Título

Kinectulum

3D Object reconstruction using Kinect and mirrors

3D Object reconstruction using a single Kinect and mirrors

3D Image Acquisition using a Static Setup

3D Information Acquisition of Small Daily Objects

3D Full Data Acquisition a Static Setup

3D Panoramic Data Acquisition using a Static Setup

3D Concentric Data Acquisition using a Static Setup

3D 360 Data Acquisition using a Static Setup

Head

Thanks

You’re welcome.

Abstract

Things

What did you do?

Why did you do it? What question were you trying to answer?

How did you do it? State methods.

What did you learn? State major results.

Why does it matter? Point out at least one significant implication.

Resumo

Pretende-se com esta dissertação desenvolver um sistema de baixo custocapaz de fazer a captura panorâmica 3D de um objeto de pequena/média dimensão. Este sistema deve ser capaz de executar essa tarefa utilizando apenas uma câmara e fazê-lo de forma estática, isto é, sem ser necessário colocar a câmara em diferentes posições e como tal, realizar várias capturas, nem mover o objeto para obter as diferentes perspetivas. São objetivos, estudar e analisar os diferentes sistemas e métodos de captura de informação em 3D, compreender o processo de reconstrução de forma a gerar os modelos desses e objetos e por fim a construção do sistema descrito.

­Introdução

Todos os dias nós, humanos, interagimos com objetos sem precisar de prestar especial atenção aos mesmos. A perceção da geometria destes objetos a distância a que estes são é feita de forma tão natural que nem nos apercebemos disso. No entanto, no mundo digital, fazer com que um computador reconheça autonomamente essa informação sobre as cenas não é um assunto trivial.

Quando capturada com uma câmara convencional (RGB), o mundo 3D é mapeado numa imagem 2D e a noção de profundidade (pode) perder-se com alguma facilidade o que pode dificultar no reconhecimento da correta geometria dos objetivos.

???

No caso dos scanners 3D, essa informação já existe o que ajuda à análise da cena e dos objetos. Este tipo de tecnologia está em constante evolução e há cada vez mais sistemas capazes de o fazer e cada vez mais acessíveis: a Kinect é um exemplo disso. No entanto, os sistemas que permitem fazer a aquisição concêntrica da informação 3D de um objeto exigem a utilização de um maior número de câmaras (no caso de um *setup* estático), o que implica um maior custo, ou o posicionamento da câmara ou objeto em diferentes perspetivas, o que implica um maior tempo para realizar a captura.

A captura de informação em 3D pode ter diferentes funções. Por exemplo, no caso de uma aplicação de Realidade Aumentada em que está a ser desenhada numa imagem do mundo real informação adicional, se esta informação for colocada no sítio errado pode fazer com que os resultados dessa aplicação sejam incorretos e como tal podem provocar uma experiência pobre para o utilizador. Desta forma, a existência de informação volumétrica pode ser vantajosa dando uma maior robustez a este tipo de sistemas e minimizando erros de posicionamento ou *tracking*.

Outro uso potencial da informação em 3D de objetos é a modelação dos mesmos para posterior utilização num mundo digital. Seja no âmbito de jogos, cinema de animação ou até mesmo em simulações em áreas tão diversas como indústria, medicina ou militar, modelos de objetos são usados para os mais diferentes fins e como tal, sistemas que possam ajudar na construção dos mesmos são úteis. Outra possível utilização destes modelos é através das impressoras 3D. A ponte entre estas duas áreas, aquisição e reprodução de informação, é aliciante e pode abrir portas a novas possibilidades.

Motivação

(Como foi mostrado) A utilização de informação em 3D é útil e pode ser usada diferentes formas. O caso específico da captura de informação de um determinado objeto a partir de uma perspetiva concêntrica, além de útil, levanta um desafio interessante em relação ao método da sua realização.

A utilização de um *setup* móvel, isto é, um *setup* em que é necessário haver o movimento da câmara ou do objeto para a recolha de toda a informação do mesmo, faz com que o processo seja de certa forma, barato (apenas é necessária uma câmara) mas lento, uma fez que a captura é incremental e não instantânea. Por outro lado, um *setup* estático permite que a captura de informação das várias perspetivas seja feita em simultâneo. Além de a aquisição de informação 3D de um objeto fixo ser feita de forma mais rápida, este *setup* possibilita ainda a aquisição e geração de informação em 3D das várias perspetivas em tempo real podendo neste caso almejar-se a captura de 3D de entidades em movimento. No entanto, este tipo de *setup* é por norma mais dispendioso uma vez que é frequente usar múltiplas câmaras para realizar a captura.

Desta forma e juntando o melhor dos dois *setups*, seria um desafio conceber um sistema capaz de fazer uma aquisição concêntrica em tempo real utilizando apenas uma câmara e um *setup* estático. Isto permitiria a aquisição de mais informação em menos tempo e de forma mais económica. Um sistema deste género poderia permitir a geração de vídeo em 3D real (não apenas com a noção de profundidade estereoscópica) que posteriormente, no prisma de um espectador, possibilitaria a visualização desse vídeo de forma dinâmica em várias perspetivas. A geração de dados com estas características poderá também ser usada mais tarde em sistemas holográficos uma vez que existe informação suficiente para criar uma vista concêntrica do objeto em foco.

Objetivos

Neste contexto, pretende-se com esta dissertação desenvolver um sistema de baixo custo capaz de realizar a captura concêntrica de informação 3D em tempo real de entidades de pequenas/médias dimensões a partir de um *setup* estático e de apenas uma câmara. Inicialmente será necessário compreender o processo de captura de informação 3D e como a partir dessa informação poderemos gerar o objeto em questão.

A captura concêntrica de informação 3D de um objeto requere a aquisição de informação de várias perspetivas. Para o conseguir fazer de forma instantânea (todas as perspetivas ao mesmo tempo) é comum usar-se vários sensores/câmaras, no entanto, além do aumento da carga computacional, isso também envolve um maior custo em *hardware.* Como tal, um dos desafios será conceber a arquitetura de um sistema que consiga obter a informação de todas essas perspetivas em simultâneo e de forma eficiente utilizando apenas um sensor.

Dada a captura e tratamento de informação em 3D envolver um grande esforço computacional, outro desafio será estudar formas de agilizar este processamento.

Em suma, os objetivos para esta dissertação são:

* Compreender os diferentes processos de captura de informação de 3D.
* Estudar e conceber uma arquitetura do sistema capaz de recolher a informação do objeto de todas as perspetivas utilizando apenas uma câmara estática e sem mover o objeto.
* Desenvolver esse sistema de captura e também a componente de visualização de informação.
* Estudar formas de melhorar o desempenho e a qualidade da informação capturada.

Estrutura do Documento

Esta dissertação está dividida em X capítulos. O primeiro capítulo, este, apresenta uma pequena introdução ao tema e ideia por detrás desta tese.

No capítulo 2 será feita um apanhado do estado da arte no campo da aquisição de informação em 3D. Serão abordados diferentes métodos de captura deste tipo de informação, primeiro de forma genérica e depois com mais detalhe os processos capazes de realizar a captura de forma concêntrica. Por fim serão ainda apresentadas diferentes……..

No capítulo 3 o problema será descrito de forma detalhada e será apresentado a arquitetura do sistema desenvolvido. Neste capítulo serão também justificadas todas as decisões tomadas que levaram à construção de tal sistema………

No capítulo 4 será descrito a arquitetura do sistema assim como o processo de desenvolvimento e funcionamento do mesmo. Serão apresentados os problemas surgidos e as soluções tomadas para os resolver………

No capítulo 5 serão apresentados e analisados os resultados…….

Finalmente, no capítulo 6 concluiu-se a conclusão. E trabalho futuro………..

Related Work

Os primeiros sistemas de captura de informação em 3D remontam à década de 1960 onde eram usadas luzes, câmaras e projetores para realizar esta tarefa. Era um processo moroso e que exigia muito esforço e tempo para conseguir ter resultados satisfatórios. Durante vários anos esta tecnologia não sofreu grandes desenvolvimentos e tal pode ser justificado por exemplo pelas limitações de largura de banda ou pela capacidade de armazenamento necessária. No fim dos anos 1980 foram criados os primeiros scanners 3D a laser foram criados

[http://www.tx.ncsu.edu/jtatm/volume5issue4/Articles/Lerch/Lerch\_full\_2 21\_07.pdf](http://www.tx.ncsu.edu/jtatm/volume5issue4/Articles/Lerch/Lerch_full_2%0921_07.pdf)

<http://fer.rb.kau.edu.sa/Pages-drmtfaen1.aspx>

The first 3D scanning technology was created in the 1960s. The early scanners used lights, cameras and projectors to perform this task. Due to limitations of the equipment it often took a lot of time and effort to scan objects accurately. After 1985 they were replaced with scanners that could use white light, lasers and shadowing to capture a given surface.

Many different technologies can be used to build these 3D scanning devices; each technology comes with its own limitations, advantages, and costs. Many limitations in the kind of objects that can be digitized are still present: for example, optical technologies encounter many difficulties with shiny, mirroring, or transparent objects.

A survey of content based 3D shape retrieval methods [pdf]

* The first 3D scanning technology was created in the 1960s. The early scanners used lights, cameras and projectors to perform this task. Due to limitations of the equipment it often took a lot of time and effort to scan objects accurately.   After 1985 they were replaced with scanners that could use white light, lasers and shadowing to capture a given surface.
* A 3D laser scanner works by first projecting laser light onto the object or surface, then detecting the reflected light. Based on where the lights fall in relation to each other, the scanner calculates their positions and creates data points. These points help a computer recreate it visually.
* 3D laser scanning is used in a variety of fields and academic research. It has benefited clothing and product design, the automotive industry and medical science. Laser scanning can also be used to record built and natural structures, especially in places that people may not be able to access due to safety hazards.

[<http://artescan.net/blog/3-d-laser-scanner-history/>]

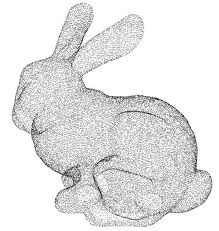
<http://3dscanningservices.net/blog/index.php/what-does-a-3d-scanner-do/>

<http://3dscanningservices.net/blog/index.php/growth-change-and-the-future-of-3d-scanning/>

## The History of Laser scanning

Posted on [September 10, 2013](http://3dscanningservices.net/blog/index.php/the-history-of-laser-scanning/) by [Gareth](http://3dscanningservices.net/blog/index.php/author/gareth/)

Laser scanning has actually been around since the 1960’s, only everyone was more concerned about going to Woodstock. The popularity of 3d scanning didn’t break through beyond the engineering field until the late 1990’s.  Which should have been expected as, laser scanning could never advance correctly until the hard drive storage and bandwidth increased.

**[](http://3dscanningservices.net/blog/wp-content/uploads/2013/09/Laser-Scanned-Rabbit-Piont-Cloud.jpg)**

Laser Scanned Rabbit cloud point example

In today’s times, point clouds are too large to be delivered via e-mail or FTP format, but are delivered instead to external hard drives or thumb drives mainly because the data files are so large. Laser scanning, typically started out in the same way [**GPS**](http://www8.garmin.com/aboutGPS/) (Global Positioning System) was introduced. When GPS was introduced only military personnel, aviators, boaters and surveyors used it. Today, GPS technology is incorporated into millions of various devices throughout the world, including, cars, computers, and mobile phones.

One example of how laser scanning has become more commonplace is the X-box gaming station. If hooked up with the Xbox [**Kinect**](http://www.wired.co.uk/news/archive/2013-03/18/kinect-sdk-update) feature, this has the ability to perform a type of 3d scanning. The various games scan the player in real-time, shape movements, size, and collects data to create a virtual reality.

The process of scanning technology is a very simple and a straight-forward science, 3d points are used to define the tangible real surfaces that are scanned. It doesn’t matter if it’s a car, house, jumbo jet or even a person. The finished result is always high quality and accurate detail.

A cloud point can be anything from 10 to 100 GB’s, meaning most CAD programs aren’t able to handle it. The power behind successful three-dimensional scanning is utilizing the point cloud and the built model. The scan takes the collected data and places it from the point cloud in the user’s selected software system. The benefits are being able to visualize both of the components it ultimately maximizes the ability for clash detection (interference in a project), which allows for faster turn-around time on every new project

[<http://3dscanningservices.net/blog/index.php/the-history-of-laser-scanning/>]

<http://3dscanningservices.net/blog/index.php/growth-change-and-the-future-of-3d-scanning/>

<http://store.makerbot.com/digitizer.html>

Many different technologies can be used to build these 3D scanning devices; each technology comes with its own limitations, advantages, and costs. Many limitations in the kind of objects that can be digitized are still present: for example, optical technologies encounter many difficulties with shiny, mirroring, or transparent objects.

História de captura 3D – gesso (?)

Today, there are three common technologies that can acquire 3D images, each with its own unique strengths and common use cases: stereoscopic vision, structured light pattern and time of flight (TOF)

Explicação de cada uma e quadro comparativo (2012):

<http://www.ti.com/lit/wp/spry199/spry199.pdf>

<http://www.pointclouds.org/news/2013/09/26/occipital-launches-structure/>

3D Scanning / Images Acquisition

Things

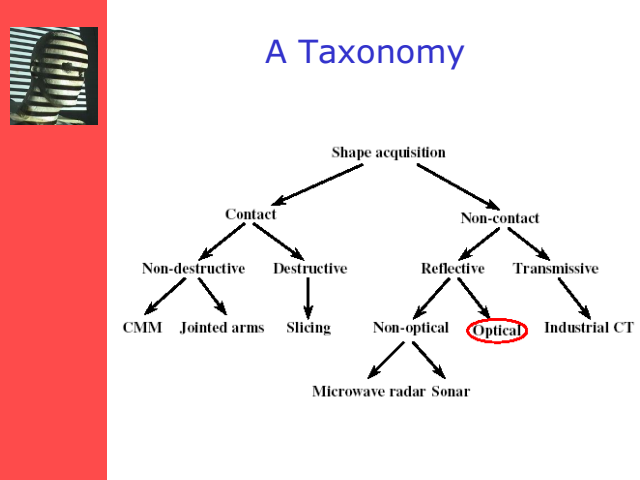


Ilustração 1 - Structured Lighting (Talk - 2012)~

“Consumer RGB-D Cameras and their Applications”

Stereoscopy

Things

Structured Light

Things

Time-Of-Flight

See Book Time-of-Flight Cameras and Microsoft Kinect

Things

Kinect

See Book Time-of-Flight Cameras and Microsoft Kinect

Kinect assigns 11 bits to each depth value returned. 10 bits are used to codify depth values, corresponding to only 1024 levels of depth and the 11th bit is used to signal a non disparity measure or a depth measure error (Khoshelham and Elberink, 2012), represented by NaN (Not-a-Number value).

The difference between two successive levels of depth values is not constant; as it is shown in ﬁgure 7, it follows a quadratic function. In other words, the empty space between layers becomes greater each time points are farther from sensor. These layers form ﬂat slices perpendicular to the Z axe (the optical axe of sensor). The distance between slices begins with few millimeters and it’s increased up to 25cm at 10m. This is the reason why, most of works take into consideration only points with depth values lower than a certain threshold; where the empty space between layers can be accepted. Most of the works, take this limit as 5m, where distance between layers is lower than 10cm; however, in other works, as in (Trevor et al., 2012), this limit was chosen equal to 3.5m; region where the distance between bands is lower than 5cm, the uncertainty of LRF.

Figure 7: Depth step discretization in function of the depth.

<http://www.uv.mx/anmarin/papers/ICINCO13.pdf>

Concentric Data Acquisition Systems

Things

Moving

Things

Static

Things

­­­­­­­Summary

Things

Concept Design (?)

Como foi brevemente descrito na introdução, o objetivo desta tese consiste em criar um sistema de baixo custo capaz de fazer a aquisição concêntrica de informação 3D de um objeto em tempo real a partir de um *setup* estático e utilizando apenas uma câmara.

Depois de analisadas várias formas de fazer a captura de informação em 3D e listadas as vantagens e desvantagens de cada uma, …

Com isto em mente, neste capítulo vamos mostrar os passos tomados que permitiram chegar ao sistema final. Primeiro os objetivos e características do sistema serão alvo de uma análise mais detalhada e de acordo com essa informação serão mostrados alguns dos possíveis casos de uso para o sistema. Depois serão mostradas e justificadas as decisões tomadas a nível de material utilizado para a realização da captura e do *setup* físico escolhido que levaram à construção deste sistema.

Descrição e Funcionalidades (?)

Características como baixo-custo, aquisição concêntrica ou tempo real, que estão na base dos requisitos para a construção desta aplicação, não são por norma compatíveis, no entanto são características necessárias para cumprir os objetivos propostos.

A aquisição concêntrica de um objeto é importante para obter a geometria do mesmo. Esta aquisição pode ser feita de forma faseada, isto é, uma perspetiva de cada vez, ou então capturando todas as perspetivas de uma só vez. Como a captura se quer também em tempo real, isto faz com que o método faseado não possa ser considerado. Desta forma, a realização de captura simultânea traz vantagens, não só a nível do tempo que a captura demora a ser feita mas também traz a capacidade de capturar objetos ou outras entidades que estejam em movimento. A aquisição deste tipo de entidades só assim é possível caso contrário a informação das diferentes perspetivas com tempos também diferentes resultaria em aquisições erradas.

A utilização de apenas uma câmara está estritamente relacionada com a questão de ser querer um sistema de baixo custo. ~~Esta característica é desejável para diferenciação num mercado que se encontra em crescimento~~. Já existem algumas soluções que permitem fazer a aquisição concêntrica de informação 3D, no entanto, estas têm características diferentes das aqui propostas. Soluções como X e Y estão orientadas a modelos com grandes dimensões e apenas à aquisição do modelo estático e funcionam como estúdios. Isto faz com que estas soluções tenham um nível de detalhe muito elevado mas em contrapartida são instalações de grande dimensão e, devido à qualidade e quantidade de material utilizado, são também soluções dispendiosas.

X – <http://ir-ltd.net/>

Y – <http://www.4ddynamics.com/3d-scanners/bodyscanner/>

<http://www.kscan3d.com/> por sua vez já oferece uma solução mais acessível (220€ + câmaras (+-150€ por unidade)) e permite realizar a aquisição completa igualmente estática mas faseada. Os modelos conseguidos também têm menos pormenor que as anteriores devido ás capacidades do hardware usado.

Outras soluções como o *Digitalizer* da *Makerbot* ouY(??), já têm objetivos diferentes e estão orientadas à portabilidade à comercialização para o público geral. Estes *scanners* conseguem fazer a aquisição 3D de objetos estáticos de pequena dimensão e está pensado para a criação de modelos completos e detalhados e para a ponte com impressoras 3D e consequente replicação.

<http://store.makerbot.com/digitizer.html>

<http://www.matterform.net/scanner>

Casos de Uso

Um sistema de captura de informação em 3D consegue de forma mais ou menos rápida adquirir a informação geométrica de objetos ou outras entidades na forma de uma nuvem de pontos. Esta nuvem tem por norma boa qualidade e representa a entidade capturada, o que pode facilmente ser visualizado via *software*. Uma vez criadas mais que uma nuvem de pontos, estas também podem ser unidas de forma a criar uma melhor representação do objeto.

<http://www.cs.princeton.edu/courses/archive/spr03/cs426/lectures/21-3Dscan.pdf>

3D Scanning Applications

•• Computer graphics

•• Product inspection

•• Robot navigation

•• As--built floorplans

•• Product design

•• Archaeology

•• Clothes fitting

•• Art history

O sistema pretendido consegue capturar a informação da nuvem de pontos de várias perspetivas em simultâneo. Esta característica aliada à possibilidade da captura de informação concêntrica em tempo real abre portas a outro tipo de aplicações além da simples captura de modelos.

Realização de vídeo em *3D* verdadeiro – o que é que isso representa para o espetador.

Captura total de modelos para extração de exa-esqueletos em tempo real.

Geração e transmissão de 3D em tempo real.

Holografia

Caso de uso 1

Caso de uso 2

Decisões

De acordo com as características pretendidas e os objetivos traçados, foi necessário tomar decisões em relação ao material para a aquisição de imagem e ao *setup* a usar.

Captura

Vários dos métodos referidos no capítulo anterior foram testados ….

Descartar projetores. Preço, portabilidade, tempo para captura, captura concêntrica faseada – não tempo real.

Utilização de câmaras + estereoscopia ? descartar

…

As câmaras de profundidade têm como uma das vantagens o acesso direto à informação 3D com boa qualidade e de forma rápida. O preço e a portabilidade são também fatores que valorizam este tipo de dispositivos. No entanto traz outras desvantagens como o alcance limitado, tanto a distâncias muito curtas (0-50cm) como longas (> 5m), e a dificuldade na aquisição de informação em condições mais adversas como é o caso da exposição direta de luz solar nas superfícies ou de materiais brilhantes/refletores ou transparentes.

Somando vantagens de desvantagens e tendo em conta os objetivos do projeto, foi decidido usar-se uma câmara de profundidade/RGB-D. Dentro desta gama de câmaras, a foram estudadas as características da Kinect, Xtion PRO Live e as Carmine 1.09 e 1.08 como se pode ver no seguinte quadro:

QUADRO COM CARACTERÍSTICAS DA KINECT E OUTRAS CAMARAS

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Kinect | Xtion PRO LIVE | Carmine 1.08 | Carmine 1.09 |
| Distance min (m) | 0,8\* | 0,8\* | 0.8 | 0.35 |
| Distance max (m) | 3,5\*\* | 3,5\*\* | 3.5 | 1.40 |
| FOV vertical | 43 | 45 | 45 | 45 |
| FOV horizontal | 57 | 58 | 57.5 | 57.5 |
| Depth Size | 640x480 (30FPS) | 640x480 (30FPS) | 640x480 (60FPS) | 640x480 (60FPS) |
| Color Size | 1280x1024 (10FPS) | 1280x1024 (15FPS) | 640x480 (30FPS) | 640x480 (30FPS) |
| 640x480 (30FPS) | 640x480 (30FPS) | - | - |
| Price | 150/250€ | 139£ - 166€ | 200$ - 148€ | 200$ - 148€ |

\*Testes realizados mostram que é possível capturar informação a uma distância de 60cm e detetar características de um objeto com detalhe de 1cm

\*Testes realizados mostram que é possível capturar informação a uma distância de 5m com um erro inferior a 10cm e a uma distância de 10m com um erro inferior a 25cm.

<http://www.asus.com/Multimedia/Xtion_PRO_LIVE/#specifications>

<http://www.primesense.com/get-your-sensor2/> <http://www.primesense.com/wp-content/uploads/2012/12/PrimeSenses_3DsensorsWeb.pdf>

<http://msdn.microsoft.com/en-us/library/jj131033.aspx>

[http://eng.au.dk/fileadmin/DJF/ENG/PDF-filer/Tekniske\_rapporter/Technical\_Report\_ECE-TR-6-samlet.pdf - pag 3](http://eng.au.dk/fileadmin/DJF/ENG/PDF-filer/Tekniske_rapporter/Technical_Report_ECE-TR-6-samlet.pdf%20-%20pag%203)

Apesar das câmaras X, Y e a Kinect serem bastante similares em termos de características e preços, decidiu-se utilizar a Kinect como dispositivo de captura para este projeto devido à sua popularidade em ambientes de desenvolvimento. Entre as duas Kinects apresentadas, a escolha recairia pela Kinect base uma vez que é mais económica. No entanto e para os testes realizados foi usada uma Kinect for Windows cedida pelo CCG[[1]](#footnote-1).

Near Mode para Kinect for Windows. allowing the sensor to see objects as close as 40 centimeters

http://blogs.msdn.com/b/kinectforwindows/archive/2012/01/20/near-mode-what-it-is-and-isn-t.aspx

Kinect 2 é que era bom mas, além de ainda não se saber o preço, também não está disponível. Só em 2014.

Foram ainda realizados testes com a utilização de duas Kinects em simultâneo

Setup (Aquisição Concêntrica)

A aquisição de informação de uma forma concêntrica pressupõe que esta seja adquirida de várias perspetivas. Como foi referido anteriormente, uma das formas de o fazer é utilizando várias fontes de captura o que tem como algumas desvantagens um maior custo em *hardware* e também o aumento da carga computacional. No caso específico da utilização de várias *Kinects*, existe ainda outro problema relacionado com a interferência entre elas. Quando, por exemplo, duas *Kinects* estão apontadas para uma determinada zona existem áreas de sobreposição, nessas regiões pode haver a introdução ruído. Isto deve-se à sobreposição dos padrões de IR que impossibilitam a correta estimativa da profundidade dessas áreas. Na figura X pode ver-se este fenómeno.

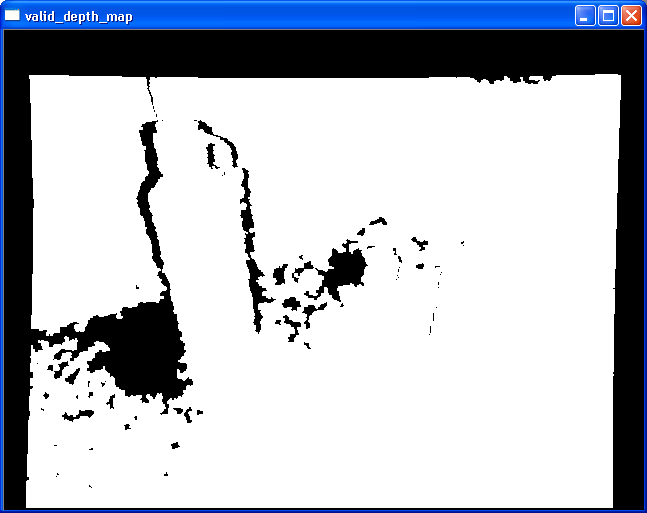
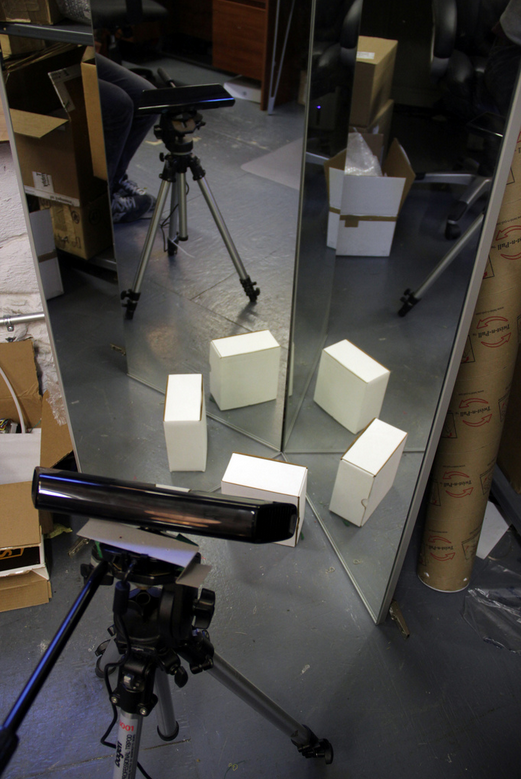
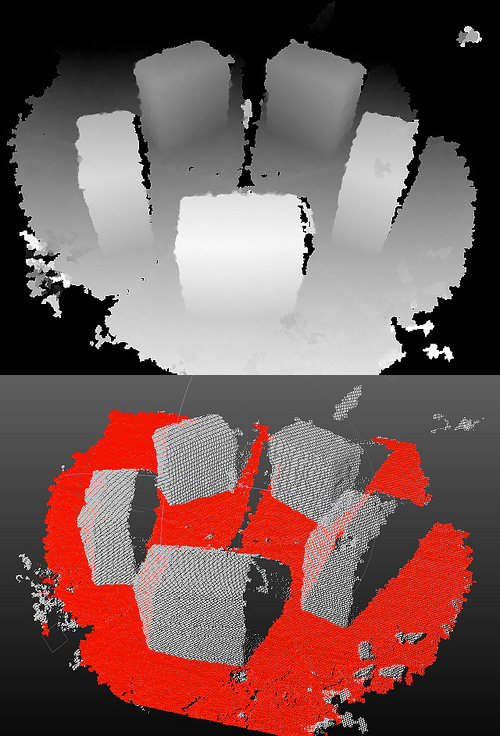


FIGURA X COM RUIDO DE DUAS KINECTS

Dada a característica “baixo custo” proposta, a utilização de mais que uma *Kinect* torna-se inviável. Desta forma foi necessário arquitetar outra forma de conseguir fazer a captura das outras perspetivas (que não a da Kinect) sem perder a propriedade de tempo real. Assim, o método proposto foi a utilização de espelhos que, ocupando a periferia da imagem (não interfere com a visão direta ao objeto), recolhe informação de outras perspetivas do objeto.

A utilização de espelhos para a realização de captura em 3D não é novidade e já foi realizada me X e Y por exemplo. No primeiro caso, foram utilizadas camaras RGB normais (VER E EXPLICAR ARTIGO DOS ESPELHOS)

No livro Making Things See, é mencionada a utilização de espelhos pelo artista e investigador Kyle McDonald como um exemplo da versatilidade da Kinect e para mostrar como é que esta se com os espelhos. No entanto, não foram encontradas evidências de este trabalho ter sido continuado.



<http://www.flickr.com/photos/kylemcdonald/5641883004/>

System Overview

De acordo com as decisões tomadas anteriormente, foi escolhido implementar um sistema que utiliza apenas uma câmara, de profundidade (Kinect) e espelhos.

Desta forma, o sistema construído consiste numa câmara colocada numa posição superior e centrada no objeto e um conjunto de N espelhos numa posição quer permita ocupar as áreas da câmara e ver as restantes partes do objeto. Este *setup* oferece alguma liberdade quanto à colocação dos espelhos e para os testes realizados foram usadas duas disposições diferentes como se pode ver na figura X.

FIGURA COM DOIS SETUPS: ARENA + SAPATOS

O primeiro *setup* tem como estrutura física a Kinect apontada ao objeto numa direção perpendicular ao chão e quatro espelhos a rodear o mesmo objeto. Como se pode observar na imagem proveniente da Kinect, os espelhos ocupam as margens da imagem e contêm informação de uma visão periférica do objeto, neste caso, de quatro perspetivas diferentes, cinco no total. Apesar deste número mais elevado de perspetivas, este *setup* tem como principal limitação a área reservada para a colocação dos objetos devido à distância entre os espelhos e dos espelhos para a câmara.

No segundo *setup,* a Kinect encontrasse centrada no objeto também mas a sua posição relativa ao chão é agora de aproximadamente 35º. Foram usados dois espelhos que se encontram no lado oposto à camara havendo assim mais duas perspetivas diferentes do objeto, três no total. Este *setup* é mais amplo o que permite uma maior liberdade a nível de objetos a serem capturados e a nível de movimentação.

Implementação

Neste capítulo começaremos por mostrar qual a arquitetura do sistema e depois a *pipeline* do mesmo

Arquitetura do Sistema

No final do capítulo anterior foram mostradas imagens do sistema em funcionamento e através dela pode-se perceber a estrutura física do mesmo.

Descrever sistema físico abstrato enquanto câmara + N espelhos. Falar sobre a posição da câmara em relação ao objecto e dos espelhos em relação ao objecto e à camara.

Chamar e definir o primeiro caso (camara perpendicular e 4espelhos) como ARENA para usar nas próximas secções.

System pipeline

Fase0

* Setup chão
* Setup espelhos

Fase1

* Captura “crua”
* Filtrar áreas (chão + espelhos)
* Fazer mirroring da área dos espelhos
* Retirar chão pela distância ao plano

Fase 2

* Remoção de outliers?
* Gerar mesh

Setup

*Setup “Manual” referred to Appendix?*

Para que o sistema funcione é necessário definir o chão e os espelhos. Esta definição passa por delimitar a área de ação de cada um deles e calcular qual o plano que os representa.

A delimitação das áreas é feita da mesma forma para os dois casos. É usada a imagem RGB da câmara e nela são selecionados pontos que delimitarão a área pretendida. No caso do chão isso corresponde á zona de ação, isto é, a área onde os objetos ou entidades deverão estar para que sejam analisados pelo sistema, e no caso dos espelhos corresponde à área do espelho onde aparecerão os mesmos objetos.

IMAGEM DEMONSTRATIVA [Imagem da cena com selecção de pontos + máscara]

A extração do plano, no caso do chão, pode ser feita de duas formas diferentes, uma automática ou a outra, manual. A extração automática do plano deve-se a uma capacidade do OpenNI que permite extrai esse plano segundo alguns pressupostos como a posição da câmara, quantidade de chão visível, etc. Como tal, este tipo de extração nem sempre é possível e como tal tem que se calcular esse plano de forma manual.

A extração manual é feita através da seleção de pontos pertencentes ao plano que ser quer extrair. Depois de selecionados, é calculado o plano que melhor se adapta a esse conjunto de pontos através de um algoritmo de Ransac e este plano será usado como o plano do chão.

IMAGEM DEMONSTRATIVA [Imagem com selecção dos pontos do chão]

No caso dos espelhos esta seleção não pode ser feita diretamente uma vez que os pontos presentes na área do espelho não contêm informação do próprio espelho. Para contornar este fato foi introduzido um artefacto, apenas presente na fase de calibração, que consiste num plano opaco (uma folha de papel ou cartolina) *colado* ao espelho. A extração manual do plano é feita da mesma forma que a do chão mas usando apenas a área ocupada por este artefacto no entanto, o plano extraído é representativo do plano total do espelho.

IMAGEM DEMONSTRATIVA [Imagem com selecção dos pontos de um espelho]

Aquisição da Nuvem de Pontos

A nuvem de pontos corresponde ao conjunto de pontos em 3D que foi capturada pela Kinect. Esta nuvem de pontos pode ser facilmente adquirida com a ajuda do OpenNI no entanto a informação obtida desta forma ainda contém muito ruído que é necessário excluir. Nesta fase de *emagrecimento* da nuvem de pontos é essencial a informação recolhida na fase anterior, a de *setup*.

O primeiro passo pode ser (e é) feito antes do cálculo efetivo dos pontos 3D e consiste em filtrar a informação através das máscaras definidas na fase de *setup,* tanto a do chão como a dos espelhos. É possível realizar este passo uma vez que apenas estas áreas contêm informação útil da cena. Por exemplo, no caso da arena, a informação contida nos cantos da imagem não é útil para a aquisição do objeto e como tal, não há necessidade de ter esse processamento extra.

IMAGEM DE EXEMPLO DAS MASCARAS EM 2D

Na imagem anterior foi ainda aplicado um filtro de distância limitando às áreas dos espelhos de forma a remover ruído desnecessário proveniente das reflecções dos outros espelhos. Depois de aplicado estes filtros os pontos em 3D são então calculados e obtemos a primeira representação tridimensional da cena.

IMAGEM 3D DA CENA

Como se pode observar ainda existe bastante informação desnecessária que terá que ser removida e a informação na zona dos espelhos ainda não é a correta. Esta última incorreção é a primeira a ser corrigida. Os pontos presentes nas áreas dos espelhos são invertidos de acordo com o plano do espelho correspondente colocando-os assim no sítio correto.

IMAGEM 3D DA CENA COM INFORMAÇÃO DOS ESPELHOS JÁ INVERTIDA

Depois da colocação dos pontos 3D nos sítios corretos, o passo seguinte é a remoção do chão. Este passo é simples mas computacionalmente pesado uma vez que é necessário comparar a distância de todos os pontos ao plano do chão e remover aqueles cuja diferença é inferior a um determinado limite. Por defeito esse limite é 1cm.

IMAGEM 3D DA CENA SEM CHAO

Neste ponto apenas se encontra presente no visualizador 3D o objeto que se encontra na zona de ação no entanto ainda é possível encontrar algum ruído na imagem resultante de pontos não removidos do chão ou áreas de espelhos definidas erradamente. Todo este processo é conseguido a framerates interativas.

Mesh Generation

Smoothing (where?)

Data Visualization / Recording

Visualizador usado é o do PCL

A gravação é feita a partir do OpenNI (ficheiros .oni) com informação *crua* da *Kinect*. Será necessário criar formato de dados para guardar o resultado 3D como “filme” ? Seria necessário muito espaço..

Problemas e Soluções

Ruído

Distance filtering e Floor removal como já foi descrito na pipeline do sistema – aquisição da nuvem de pontos. No entanto ainda existe ruído na núvem e este tem 3 formas: pontos a mais, buracos na informação e informação “desnivelada” (os pontos precisam de ser alisados)

Hole filling

MeshLab para modelos fixos

“Multiple Kinect Studies - Technical Report”

“DEPTH CAMERA IMAGE PROCESSING AND APPLICATIONS” - In this paper, we introduce various systematic and non-systematic depth errors and state of the art enhancement methods

“Incremental 3D Model Generation using Depth Camera” - we propose a method of retaining knowledge of surfaces from depth camera images acquired over time

Tecnologia

Used Technologies: OpenNI, OpenCV, (C++) Boost, PCL, Qt

<http://www.openni.org/>

<http://opencv.org/>

<http://www.boost.org/>

<http://pointclouds.org/>

[http://qt-project.org](http://qt-project.org/)

Results

Object Acquisition

Show examples of Point Clouds and Meshes.

Continuous acquisition for missing data

Hole filtering technics/3D processing

Real-time Acquisition

Show examples of Point Clouds videos and Fast Mesh Generation

Quality analysis

Performance Analysis.

Conclusion and Future Work

Conclusions

Future Work

Kinect 2 – mais resolução (profundidade e imagem)

Use of Leap Motion – Show the advantages of Leap and a possible integration with a RGB camera to complement depth and rgb image.

Appendix

User Guide

Related Work (OLD)

3D Scanning / Images Aquisition

Things

Technology (Instruments)

Describe each technology and point the pros and cons.

Structured Light

Stratified light?

Depth Cameras (emphasis on Kinect)

Ranging Cameras

Triangulation Scanners

**Mesa Imaging SwissRanger 4000 (SR4000)**

<http://www.acroname.com/robotics/parts/R317-SR4000-CW.html>

**PMD Technologies CamCube 2.0**

http://www.geometh.ethz.ch/people/kohtobia/DGPF2011

<http://openni-discussions.979934.n3.nabble.com/OpenNI-dev-Minimum-Depth-td4015339.html>

**PMD Nano**

http://www.pmdtec.com/html/pdf/order\_CamBoard\_nano.pdf

**Bumblebee 2 e XB3 specs**

http://uprt.vscht.cz/kubicekm/Novinky%20ze%20sv%C4%9Bta/Stereo\_Vision\_Introduction\_and\_Applications.pdf

**Bumblebee 2**

http://3dvision-blog.com/forum/viewtopic.php?f=23&t=2655

**Bumblebee XB3**

http://www.ece.gatech.edu/academic/courses/ece4007/11fall/ECE4007GTS/sv1/documents/ASEDProposal.pdf

**Kinect**

**Primesence**

**Other Cameras**

<http://en.wikipedia.org/wiki/Time-of-flight_camera#Brands>

<http://dinast.com/ipa-1110-cyclopes-ii/>

http://www.3d3solutions.com/products/3d-scanner/

Systems

Show some examples of working systems. Point the technologies that they use, their characteristics and for what purposes where they developed and explain how they are used (static or moving camera, p.e.)

From Capture to Models

Mesh construction from point clouds.

Smoothing

Color issue.

Normal, reflection, etc.

­­­­­­­Summary

Things

1. Centro de Computação Gráfica (www.ccg.pt) [↑](#footnote-ref-1)