con una “densità olografica” di 2500 radianti o “punti luce per radiante”.

Il risultato è una qualità senza precedenti.

Le specifiche complete dell’hardware di Hololens sono le seguenti:

Optics	Sensors	Processors	Power	Input / Output / Connectivity
<ul style="list-style-type: none"> • See-through holographic lenses (waveguides) • 2 HD 16:9 light engines • Holographic Resolution: 2.3M total light points • Holographic Density: >2.5k radiants (light points per radian) 	<ul style="list-style-type: none"> • 1 IMU • 4 environment understanding cameras • 1 depth camera • 1 2MP photo / HD video camera • Mixed reality capture • 4 microphones • 1 ambient light sensor 	<ul style="list-style-type: none"> • Intel 32 bit architecture • Custom-built Microsoft Holographic Processing Unit (HPU 1.0) 	<ul style="list-style-type: none"> • Battery Life <ul style="list-style-type: none"> - 2-3 hours of active use - Up to 2 weeks of standby time • Passively cooled (no fans) 	<ul style="list-style-type: none"> • Built-in speakers • Audio 3.5mm jack • Volume up/down • Brightness up/down • Power button • Battery status LEDs • Wi-Fi 802.11ac • Micro USB 2.0 • Bluetooth 4.1 LE

Fig. 2.4: Specifiche complete hardware di Hololens

¹¹ <http://newatlas.com/hololens-fov-field-of-view-illustrated/44903/>

¹² https://www.ted.com/talks/alex_kipman_the_dawn_of_the_age_of_holograms/transcript?language=en

2.3. Output

L'output fondamentale dell'HoloLens è l'ologramma.

Un ologramma è un oggetto come qualunque altro nel mondo reale; l'unica sostanziale differenza è l'essere interamente fatto di luce.

Gli oggetti olografici possono essere visti da qualsiasi angolazione e distanza, come qualsiasi oggetto fisico e ovviamente abbiamo la possibilità di interagire con essi.

Possono produrre suoni senza coprire quelli provenienti dall'ambiente grazie ai due speaker localizzati direttamente sopra le orecchie.

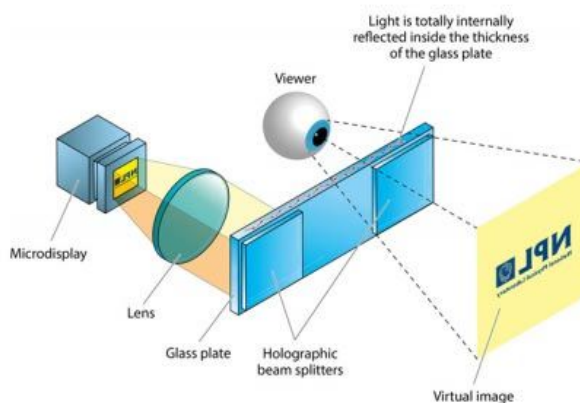
Gli ologrammi di HoloLens sono definiti così perché l'effetto è assolutamente assimilabile a immagini olografiche, ma l'uso del termine non è propriamente e tecnicamente corretto.

Con "ologramma" solitamente intendiamo la riproduzione dell'immagine di un oggetto reale creata tramite l'olografia. Un fascio di luce laser sparato sia verso l'oggetto sia verso una lastra grazie a un gioco di specchi: su questa lastra vengono a convergere sia la luce che arriva dal laser che quella riflessa dall'oggetto, venendosi a creare delle increspature contenenti l'informazione sulla tridimensionalità dell'oggetto.

Puntando un altro raggio laser sulla lastra viene poi a decodificarsi l'informazione, ricostruendo l'oggetto tramite appunto l'ologramma.

Questo è generalmente creato nello spazio da un gioco di riflessi senza che

la persona debba indossare visori, mentre in HoloLens è stata sviluppata una tecnologia diversa¹³: le particelle di luce rimbalzano milioni di volte nella struttura di "light engine"¹⁴ implementata frontalmente per creare le immagini tridimensionali. Questo sistema è



¹³ <http://www.imaginativeuniversal.com/blog/2015/10/18/how-hololens-displays-work/>

¹⁴ https://en.wikipedia.org/wiki/Microsoft_HoloLens

poi inserito all'interno di una tecnologia che permette di mettere in comunicazione le immagini con gli altri sensori e componenti del device e conseguentemente con i gesti della persona, ottenendo l'inedito effetto di interagire con le immagini.¹⁵

2.4. Prestazioni

Hololens è capace di dare un'esperienza olografica molto immersiva senza l'uso di cavi collegati ad un PC (come vedremo avanti in altri modelli).

Per riuscire a processare miliardi di calcoli al secondo, dovuti principalmente alla sensoristica per lo spatial mapping e la gesture recognition, Microsoft ha affidato il compito al suo processore di ultima generazione chiamato Holographic Processing Unit¹⁶.

Le applicazioni create con Hololens hanno bisogno di bilanciare la complessità visiva, la frequenza dei frame, la latenza di rendering, la latenza di input, il consumo di energia e la durata della batteria per ottenere un'esperienza ottimale.

Per la comodità e l'usabilità degli utenti, è importante ottenere livelli di frame costanti ed elevati con una latenza minima.

Le applicazioni devono raggiungere 60 fotogrammi al secondo (fps) con una latenza di 1 frame.

Una volta che l'applicazione raggiunge la performance desiderata, è importante monitorare il consumo energetico e assicurare che l'applicazione funzioni correttamente senza sbalzi di energia e di frame.¹⁷

Sia la frequenza di fotogrammi sia il consumo di energia possono essere osservati nel portale web del dispositivo collegato alla rete wifi raggiungibile tramite indirizzo ip.

¹⁵ Fig. 2.5: Funzionamento lenti di Hololens

¹⁶<https://www.theverge.com/circuitbreaker/2016/8/23/12602516/microsoft-hololens-holographic-processing-unit-specifications>

¹⁷https://developer.microsoft.com/en-us/windows/mixed-reality/performance_recommendations_for_hololens_apps

Questo portale offre diversi servizi tra cui il monitoraggio di sistema del dispositivo.

Il portale del device offre i seguenti strumenti per la performance:

- *Tracciamento Performance* - questa pagine offre la possibilità di catturare dati di performance e analizzarli con Windows Performance Analyzer.
- *Processi* - mostra i processi attualmente attivi con le relative informazioni.
- *Performance di sistema* - per il tracciamento in tempo reale durante l'utilizzo del dispositivo della potenza, utilizzo cpu e gpu, disco I/O, memoria, frame rate e altre metriche.

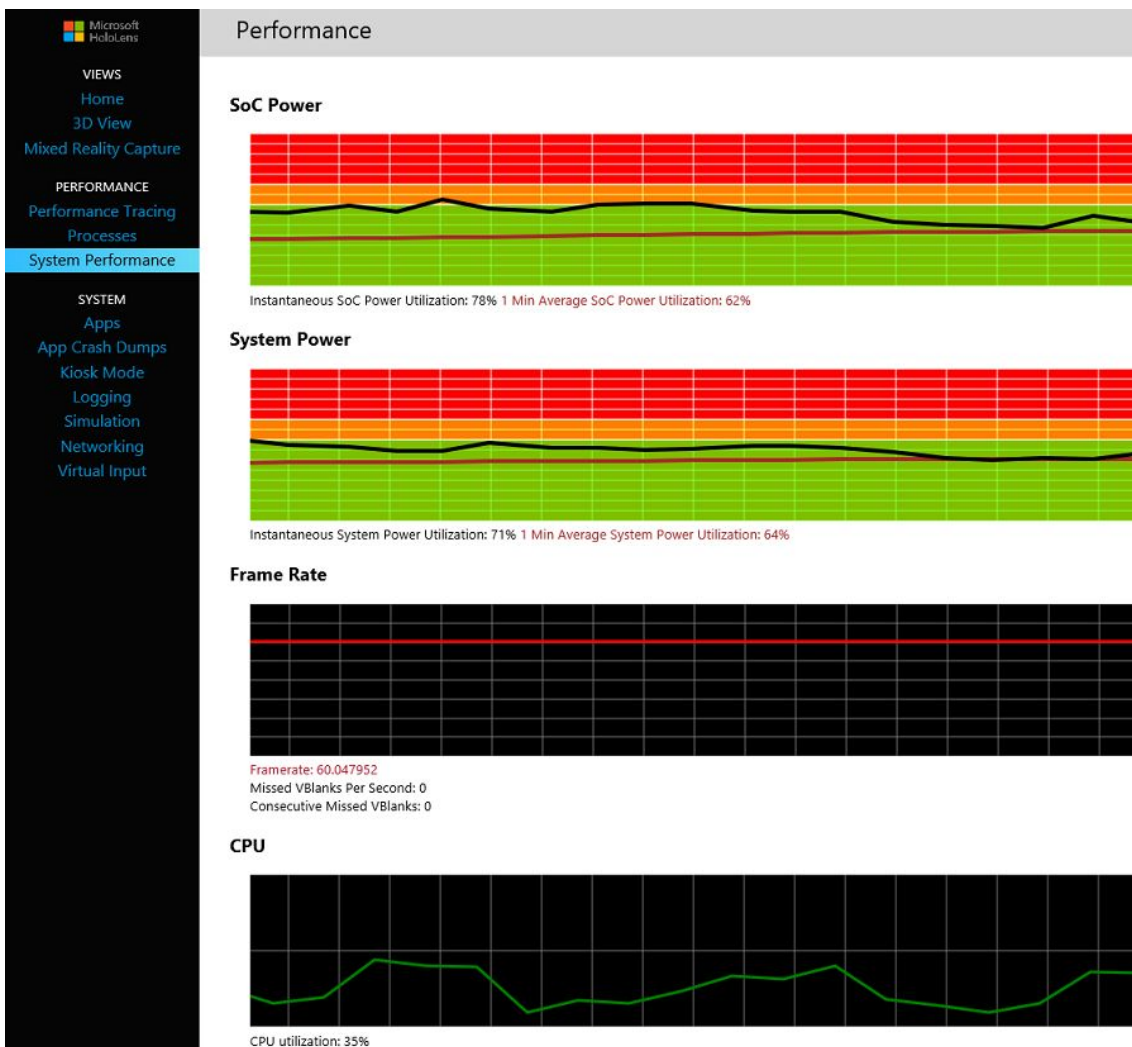


Fig. 2.6: Sistema di Performance nel portale web di Hololens

In particolare nella pagine di performance del sistema è possibile visualizzare le seguenti informazioni:

- Potenza SoC: è il consumo di energia combinata tra CPU, GPU e memoria.
- Potenza Sistema: mostra il consumo totale di energia di Hololens. Questo include sensori attivi, display, speakers, GPU, CPU, memoria e tutti gli altri componenti.
- Frame Rate: Sicuramente un indicatore importante per gli sviluppatori.
- CPU: Percentuale di tempo in cui i core della CPU stanno funzionando. Rappresenta quindi l'utilizzo completo di tutti i core del processore.
- GPU: Percentuale di tempo in cui la GPU è attiva.
- I/O: Lettura e Scrittura del disco per tutti i processi
- Memoria: Mostra la memoria occupata per ogni processo nel device. Il limite di memoria di un'applicazione per Hololens è di 900 MB. Se l'applicazione eccede questo limite verrà terminata.
- Network: Rete utilizzata per ogni processo in corso.

2.5. Limitazioni

Prima di iniziare ad elencare le limitazioni di questo dispositivo è necessario ribadire che si tratta di un prototipo uscito solamente da un anno, ancora acerbo sotto molti punti di vista.

Sicuramente verranno superati con il passare degli anni visto che molte di queste si riferiscono soprattutto, come vedremo, alla parte hardware del dispositivo.

- La prima su tutte, il **Field of View** (FOV) che limita in modo decisivo l'esperienza dell'utente.

Limitato e ristretto per una questione di batteria, Hololens si concentra più nell'essere un dispositivo autonomo che vuole avere come punto di forza la portabilità e l'essere senza cavi così da poterlo utilizzare in qualunque contesto e luogo anche se è ottimizzato per un uso interno.

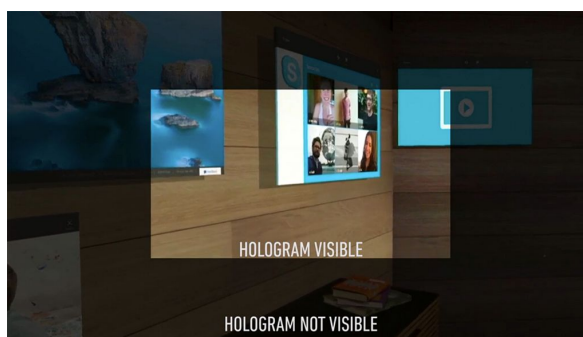


Fig. 2.7: FOV Hololens

- Come accennato prima, la batteria è un problema con cui bisogna avere a che fare soprattutto se pensato per un uso commerciale.

Considerando la potenza computazionale necessaria e la costante mappatura dell'ambiente circostante è comprensibile immaginare il dispendio energetico elevato in caso di uso frequente.

Il problema è bypassabile usando il dispositivo sotto carica con il cavo USB attaccato al computer che ne limita l'usabilità ma ne permette l'impiego.

- Altra carenza hardware è sicuramente la risoluzione.

La massima risoluzione supportata da Hololens è infatti di 1268x720 per occhio (720p) mentre la minima è 634x360 (360p)¹⁸. Microsoft insiste comunque nel confermare i 60 fps quando viene utilizzato Hololens per sviluppare anche se calerà a 30 in caso di registrazione video.

Questo limite risulta abbastanza tangibile considerando sia i tempi attuali in cui si sta raggiungendo risoluzioni significative sia confrontandolo con diversi dispositivi di realtà virtuale che permettono di avere, come nel caso di Oculus, una risoluzione per occhio di 1080x1200 o come nel caso di Meta 2, una risoluzione di 2560x1440.

- Un'altra restrizione da considerare riguarda la percezione degli ologrammi su diverse distanze. Infatti quando ci si avvicina troppo ad un ologramma questo sparisce perchè a distanza ravvicinata può comportare malessere.

Per questo è stata settata una distanza ottimale come riferimento in cui è preferibile vedere l'oggetto virtuale relativo.

Microsoft ci conferma appunto che la visualizzazione ottimale, come mostra l'immagine sottostante è a due metri dall'utente.

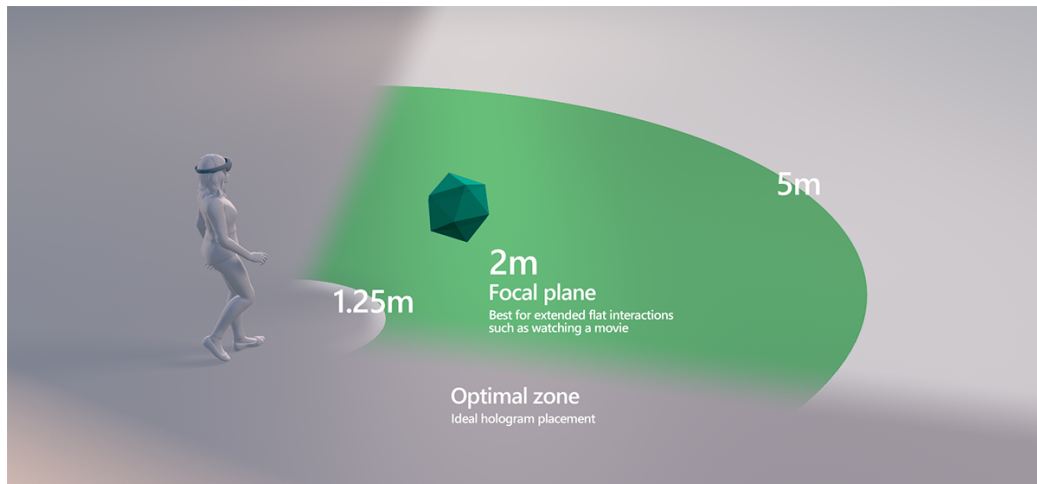


Fig. 2.8: Distanza ottimale per l'utilizzo¹⁹

¹⁸<http://www.pcworld.com/article/3039822/consumer-electronics/we-found-7-critical-hololens-details-that-microsoft-hid-inside-its-developer-docs.html>

¹⁹ https://developer.microsoft.com/en-us/windows/mixed-reality/hologram_stability

- Restrizione importante è la bassa quantità di poligoni renderizzabili dal dispositivo.

Come riscontrato da me tramite vari import di modelli nelle scene e da vari utenti nel web, il limite raggiungibile in termine di poligoni è intorno ai 100mila per avere un'esperienza fluida senza far calare gli FPS²⁰.

Considerando che in una scena si possono presentare una moltitudine di oggetti allo stesso tempo si capisce subito che deve essere considerata primaria la scelta dei modelli in termini di qualità/quantità poligoni per avere la resa desiderata.

Questo significa che dev'essere presente o un'ottimizzazione del modello per quel dispositivo dedicato, risolvibile anche tramite dei software dedicati, o direttamente la creazione dell'oggetto con la quantità giusta di poligoni.

Sono ovviamente presenti dei consigli tecnici che cercano di assottigliare il problema cambiando impostazioni del progetto o creando limitazioni come ad esempio il non utilizzo delle ombre che si è dimostrato molto dispendioso in termini di calcolabilità e che può avere un grosso impatto a livello di FPS.

- La luce ha un grosso impatto nell'esperienza di utilizzo.

Risulta avere un ruolo primario in quanto influenza il rendering dell'ologramma. Come spiegato nel capitolo 2.3, essendo un ologramma fatto di luce riflessa nelle lenti se questa risulta essere troppo intensa o troppo soffusa il risultato sarà molto insoddisfacente. Questo è subito visibile con l'uso esterno dell'Hololens.

²⁰ http://www.wikiholo.net/index.php?title=Performance_Tips

2.6. Altri Device (Competitors)

Sono diversi i dispositivi che ultimamente stanno uscendo nel mercato e che soprattutto cercano di competere con Microsoft ma come vedremo per ora HoloLens risulta un passo avanti rispetto a tutte le altre.

Di tutti i visori a realtà mista, HoloLens è l'unico utilizzabile al 100% con una piattaforma completa di utilizzo che è Windows 10.

L'altro visore che sta prendendo visibilità nell'ultimo anno è Meta 2, nato grazie alla campagna Kickstarter creata dal fondatore.

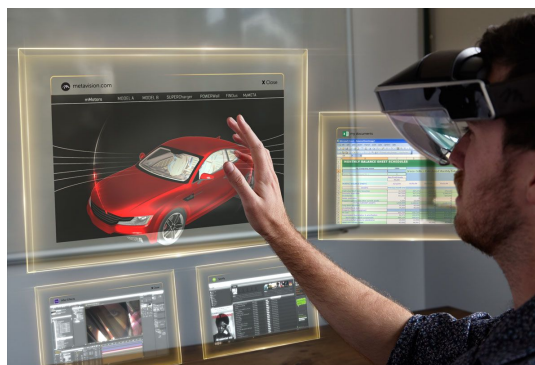
Questo visore, molto simile per diversi aspetti ad HoloLens, proietta nello stesso modo oggetti virtuali nelle lenti attaccate al visore e ne permette l'interazione con le mani.

La differenza sostanziale sta nell'approccio con cui avviene l'interazione.

Pensato per avere un'interazione il più naturale possibile, a causa del background di neuroscienze del fondatore, Meta 2 vuole un ambiente semplice ed efficace che il nostro cervello possa capire immediatamente con una curva di apprendimento più bassa possibile.²¹

Per questo motivo gli oggetti virtuali vengono considerati e trattati come oggetti fisici ed è possibile manipolarli senza gesti particolari come nel caso di HoloLens.²²

Meta 2 è riuscito a prendersi una fetta importante nel mondo della mixed reality in



quanto cerca di offrire novità rispetto a quanto già visto fino ad oggi.

Tutto ad un prezzo di 1000 dollari che, comparato con i 3000 di HoloLens lasciano sperare ad un futuro più economico.

Il vantaggio di Microsoft è stato puntare prima di tutti su questo tipo di dispositivi e

²¹ Fig. 2.9: Meta 2 Gesture

²²<https://techcrunch.com/2016/03/02/hands-on-with-the-949-mind-bending-meta-2-augmented-reality-headset/>

questo gli ha permesso di avere l'appoggio di importanti aziende tra cui l'ambiente Unity per lo sviluppo di applicazioni.

Meta 2 essendo ancora in fase di preordine e utilizzato da pochi non è stato ancora sfruttato appieno per essere considerato un dispositivo di riferimento.

Magic Leap, altro competitor che promette un'esperienza di mixed reality senza paragoni, a differenza di Meta 2 ha avuto immediatamente grossi finanziamenti da parte di aziende importanti come Google e Alibaba Group senza avere in mano prodotti o servizi.²³

Solo grazie a promesse e sorprendenti demo di alcuni minuti questa startup si conferma come uno dei prossimi passi tecnologici più interessanti dei prossimi anni.

Il visore di Magic Leap cerca di combinare la maggior potenza, risoluzione e usabilità di Meta 2 con la portabilità di Hololens per ottenere un'esperienza senza precedenti.

Già dalle poche demo presenti è possibile constatare i livelli di dettaglio strepitosi raggiunti nella definizione degli ologrammi e soprattutto l'incredibile praticità nel manipolarli e interagirci.

Per quanto bello possa sembrare appare comunque una realizzazione ancora lontana dall'utilizzo commerciale in quanto rimane nel mistero ogni dettaglio a riguardo come il prezzo e una data di uscita.

L'immagine sottostante cerca di elencare tramite poche proprietà le varie differenze tra questi tre visori.

	 HoloLens	 Meta 2	 Magic Leap
Standalone Unit?	Yes	No	Yes (but not yet)
Screen Resolution	1268x720	2560x1440	Unknown
Viewing Area	Limited Viewport	Full Area of Glasses	Full Area of Glasses
Battery Life	Approximately 3 Hours	N/A	Unknown
Operating System	Windows 10 (Holographic)	Prioprietary	Unknown
Work/Life Applications	Yes	Yes	Yes
Games	Yes	No	Yes
Cost (Dev Kit)	\$3,000	\$949	Unknown

Fig. 2.10: Confronto dispositivi di Mixed Reality

²³<https://www.forbes.com/sites/davidewalt/2016/11/02/inside-magic-leap-the-secretive-4-5-billion-startup-changing-computing-forever/#5da4cf7d4223>

3. Sviluppo con Hololens

Prima di iniziare a parlare di sviluppo, è necessario dire che Microsoft mette a disposizione una serie di risorse web che presentano un quadro d'insieme della tecnologia.

Tra queste troviamo la pagina per scaricare i software indispensabili (Visual Studio 2015 Update 3, l'Emulatore dell'Hololens²⁴ e una versione customizzata del motore grafico Unity) o i requisiti hardware e software necessari per riuscire ad iniziare il lavoro di sviluppo e molte altre ancora.

Nello spazio web chiamato "Holographic Academy²⁵" troviamo poi le indicazioni per approcciarsi allo sviluppo con questa tecnologia e familiarizzare step-by-step con le nuove caratteristiche della Mixed Reality, tramite video tutorial e mini corsi.

3.1. Unity

Per sviluppare viene usato principalmente Unity, un motore grafico nato per il gaming che permette di creare, importare e manipolare oggetti e modelli 3D, gestire eventi e creare scripts per gestire i comportamenti (behaviours) di ogni oggetto.

Inoltre è possibile controllare l'illuminazione, l'audio, le animazioni e ogni altro aspetto che si vuole rendere visibile all'esperienza dell'utente. Queste funzionalità permettono uno sviluppo eccellente non solo nell'ambito per cui è stato creato, cioè i videogiochi 2D/3D, ma anche per Hololens.

Con l'integrazione sulla versione 5.1 di Unity di un plugin di supporto VR si è aperta la strada dello sviluppo per visori.

Grazie a un accordo tra Microsoft e Unity è stato possibile installare su quest'ultimo un plugin dedicato che permette alcune funzionalità aggiuntive per lo sviluppo con Hololens, senza bisogno di extra SDK.

Questo mostra come Microsoft stia investendo nella crescita di questo visore.

²⁴ https://developer.microsoft.com/en-us/windows/mixed-reality/using_the_hololens_emulator

²⁵ <https://developer.microsoft.com/en-us/windows/mixed-reality/academy>

In più l'uso personale è gratis così da permettere a chiunque di poter cimentarsi senza problemi.

3.2. Applicazioni Olografiche per Hololens

Una volta creata una nuova scena in cui verranno visualizzati gli oggetti, la prima cosa da modificare è la Main Camera in quanto è impostata di default con dei valori di posizione e rotazione che andranno invece portati a 0, cioè corrispondenti ai valori richiesti del visore che diverrà il centro del proprio mondo olografico.

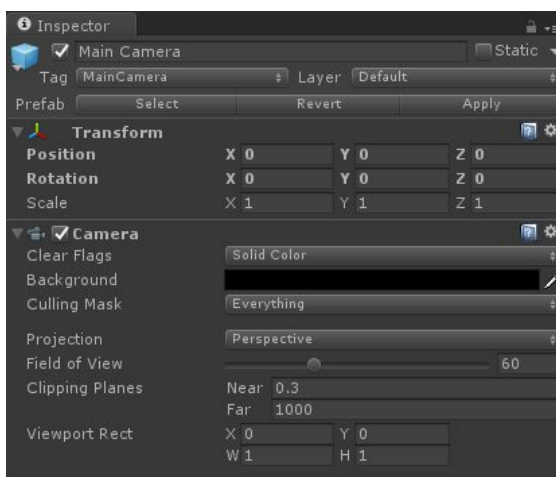
Inoltre, le impostazioni di default della Camera sono settate in automatico su quelle delle applicazioni 3D, in particolare è previsto un background (sfondo) su cui si vanno ad inserire gli oggetti. In Hololens invece è essenziale settare il background della camera sul colore nero in quanto il dispositivo renderizza il nero come “trasparenza” e in questo modo possano apparire solo gli oggetti che andremo ad inserire noi. In questo modo la mixed reality prenderà in considerazione solamente l'oggetto e non anche la “cornice” che contiene l'oggetto.

26

Inoltre Hololens ha un'esperienza di visione migliore se gli oggetti vengono renderizzati ad una distanza di circa 2

metri: Microsoft consiglia di impostare la *clipping render distance* (cioè la distanza minima e massima in cui un'ologramma viene visualizzato) tra 0,85 m e 5 m in modo tale da limitare l'affaticamento visivo. Su Unity questa impostazione può essere trovata nel menù della camera sotto il nome “Clipping Planes.”

La modifica più importante è però l'aggiunta dal menù “Player Settings” del supporto



²⁶ Fig. 3.1: Proprietà della Main Camera in Unity

alla realtà virtuale in modo da includere nel motore la Windows Holographic Sdk²⁷ che permette di aggiungere o rimuovere certe features che poi saranno necessarie per lo sviluppo dell'applicazione, come l'uso del microfono per lo Speech Recognition e la percezione dello spazio per lo Spatial Mapping.

Una volta costruita la *scena*, creati e posizionati gli oggetti necessari e aggiunti gli script per determinare certi comportamenti, è possibile estrarre il lavoro svolto su Unity per generare un progetto Visual Studio corrispondente. In questo progetto sono contenuti i tutti i dati del lavoro di sviluppo, quindi rappresenta l'archivio di codice, immagini e configurazioni che in Unity vengono invece mostrati all'utente in forma visiva. Da quel progetto è poi possibile lanciare l'applicazione creata, sia direttamente per mezzo del cavo USB sia inserendo l'indirizzo IP del visore ed eseguendola tramite wifi.

Oltre ad aiutare gli utenti alle prime armi con l'accademia citata prima, Microsoft mette a disposizione su Github, il servizio di hosting per progetti software, l'HoloToolkit²⁸ (con wiki annessa) cioè un insieme di script, oggetti preimpostati, materials e ombre così da semplificare ancora di più lo studio e lo sviluppo.

Visto il prezzo del visore non proprio accessibile a chiunque, Microsoft ha pensato di permettere ai programmatori di iniziare a sviluppare in modo completo, seppure limitato, anche senza un dispositivo fisico.

Infatti poco dopo l'uscita del device e dei software di sviluppo è stato rilasciato l'*Hololens Emulator* che simula il dispositivo, compresi i sensori di input attraverso l'uso della tastiera, del mouse o di un controller Xbox.

Dopo aver guardato al processo di creazione, una questione che molti vorranno approfondire riguarda le scelte progettuali da trattare, nel senso pratico di come impostare i comportamenti (behaviours) degli oggetti virtuali creati, in relazione ai vari metodi di interazione tra essi e rispetto alle gestures a disposizione.

Questo comporta di ripensare in modo diverso l'esperienza utente.

²⁷ https://developer.microsoft.com/en-us/windows/mixed-reality/install_the_tools

²⁸ <https://github.com/Microsoft/HoloToolkit-Unity>

3.3. Primi Sviluppi²⁹

In una delle mie prime esperienze, lo sviluppo di un'applicazione ha incontrato alcune problematiche in questo contesto. Davanti alla necessità di progetto di ricondurre ad un solo oggetto (un'auto) più funzionalità (per esempio l'apertura di un menù che permettesse l'interazione con altre funzionalità e lo spostamento di alcuni



componenti dell'auto) è emerso il limite di poter utilizzare solo una gesture (la più semplice e immediata chiamata "air-tap") e quindi non potervi associare le diverse funzioni senza compromettere l'univocità del comando.

Davanti a questa complicazione ho elaborato la soluzione di inserire un nuovo oggetto complementare (una piattaforma sotto la macchina) così da dividere le funzionalità in diversi sotto-oggetti e non creare più sovrapposizioni.

Questa difficoltà nel determinare una struttura precisa si incontra nelle applicazioni sviluppate fino ad adesso (soprattutto firmate Microsoft) che sono ancora piuttosto semplici e molto lineari, pur mostrando un innegabile (e innovativa) qualità del grande ed interessante investimento sul lato della costruzione e definizione degli oggetti che coinvolge un team molto consistente formato da visual designers, artisti e sviluppatori. Il lato tecnico verrà implementato a breve, ma l'impatto della visione ci sta già mostrando le incredibili prospettive di questa tecnologia.

3.4. Usi

Adesso saranno elencati esempi di settori in cui la rappresentazione 3d di oggetti ha permesso di capire ad un livello più profondo cosa è possibile fare.

²⁹ Fig. 3.2: Ologramma di una macchina

- Educativo/training

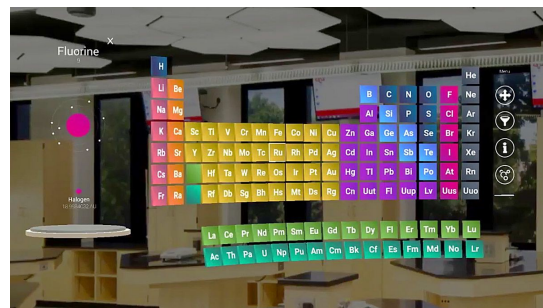
E' facile immaginare come nell'ambito in questione sia un forte vantaggio l'uso della mixed reality.

Come il detto “un'immagine vale più di mille parole” è anche vero che una rappresentazione 3d che è possibile vedere in ogni angolazione e interagirci come se fosse un oggetto reale, vale più di un'immagine.

Infatti è già utilizzato per questo scopo da diverse Università che stanno cercando di sfruttare il massimo potenziale in questo ambito, grazie anche alla possibilità di condividere quello che si sta vedendo con altre persone.³⁰

Non è un caso che, oltre all'educazione, anche nei contesti di training abbia un ampio utilizzo.

Volvo e Japan Airlines fanno utilizzare ai dipendenti il dispositivo per capire l'uso di complesse macchine industriali in quanto incrementa significativamente la curva di apprendimento grazie ad un approccio immediato e finalizzato a quello scopo.³¹



- Videoludico

Il videoludico è un settore in fortissima crescita soprattutto negli ultimi anni e l'utilizzo della realtà aumentata si accoppia perfettamente a questo scopo.

Come abbiamo visto con il successo di Pokémon Go, l'impiego di questa realtà porta ad un livello di esperienza maggiore e soprattutto diverso da quanto visto fino ad adesso.

Hololens rompe i confini di un semplice smartphone portando sempre il solito tipo di realtà ma con un'immersione molto più profonda.

La domanda è quanto riuscirà a spopolare in questa direzione visto l'utilizzo principale non proprio pensato a questo campo e soprattutto data la disponibilità e usabilità ai soli sviluppatori. Infatti la questione del prezzo di 3000\$ fa tenere in considerazione come

³⁰ <http://case.edu/hololens/>

³¹ Fig. 3.3: Ologramma della tavola periodica degli elementi

attualmente sia sia un dispositivo riservato a poche persone e soprattutto aziende che si possono permettere questa cifra ingente.

- Medico

Tante sono le sperimentazioni in corso riguardanti questo campo nell'ultimo anno.³²

Nell'ambito della diagnosi è facile pensare come sia molto più semplice ed efficace l'utilizzo di ologrammi che mostrano in maniera molto meno limitata rispetto al cartaceo o anche solo a schermo.

Molti ospedali stanno cercando di implementare e accompagnare un'operazione con questi tipi di dispositivi in modo da renderla più sicura e praticabile.



Fig. 3.4: Utilizzo Hololens in sala operatoria

- Architettura e Design

Anche in questi settori sta prendendo piede molto velocemente l'uso di Hololens.³³

Gli architetti in particolare possono usare questa tecnologia per visualizzare gli interni di un edificio quando ancora non è stato definito e formato totalmente.

A differenza della realtà virtuale, che crea un ambiente totalmente immersivo, la mixed reality porta una migliorata percezione e un modo di interagire molto più immediato così da fare delle scelte implementative, di design e non, migliori.

³² <https://www.theverge.com/2017/5/5/15557790/scopis-medical-microsoft-hololens-ar-spine-surgery>

³³ <https://news.microsoft.com/en-gb/2017/03/09/world-famous-architect-sees-his-buildings-recreated-using-hololens/>

L'applicazione più rilevante in questo contesto è sicuramente SketchUp, software conosciuto per il 3d Modeling utilizzato da milioni di persone.



Fig. 3.5: Utilizzo Hololens nel settore del design

- Militare

Microsoft potrebbe non aver pensato a questo utilizzo ma c'è già chi se ne sta occupando facendo sperimentazioni.³⁴

Infatti già diverse aziende, tra cui LimpidArmor, ha iniziato ad includere nel suo equipaggiamento militare questo dispositivo.

Questo elmetto speciale include un puntamento automatico del target, l'individuazione dei nemici con la particolarità di etichettarli in modo differente dagli alleati e mostrare posizioni condivisibili con altri.

Anche se il sistema è ancora in fase concettuale capiamo che in futuro una realtà simile potrebbe vedere la luce anche nei campi di battaglia.



Fig. 3.6: Casco con Hololens integrato per utilizzo militare

³⁴ <https://www.inverse.com/article/23183-limpid-armor-ukraine-military>

4. Progetto: LandMark Recognition

4.1. Introduzione

Lo scopo di questo progetto è la realizzazione di un'applicazione per Hololens che permetta di riconoscere un monumento, visto sia realmente sia in foto, e far visualizzare all'utente che indossa il dispositivo delle informazioni riguardanti esso.

Grazie alla presenza di una fotocamera incorporata nel dispositivo è stato considerato di scattare in tempo reale una foto ad un monumento per poi passarla ad un servizio cloud di riconoscimento che capisca di quale struttura si tratti così da estrapolare le informazioni relative da un altro servizio.

Più orientata verso le realtà aumentata, vedremo che per questa applicazione si è deciso di sfruttare una tecnologia molto recente per l'immagine recognition, Google Cloud Vision. In questo capitolo si descriverà il suo funzionamento e il motivo per la quale è stato utilizzato in particolare le API di Google Cloud Vision, le API utilizzate per il prelevamento delle informazioni e il progetto Unity relativo creato per l'applicazione.

4.2. API per il riconoscimento di immagini

Con la moltitudine di servizi di Image Recognition ultimamente presenti nel mercato è stato deciso di ridurre la scelta ai colossi dell'IT del momento: Microsoft e Google.

Entrambi offrono e mettono a disposizione servizi di interpretazione e classificazione delle immagini che sfruttano il Machine Learning.³⁵

L'idea più semplice per decidere quale delle due utilizzare è stata provare le API di entrambi. Questo ha permesso di capirne le funzionalità, l'accuratezza e la precisione di ciò che viene rilevato.

Mettiamo adesso a confronto i due servizi utilizzando l'immagine delle due Torri di Bologna.

³⁵<https://www.altexsoft.com/blog/datascience/comparing-machine-learning-as-a-service-amazon-microsoft-azure-google-prediction-api/>

Microsoft Computer Vision API:



Fig. 4.1: Utilizzo delle API Microsoft

Google Cloud Vision API:

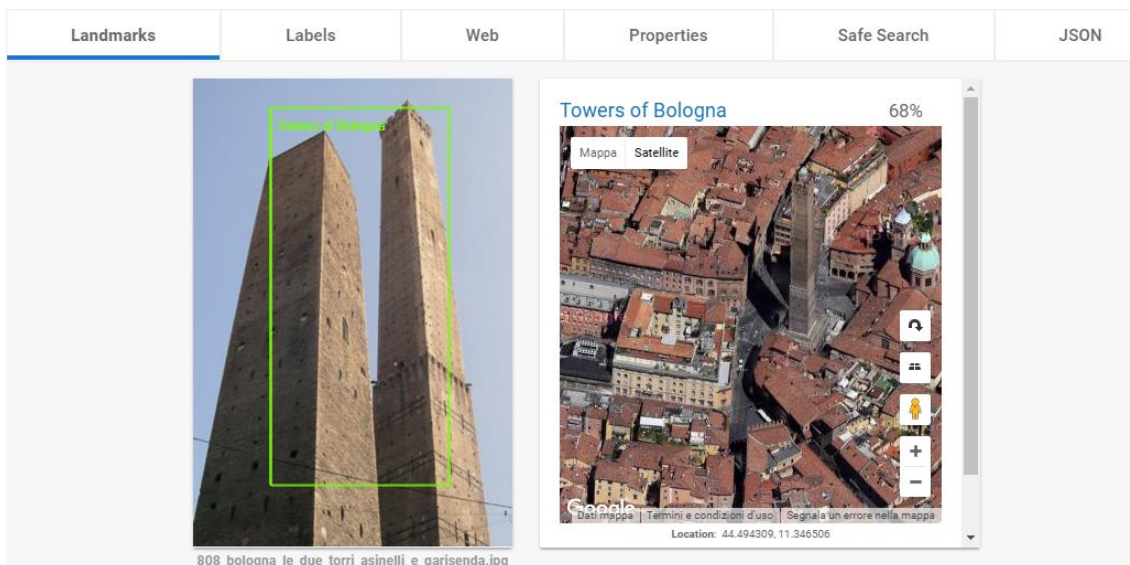


Fig. 4.2: Utilizzo delle API Google

Come riscontriamo subito facendo un veloce confronto, le API di Microsoft riescono a riconoscere che si tratta di una costruzione ma non ad etichettarlo e identificarlo esattamente come previsto.

Al contrario, molto più efficacemente, Google Cloud Vision non solo riconosce di quale monumento si tratta ma riesce anche a localizzarlo nell'immagine (come vediamo a

sinistra dal riquadro verde) e nella mappa grazie alle coordinate geografiche che si ricava (come vediamo a destra).

Questa forte limitazione di Microsoft è dovuta al loro modello di “landmark recognition” che riconosce solamente 9000 monumenti (luoghi storici) sia naturali che fatti dall’uomo.³⁶

Proprio a causa di questa restrizione è stata presa la decisione di utilizzare Google Cloud Vision.

4.3. API: Google Cloud Vision³⁷

Per il motivo scritto in precedenza sarà approfondito nel dettaglio questo strumento.



Google Cloud Vision API permette di “comprendere” il contenuto delle immagini attraverso delle REST API.

Per fare ciò vengono utilizzati dei performanti modelli di Machine Learning che non hanno rivali per accuratezza e performance ad altri sistemi di deep learning.

Gli stessi modelli che sono utilizzati per il “cerca” in Google Photos o per il “ricerca tramite immagine” di Google Images.³⁸

Per riuscire a realizzare il riconoscimento in particolare dei monumenti, le API di Google fanno riferimento ad un database di oltre 70 mila monumenti utilizzando un “Knowledge Graph” popolato analizzando le foto pubbliche.³⁹

Per “Knowledge Graph” si intende un sistema che mette immagini, persone e fatti insieme per creare risultati di ricerca che sono più indicati ed accurati.

Ogni riga di questo database include il nome, la posizione e il “Knowledge Graph” di ogni monumento. Questo dati vengono continuamente aggiornati per includere nuovi monumenti.

³⁶ <https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/>

³⁷ Fig. 4.3: Logo di Google Cloud Platform

³⁸ <https://cloud.google.com/products/>

³⁹ <https://cloud.google.com/blog/big-data/2016/09/around-the-world-landmark-detection-with-the-cloud-vision-api>

Il processo di funzionamento è rappresentato da questa immagine:

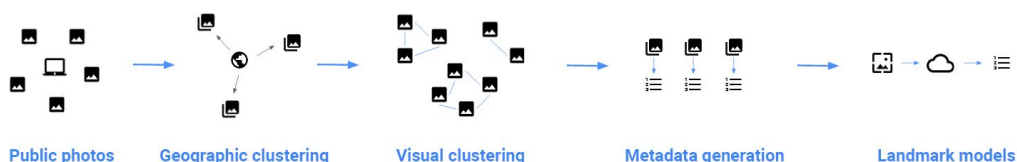


Fig. 4.4: Sistema per il riconoscimento di luoghi storici

Le funzionalità principali di Google Cloud Vision API

Di seguito le funzionalità più importanti che vengono offerte:

- riconoscimento di oggetti, monumenti, animali, celebrità, loghi, eventi e cose
- analisi dei sentimenti, in quanto rilevata una faccia di una persona può individuare felicità, rabbia, tristezza e altri.
- estrazione di testi all'interno dell'immagine in stile OCR (Optical Character Recognition)
- analisi dei contenuti in modo tale da riconoscere quello inappropriato come immagini offensive, violente o solamente per pubblico adulto

Per compiere il riconoscimento di monumenti è necessario fare una richiesta in POST attraverso l'uso della propria chiave API:

POST https://vision.googleapis.com/v1/images:annotate?key=YOUR_API_KEY

La richiesta da effettuare è un JSON, un formato adatto all'interscambio di dati fra applicazioni client-server, con questa forma:

```

{
  "requests": [
    {
      "image": {
        "content": "...",
      },
      "features": [
        {
          "type": "LANDMARK_DETECTION"
        }
      ]
    }
  ]
}

```

Nella richiesta è da notare in particolare la chiave “content” che conterrà l’immagine codificata in base 64 per poter essere mandata al servizio e la chiave “type” che specificherà il tipo di riconoscimento da effettuare.

Se la richiesta ha successo, il server ritorna un codice di stato HTTP **200 OK** e la risposta sarà, nello stesso modo, un JSON come il seguente:

```

{
  "responses": [
    {
      "landmarkAnnotations": [
        {
          "mid": "/m/02qv838",
          "description": "Towers of Bologna",
          "score": 0.6832165,
          "boundingPoly": {
            "vertices": [
              {
                "x": 259,
                "y": 129
              },
              ...
            ]
          }
        }
      ]
    }
  ]
}

```

```

    },
    "locations": [
      {
        "latLng": {
          "latitude": 44.494309,

          "longitude": 11.346506
        }
      }
    ]
  }
]
}
]
}
]
}
]
}
]
}

```

La chiave “mid” è un ID per l’entità nel “Knowledge Graph” di Google, “description” è il nome del monumento e “score” è un valore che oscilla da 0 a 1 che indica la sicurezza dell’API che quel monumento si tratta veramente di quello esatto.

La chiave “boundingPoly” contiene i vertici che identificano l’esatta regione dell’immagine in cui si trova esattamente il monumento trovato.

La latitudine e la longitudine indica le esatte coordinate geografiche del monumento.

Il JSON sopra scritto non è ovviamente completo in quanto molti altri campi sono specificate dalle API ma che al momento non sono rilevanti per lo scopo del progetto in questione.

Tali campi sono ad esempio “labelAnnotations”, “safeSearchAnnotation”, “dominantColors”, “fullMatchingImages” etc.

Nell’applicazione esposta successivamente sarà estrapolato dal JSON di ritorno la chiave “description” che ha come valore il nome del monumento per usarlo come chiave di ricerca per estrapolare in seguito delle informazioni riguardanti esso.

4.4. API: Wikipedia⁴⁰

Tutti sanno che Wikipedia è l'enciclopedia multilingue online, gratuita e libera più utilizzata al mondo.

Per quanto sia stata criticata per i contenuti non del tutto affidabili, in quanto non può assicurare l'attendibilità dei testi perché per sua natura è uno strumento dinamico, è stato



deciso di affidarsi a questo strumento in quanto è da molti anni ormai che è possibile accedere ai contenuti di Wikipedia anche tramite API.

E' per questo motivo che per lo sviluppo dell'applicazione viene utilizzato questo servizio web per scaricare i contenuti che ci servono riguardo ad un monumento.

E' doveroso dire che l'applicazione creata è fondamentalmente un concept funzionante su una tecnologia molto giovane e acerba che potrebbe appoggiarsi ad altri servizi e che ha molte limitazioni.

Di seguito si spiegherà l'utilizzo fatto delle API di Wikipedia per riuscire ad estrapolare le informazioni necessarie per essere visualizzate a video dall'utente.

L'API di azione MediaWiki è un servizio web che fornisce un conveniente accesso alle funzioni della wiki, ai dati ed ai meta-dati attraverso il protocollo HTTP, utilizzando generalmente un URL in `api.php`

L'URL utilizzato per fare una GET che ritornasse nuovamente un JSON è:

```
https://en.wikipedia.org/w/api.php?format=json&action=query&prop=extracts&explaintext=1&titles=Towers_of_Bologna
```

Andiamo adesso ad analizzare i parametri utilizzati per ricavare esattamente il JSON voluto:

- `format`: MediaWiki offre diversi tipi di formato in output come JSON, JSONFM, PHP, PHPFM, XML, XMLFM e RAWFM.

Per semplicità è stato deciso di utilizzare JSON.

⁴⁰ Fig. 4.5: Logo Wikipedia

- action: Permette di compiere diverse azioni. Le principali sono “help”, “login”, “opensearch”, “feedwatchlist”, “query”.

Il modulo action=query permette di ricevere informazione su una wiki e le informazioni immagazzinate in esso come ad esempio il testo di una particolare pagina, lin e categorie di una serie di pagine etc.

Quindi il modulo in questione ha appunto dei sottomoduli che possono essere utilizzati. Tra questi i principali sono “Meta Information”, “Properties”, “Lists”.

- prop: Una delle tante proprietà che sono disponibili è “extracts” che permette di fatto di far ritornare nel JSON anche il testo contenuto nella pagine wiki corrispondente.
- explaintext: Fa visualizzare il testo in formato normale e non in HTML in quanto non c’era necessità di utilizzarlo.
- titles: Decisamente il campo più importante, decide quale pagina di wiki scegliere in base alla chiave di ricerca che viene messo proprio in questo campo. E’ esattamente l’output che viene dal JSON manipolato di Google Cloud Vision che abbiamo visto nel capitolo precedente. In base a quella risposta verrà poi inserito la chiave da cercare e poi visualizzare.

Ovviamente nel caso in cui il riconoscimento dell’immagine con Google non dovesse funzionare, la chiamata a Wikipedia non restituirà nulla e quindi andrà gestito di conseguenza il comportamento del software.

4.5. Descrizione dell’applicazione

In questo capitolo sarà analizzato il progetto Unity creato per l’applicazione passo per passo delineando i requisiti, l’architettura, il settaggio del progetto e della scena e il codice principale sviluppato.

Tutto il progetto e in particolare il codice che sarà visualizzabile di seguito è possibile trovarlo al profilo personale di Github:

<https://github.com/carloFanc/LandMarkRecognition>

I requisiti necessari per sviluppare l'applicazione sono:

- Microsoft Hololens
- Emulatore dell'Hololens
- Unity 5.6.1f1
- Account di Google Cloud Platform
- HoloToolKit
- Framework "Json.NET"⁴¹

L'architettura con il relativo diagramma di sequenza:

1. Visualizzazione di una semplice interfaccia che mostra all'utente la possibilità di scattare una foto. Inizializzazione della camera di Hololens.
2. Attraverso un comando vocale viene scattata la foto che comprende ciò che l'utente sta visualizzando con il dispositivo.
3. Viene disattivata l'interfaccia primaria, preparata e inviata la richiesta al servizio cloud che ritorna il JSON da manipolare.
4. Attraverso il framework Json.NET viene estrapolato il nome del monumento e visualizzato a video.
5. Viene preparata e inviata la richiesta di Wikipedia che avrà come parametro il nome del monumento precedente.
6. Viene visualizzato a video, tramite un pannello, il testo contenuto nel JSON di ritorno della chiamata precedente.
7. Tramite un altro comando vocale è possibile ritornare all'interfaccia primaria che permette di scattare un'altra foto.

⁴¹ <http://www.newtonsoft.com/json>

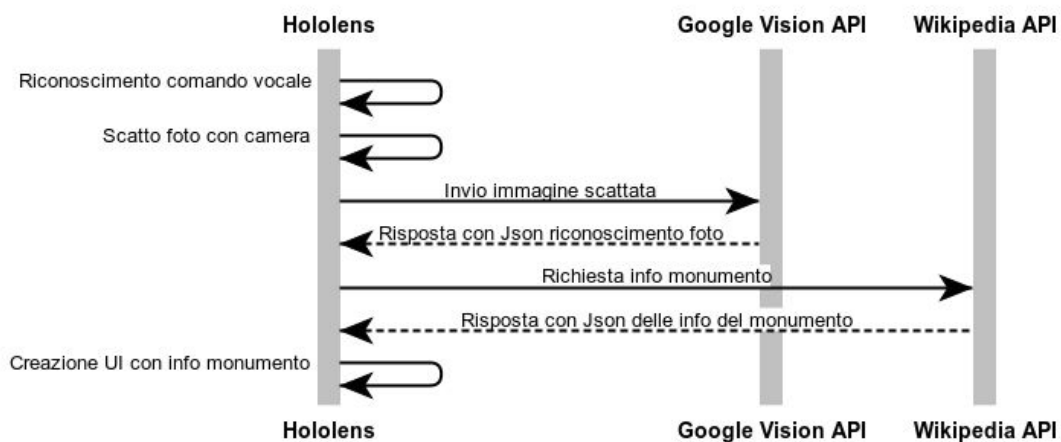


Fig. 4.6: Diagramma di sequenza dell'applicazione

Nella foto di seguito vediamo l'IDE di Unity formato dalla gerarchia degli oggetti a sinistra, la scena nella schermata centrale, i componenti del progetto in basso.

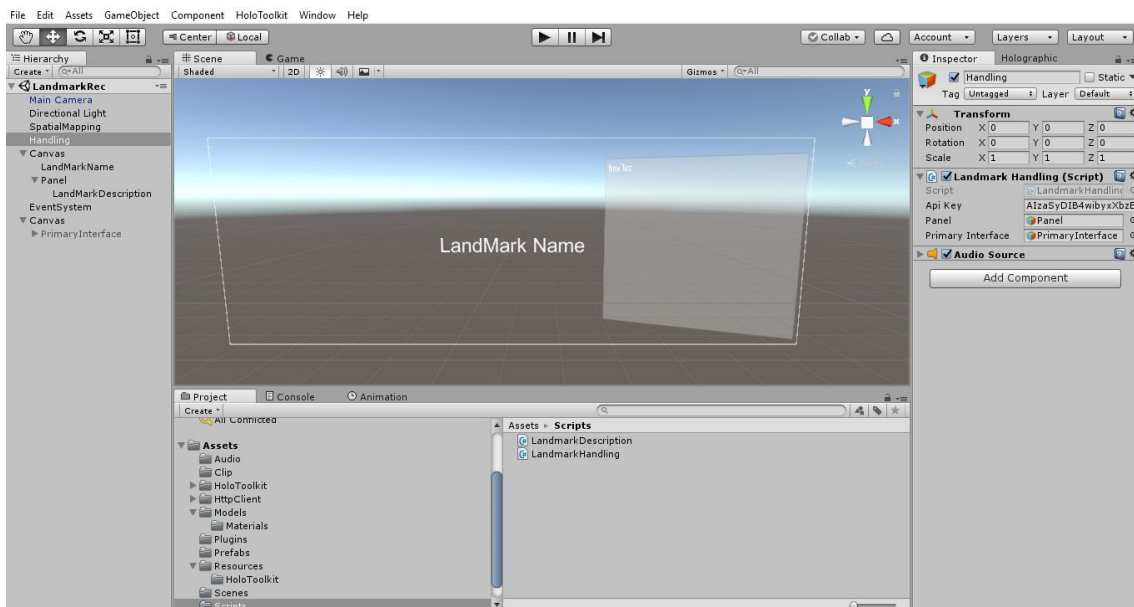


Fig. 4.7: Interfaccia del progetto Unity

Dopo la creazione del progetto Unity è necessario definirlo come app Universal Windows Platform attraverso i seguenti passaggi:

1. Selezionare File > Build Settings
2. Selezionare Windows Store nella lista delle piattaforme e aggiungerla come predefinita

3. Scegliere come SDK “Universal 10”
4. Scegliere come dispositivo utilizzato “Hololens”
5. Scegliere come tipo di Build “D3D”
6. Scegliere come UWP SDK “Last Installed”

Successivamente è necessario far supportare al progetto la Realtà Virtuale per creare contenuti 3d. Per fare ciò:

1. Selezionare File > Build Settings > Project Settings
2. Selezionare le impostazioni nella tab per Windows Store
3. Espandere “Other Settings”
4. Nella sezione Rendering spuntare il checkbox “Virtual Reality Supported” e avere la conferma che “Windows Holographic” sia nella lista.

Per poter permettere l’utilizzo di certe funzionalità nel progetto è opportuno seguire i seguenti passi:

1. Selezionare Edit > Project Settings > Player > Windows Store > Publishing Settings > Capabilities

Le funzionalità inserite per rendere operativa l’applicazione sono le seguenti:

- InternetClientServer
- Microphone
- Webcam
- Spatial Perception

Brevemente verranno ora descritti i passaggi per settare la scena con i giusti componenti nella gerarchia:

- Import della libreria “HoloToolkit” da utilizzare per l’inserimento sia della Main Camera configurata nel modo adatto a funzionare con Hololens sia dello script “Spatial Mapping Collider” che permette l’interazione degli ologrammi con la realtà rilevata dai sensori.

- Import della versione per Unity del framework “Json.NET” per manipolare i JSON ed estrarne i dati contenuti.
- Creazione di un GameObject vuoto chiamato “SpatialMapping” dove inserire lo script sopra citato e di un altro GameObject vuoto chiamato “Handling” dove inserire lo script principale per la comunicazione con il servizio di Google Cloud.
- Creazione di un Canvas dove verrà inserita l’interfaccia primaria e di un altro Canvas contenente il GameObject con il nome del Monumento e un pannello per la descrizione.

Il pannello conterrà il secondo script con all’interno la chiamata al servizio di Wikipedia.

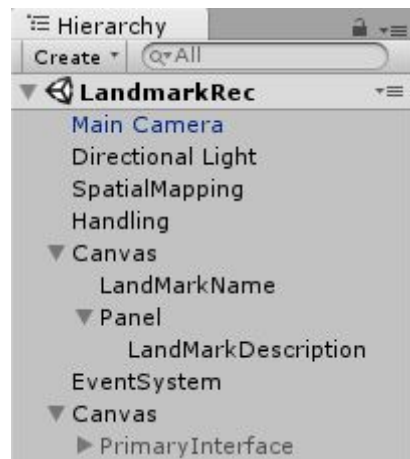


Fig. 4.8: Gerarchia del progetto Unity

Analizziamo adesso lo script principale chiamato “LandMarkHandling”.

In questo script C# viene scattata la foto da Hololens e mandata la richiesta alle API di Google:

```
public void Start()
{
    panel.SetActive(false);
    text = GameObject.Find("LandMarkName").GetComponent<Text>();
    text.text = "";
    PrimaryInterface.SetActive(true);
    audio = GetComponent<AudioSource>();
}
```

In questo metodo, utilizzato in Unity quando viene eseguito la prima volta lo script, si inizializza le variabili principali “text” e “audio” rispettivamente per settare a vuoto l’oggetto che sarà popolato dal nome effettivo del monumento e per settare il componente audio che verrà poi utilizzato per produrre un suono che farà capire che lo scatto della foto è avvenuto.

Viene attivata, per essere visualizzata dall’utente, l’interfaccia primaria e disattivato il pannello che conterrà la descrizione del monumento.

```
public void StartTheFunction()
{
    audio.Play();
    PrimaryInterface.SetActive(false);
    cameraResolution =
    PhotoCapture.SupportedResolutions.OrderByDescending((res) => res.width
    * res.height).First();
    PhotoCapture.CreateAsync(false, delegate (PhotoCapture
    captureObject)
    {
        photoCapt = captureObject;

        CameraParameters cameraPar = new CameraParameters();
        cameraPar.hologramOpacity = 0.0f;
        cameraPar.cameraResolutionWidth = cameraResolution.width;
        cameraPar.cameraResolutionHeight =
        cameraResolution.height;
        cameraPar.pixelFormat = CapturePixelFormat.JPEG;

        photoCapt.StartPhotoModeAsync(cameraPar, delegate
        (PhotoCapture.PhotoCaptureResult result)
        {
            PictureShoot();
        });
    });
}
```

In questa funzione viene riprodotta la traccia mp3, che simula il suono dello scatto di una foto, disattivata l’interfaccia primaria e attivata la fotocamera con i vari settaggi specificati in “CameraParameters”.

Può esistere solo un’istanza alla volta di PhotoCapture, per questo una volta utilizzata

tramite un'altro metodo qua non descritto viene terminata la modalità foto tramite `StopPhotoModeAsync`.

```
void PictureShoot()
{
    if (photoCapt != null)
    {
        photoCapt.TakePhotoAsync(delegate
(PhotoCapture.PhotoCaptureResult result, PhotoCaptureFrame
photoCaptureFrame)
        {
            List<byte> buffer = new List<byte>();
            Matrix4x4 cameraToWorldMatrix;
            photoCaptureFrame.CopyRawImageDataIntoBuffer(buffer);

            if (!photoCaptureFrame.TryGetCameraToWorldMatrix(out
cameraToWorldMatrix))
            {
                return;
            }
            StartCoroutine(PhotoUploading(buffer.ToArray()));
        });
    }
}
```

E' in questa funzione che viene scattata e salvata la foto tramite il metodo `TakePhotoAsync`, in un buffer come lista di byte.

L'immagine sarà poi utilizzata come parametro per la funzione successiva "PhotoUploading".

Prima di passare però alla descrizione di quella funzione, analizziamo la `UnityWebRequest` che verrà richiamata successivamente.

```
UnityWebRequest RequestCreation(byte[] photo)
{
    DownloadHandler download = new DownloadHandlerBuffer();

    string base64image = Convert.ToBase64String(photo);
    string json = "{\"requests\": [{\"image\": {\"content\": \"" +
base64image + "\"}, \"features\": [{\"type\":
\"LANDMARK_DETECTION\", \"maxResults\": 5}]}}\"";
```

```

        byte[] content = Encoding.UTF8.GetBytes(json);
        UploadHandler upload = new UploadHandlerRaw(content);
        string url =
"https://vision.googleapis.com/v1/images:annotate?key=" + m_apiKey;
        UnityWebRequest www = new UnityWebRequest(url, "POST",
download, upload);
        www.SetRequestHeader("Content-Type", "application/json");

        return www;
    }
}

```

Anche qua, viene passata come parametro la foto scattata così da convertirla in base 64 in modo tale da essere messa nel JSON da inviare al servizio cloud.

L'operazione più importante è sicuramente la UnityWebRequest che tramite i tre parametri, che sono l'url con l'apposita chiave, il DownloadHandler e l'Upload con il JSON contenente l'immagine, viene spedita per ritornare successivamente il JSON finale.

```

IEnumerator PhotoUploading(byte[] photo)
{
    using (UnityWebRequest www = RequestCreation(photo))
    {
        yield return www.Send();

        if (www.isError)
        {
            Debug.Log("errore" + www.error);
        }
        else
        {
            string jsonString = www.downloadHandler.text;

            JsonTextReader reader = new JsonTextReader(new
StringReader(jsonString));

            int count = 0;
            mon = null;
            while (reader.Read())
            {
                if (reader.Value != null)
                {
                    count++;
                }
            }
        }
    }
}

```

```
                if (count == 6)
                    mon = reader.Value.ToString();
            }
        }

        if (mon != null) {
            text.text = mon;
        }else
        {
            text.text = "Not found, Sorry. Retry.";
        }
        if (!text.text.Equals("Not found, Sorry. Retry.")) {
            OnDestroy();
            StartCoroutine(LateCall());
        }
    }
}
}
```

In questo metodo viene ritornato il JSON in risposta al metodo visto in precedenza e manipolato tramite il framework Json.NET.

In questo caso reader.Value.ToString() conterrà esattamente il nome del monumento, ovviamente nel caso in cui il riconoscimento è andato a buon fine.

Se quindi la variabile “mon” includerà il nome allora sarà assegnato al GameObject “text” che sarà visualizzabile all’utente.

```
IEnumerator LateCall()
{
    yield return new WaitForSeconds(2);
    panel.SetActive(true);
}
```

In seguito viene lanciata l’ultima funzione di questo script che andrà ad attivare il pannello.

```
void OnEnable()
{
    text = GameObject.Find("LandMarkName").GetComponent<Text>();
    txtRef = GetComponent<Text>();
    txtRef.text = "";
    StartCoroutine(GetText());
}
```

L'attivazione del pannello farà partire lo script relativo con il metodo OnEnable che si esegue esattamente quando l'oggetto correlato viene attivato.

E' proprio in questa funzione che viene ripresa la variabile utilizzata nello script precedente che racchiude il nome del monumento.

Tale variabile sarà usata come parametro e parola chiave su Wikipedia.

A seguire viene eseguito GetText().

```
IEnumerator GetText()
{
    UnityWebRequest www =
UnityWebRequest.Get("http://en.wikipedia.org/w/api.php?format=json&act
ion=query&prop=extracts&explaintext=1&titles="+ text.text );
    yield return www.Send();

    if (www.isError)
    {
        Debug.Log(www.error);
    }
    else
    {
        JObject o = JObject.Parse(www.downloadHandler.text);

        txtRef.text = "\n" +
o["query"]["pages"].First.Last.Last.Last;

    }
}
```

Viene preparata e inviata la richiesta al servizio web che ritornerà sempre un JSON che sarà estratto per prelevare il testo relativo.

Infine il pannello conterrà tale testo grazie ad una semplice assegnazione.

4.6. User Experience

In questa figura si nota l'interfaccia primaria che inquadra una foto presa da Internet delle due Torri di Bologna. E' stato deciso di utilizzare una foto a causa di difficoltà di utilizzo dell'Hololens al di fuori della sede di lavoro.



Fig. 4.9: Interfaccia primaria per lo scatto della foto

Disattivata l'interfaccia primaria compare il nome del Monumento con il pannello descrittivo sulla destra.



Fig. 4.10: Visualizzazione in primo piano del nome del monumento identificato

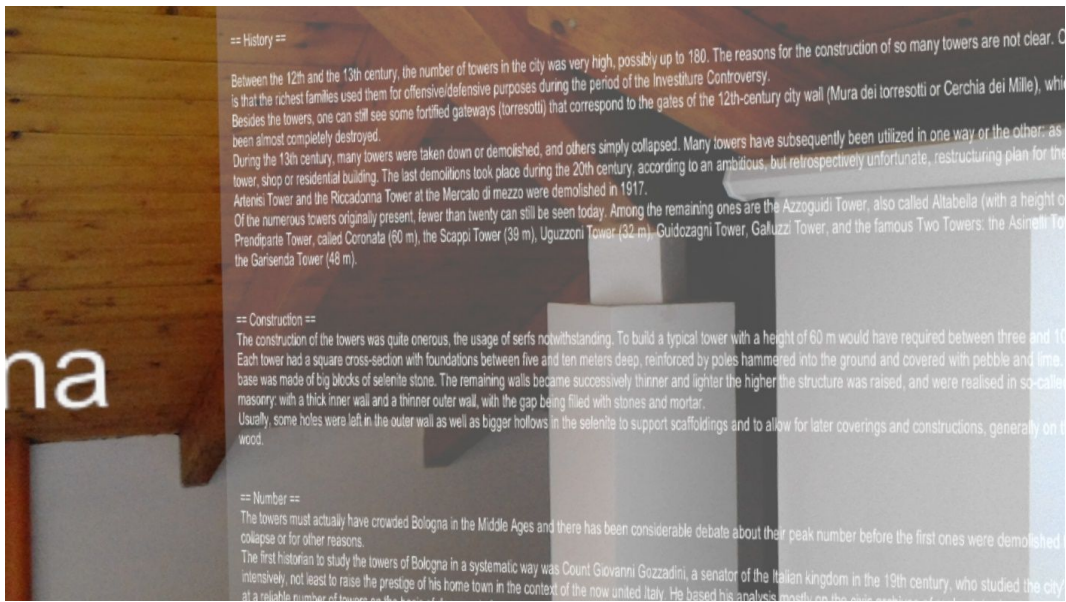


Fig. 4.11: Pannello contenente tutte le informazioni prese da Wikipedia

Conclusioni

Grazie ad un tirocinio svolto in azienda mi è stata data un'opportunità che pochissime persone hanno potuto avere, specialmente in Italia.

Grazie ad esso ho potuto scoprire una nuova tecnologia che si sta diffondendo sempre di più nel mondo e che tra pochi anni avrà modo di espandersi sempre di più.

Come abbiamo visto, l'interesse è alto e le aspettative di tutti sono molte e ad un solo anno dall'uscita del dispositivo.

Benché Hololens non sia il formato conclusivo, o comunque l'obiettivo finale, possiamo confermare che è l'inizio di una nuova realtà che potrebbe entrare a far parte della nostra vita tra pochissimi anni.

Già adesso è parte integrante di moltissime aziende che hanno deciso di investirci concretamente per migliorare aspetti per cui finora era impossibile o molto difficoltoso.

Tutto questo mi ha permesso di lavorare con una tecnologia mai vista prima e imparare software e linguaggi che non conoscevo, con annesso le difficoltà che l'innovazione porta avendo a che fare con problemi sempre nuovi e mai visti in precedenza.

Per questo ringrazio chi mi ha permesso di intraprendere questo viaggio sperando di poter continuare e approfondire il mondo della realtà aumentata e virtuale.

Bibliografia

Allen G. Taylor, Develop Microsoft HoloLens Apps now

<https://docs.unity3d.com/Manual/index.html>

[https://en.wikipedia.org/wiki/Unity_\(game_engine\)](https://en.wikipedia.org/wiki/Unity_(game_engine))

<https://www.microsoft.com/en-us/hololens>

<http://www.getconnected.it>

<http://www.imaginativeuniversal.com/blog/2015/10/18/how-hololens-displays-work/>

<https://cloud.google.com/vision/>

<https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/>

<https://azure.microsoft.com/it-it/blog/microsoft-cognitive-services-general-availability-for-face-api-computer-vision-api-and-content-moderator/>

Will Goldstone, Unity Game Development Essentials

<https://developer.microsoft.com/it-it/windows/mixed-reality/>

https://en.wikipedia.org/wiki/Microsoft_HoloLens

<https://www.foundry.com/industries/virtual-reality/vr-mr-ar-confused>

<https://www.theverge.com/2016/1/16/10778864/microsoft-hololens-battery-life-features>

<https://support.microsoft.com/en-us/help/12630/hololens-cortana-on-hololens>

<https://www.theverge.com/2016/3/30/11333192/microsoft-hololens-partners-augmented-reality-boring>

<https://techcrunch.com/2016/10/10/duke-neurosurgeons-test-hololens-as-an-ar-assist-on-tricky-procedures/>