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 360º em tempo real utilizando apenas uma câmara e um *setup* estático. Isto permitiria a aquisição de mais informação em menos tempo e de uma forma 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 360º 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 360º 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………..

3D Scan – Related Work

Os primeiros sistemas de captura de informação em 3D remontam à década de 1960 e 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 que usavam luz branca, lasers e sombras para capturar a superfície de objeto.

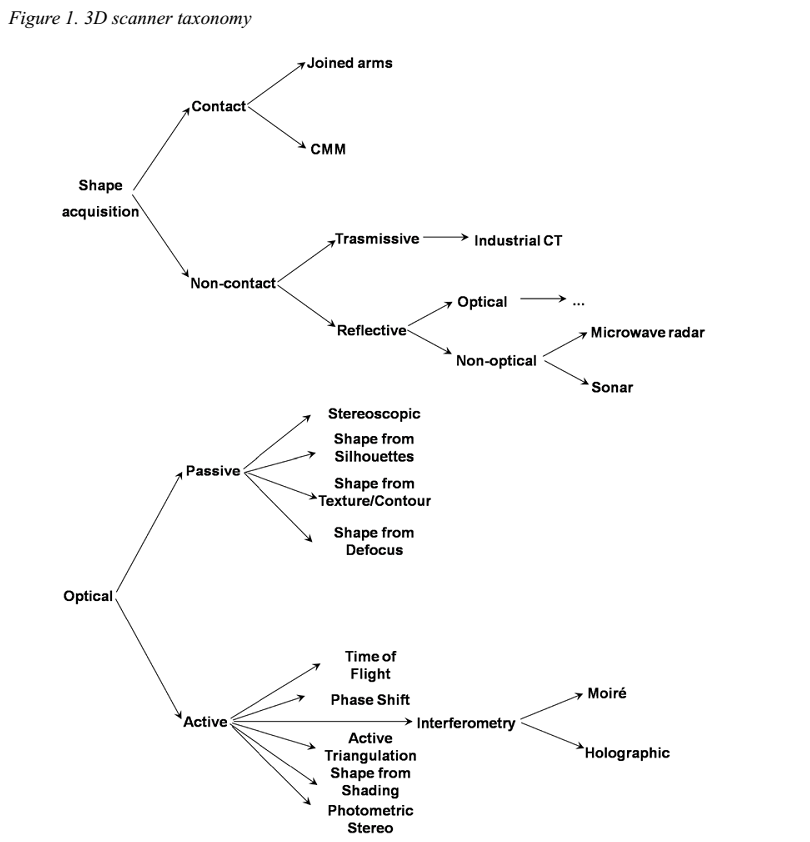
Desde então a tecnologia tem evoluído a passos largos e têm surgido vários sistemas utilizando técnicas diferentes para o mesmo fim: fazer o scan 3D de informação. Apareceram vários sistemas com características distintas e como tal, tinham propósitos diferentes como a captura a longa ou curta distância ou a aquisição de uma qualidade perfeccionista ou a preferência pela prototipagem rápida. O aperfeiçoamento e difusão destes instrumentos serviu também de alavanca para algumas áreas como a Antropometria ou a preservação digital.

Atualmente já existem vários dispositivos capazes de fazer aquisição 3d de forma fácil e rápida. Além dos sensores industriais orientados a capturas de grandes dimensões, atualmente também já existem sistemas que permitem fazer esse tipo de aquisições em casa. Produtos como o matterform[[1]](#footnote-1) e o digitalizer[[2]](#footnote-2) permitem capturar o modelo 3D com grande qualidade (erros nas ordens dos milímetros) de um objeto de pequenas dimensões em poucos minutos. A Kinect é outro exemplo que prima pela sua versatilidade e consegue tanto fazer a captura de objetos como de realizar a captura e criar o modelo de um compartimento completo. A vertente móvel deste tipo de sensores tem-se tornado cada vez mais apelativa. Soluções como o Capri da primesense[[3]](#footnote-3) ou projetos inovadores como o Structure Sensor[[4]](#footnote-4) ou o CADScan[[5]](#footnote-5) poderão levar à proliferação destes sensores em ambiente de mobilidade o que poderá levar a avanços significativos no campo da Realidade Aumentada.

Neste capítulo serão descritos diferentes métodos de captura de informação 3D, focando a análise nos processos óticos de Estereoscopia, Luz Estruturada e Time-of-Flight dado serem os que mais se enquadram com a ideia proposta. Depois dos métodos serão mostrados exemplos de aplicações que conseguem realizar a aquisição 360º e por fim serão listadas algumas das possíveis aplicações práticas destes sistemas de captura 3D.

Métodos de Captura

Existem várias tecnologias que podem ser usadas para fazer a captura de informação 3D e não existe uma que seja melhor que as outras. Todas elas têm vantagens e desvantagens de acordo com os objetivos pretendidos e um custo associado varia consoante as características dos sistemas. Se o objetivo for reconstrução de artefactos arqueológicos, a captura não poderá ser evasiva e terá que ter um elevado nível de detalhe, no entanto, o tempo de aquisição não será uma limitação. Por outro lado, se se quiser fazer videoconferência, o tempo de aquisição terá que ser mínimo enquanto a qualidade dos modelos passa a ser secundária e pode ser até aproximada com modelos já conhecidos.

Os diferentes requisitos das aplicações fizeram com que fossem criados diferentes tipos de sistemas de captura. O esquema presente na imagem XX representa uma possível taxonomia dos sensores de aquisição 3D.

A primeira grande divisão prende-se com o facto de a captura ser feita através de contacto ou não. Por norma, técnicas de captura por contacto são intrusivas e como tal podem chegar a alterar as características físicas do objeto. As técnicas mais usadas neste campo são CMM (Coordinate Measuring Machine) e Jointed Arm. Do outro lado, as técnicas que não utilizam o contacto funcionam a partir da interação entre a superfície do objeto e algum tipo de radiação. Contudo, no caso da radiação do tipo transmissiva, esta também é intrusiva uma vez que é absorvida pelo objeto e como tal pode alterar as suas propriedades. Exemplos deste tipo de sistemas são a Tomografia Computadorizada, Ressonância Magnética ou os Ultrassons.

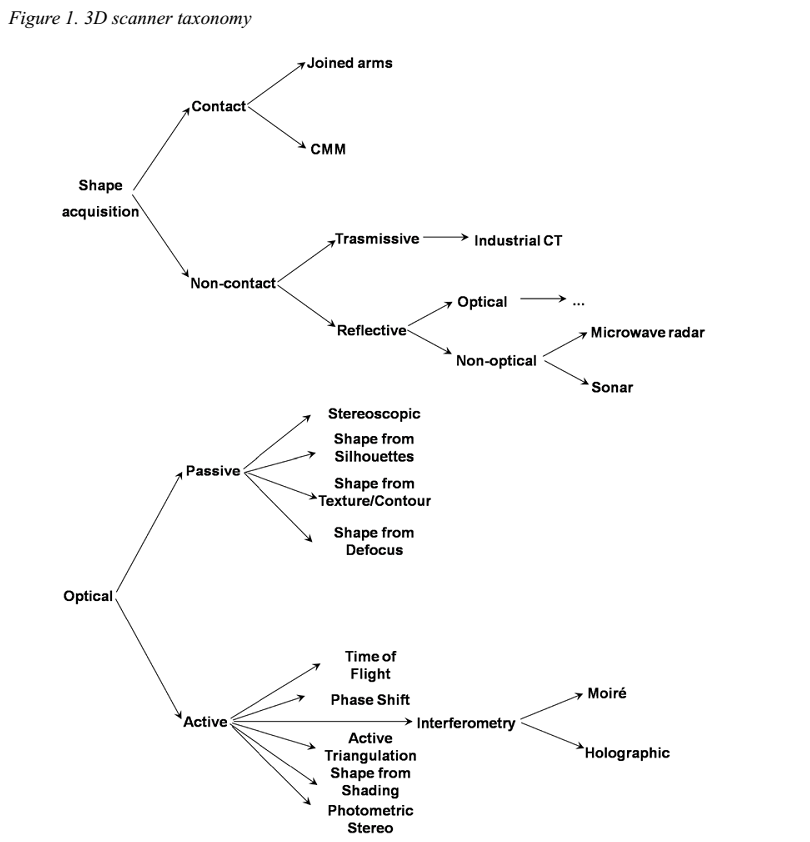
As tecnologias reflexivas, como o nome indica, exploram a radiação refletida pelos objetos para inferir a posição dos pontos. Dependendo do comprimento de onda da radiação utilizada, estas podem ainda ser óticas quando estas se encontram no espectro visível (100nm – 300um) ou não-óticas quando são utlizadas ondas sonoras ou radiação eletromagnética.

O caso das tecnologias óticas requere um maior detalhe uma vez que é o que apresenta mais oferta. Isto deve-se ao facto de serem tecnologias baratas e permitem realizar o scan da informação de forma rápida e em grande escala.

A divisão feita nas tecnologias óticas prende-se nestas serem passivas ou ativas.

Uma das desvantagens das tecnologias óticas é por exemplo a dificuldade em capturar a informação de superfícies brilhantes, reflexivas ou transparentes.

Must be simultaneously illuminated and imaged (occlusion problems)

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> [TEXAS]

Book: “Depth Map and 3D Imaging Applications”

Laser Scanning ?

Stereoscopy

The most common 3D acquisition system is the stereoscopic vision system, which uses two cameras to obtain a left and right stereo image. These images are slightly offset on the same order as the human eyes are. As the computer compares the two images, it develops a disparity image that relates the displacement of objects in the images. Commonly used in 3D movies, stereoscopic vision systems enable exciting and low-cost entertainment. It is ideal for 3D movies and mobile devices, including smartphones and tablets. [TEXAS]

Structured Light

Structured light illuminates patterns to measure or scan 3D objects. Light patterns are created using either a projection of laser or LED light interference or a series of projected images. By replacing one of sensors of a stereoscopic vision system with a light source, structured-light-based technology basically exploits the same triangulation as a stereoscopic system does to acquire the 3D coordinates of the object. Single 2D camera systems with an IR- or RGB-based sensor can be used to measure the displacement of any single stripe of visible or IR light, and then the coordinates can be obtained through software analysis. These coordinates can then be used to create a digital 3D image of the shape. [TEXAS]

Structured-light 3D surface imaging: a tutorial [pdf]

Time-Of-Flight

Relatively new among depth information systems, time of flight (TOF) sensors are a type of light detection and ranging (LIDAR) system that transmit a light pulse from an emitter to an object. A receiver determines the distance of the measured object by calculating the travel time of the light pulse from the emitter to the object and back to the receiver in a pixel format.

TOF systems are not scanners, as they do not measure point to point. Instead, TOF systems perceive the entire scene simultaneously to determine the 3D range image. With the measured coordinates of an object, a 3D image can be generated and used in systems such as device control in areas like manufacturing, robotics, medical technologies and digital photography. TOF systems require a significant amount of processing, and embedded systems have only recently provided the amount of processing performance and bandwidth needed these systems. [TEXAS]

See Book Time-of-Flight Cameras and Microsoft Kinect

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>

Sistemas de Aquisição Concêntrica/360

Explicar o que é aquisição concêntrica.

Pode também ser chamada de aquisição 360º? Aquisição de todas as perspetivas visíveis?

A aquisição de informação de uma forma concêntrica pressupõe que esta seja adquirida de várias perspetivas.

Pode ser apenas dois pontos de vista o mais para mais detalhe e evitar oclusões

IMAGEM

Long range:

<http://www.3dscanco.com/products/3d-scanners/3d-laser-scanners/faro/focus.cfm>

Aquisição Estática

Estática tanto a câmara como a entidade a digitalizar (não é necessário movê-la para capturar toda a informação.

O que é a aquisição estática e o que implica para a aquisição concêntrica

Body scanning!

<http://ir-ltd.net/>

IR-Entertainment



360° Gigapixel Full Body Scanning, Anel com 115 camaras Canon DSLR 18mp. Usa Photogrammetry e structured light.

360° Gigapixel Face Scanning, Anel com 52 camaras Canon DSLR 18mp

Serviços: Virtual Character Creation, Virtual Tailored Hair, Digital Hair Styles, Basic Character Rigging, Virtual Reality, Collective Intelligence, Interactive Experiences

Captura instantânea permite poses estáticas ou em movimento mas não permite captura contínua.

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



near simultaneous 360 degree scanning - for example body scanning for medical or Film entertainment industries

Two to eight scanners are supported in the system.

The Mephisto Body scanner will perform each scan at a speed of 0,24 sec, totalling a full   at 1-2 sec's depending on the settings.

For a body scanning application short-throw projectors could be used, therefore reducing the space needed to rig the system.

<http://www.space-vision.jp/EP-Body_Scanner.html>



The Cartesia Series Portable 3D Body Scanner is the world's first portable body scanner.

nine cameras mounted onto three separate towers.

The scanner photographs the subject image within 2.0 seconds

Measurements can be made up to 1 million points or more so that the average margin of error is 3 mm or less.

the light source utilizes a laser

<http://tc2.com/index_3dbodyscan.html>

16Kinects (4angulos, 4 alturas)

Health/medical and fitness management; Body shape analysis; Sizing surveys; 3D product development for apparel and equipment; Made-to-measure clothing; Clothing size and style recommendation; Virtual Fashion

Precisão 3mm

Duração 7seg

<http://www.tc2.com/pdf/kx16.pdf>

 price of $10,000,

Asus XTion 8 <http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XXXIX-B3/531/2012/isprsarchives-XXXIX-B3-531-2012.pdf>

3D Face Reconstruction from a Single Camera Using a Multi-Mirror Set-up [pdf]

Moving

Modelos Pequenos:

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

Scan volume: Up to 20.3 (diameter) x 20.3 cm (H)

Dimensional accuracy: ± 2.0 mm

Detail resolution; 0.5 mm [0.0197 in]

Scan speed: Approx. 12 minutes

Two eye-safe Class 1 laser line generators, 1.3 megapixel CMOS image sensor

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



Maximum object size and weight

Height: 25 cm (9.8 in); Diameter: 18 cm (7.0 in); Weight: 3.0kg (6.6 lbs)

Scan Speed:

High Speed: 5 minutes; High Quality: > 10 minutes

Scan accuracy:

Capture details as small as 0.43mm; Capture size within ±0.25mm

HD CMOS sensor and 2 lasers

<http://www.indiegogo.com/projects/rubicon-3d-scanner/>

Handheld:

<http://www.goscan3d.com/pt/faq>



Priced at US$25,000, the Go!SCAN 3D scanner isn't cheap

O scanner projeta um padrão codificado de luz a partir de uma fonte de luz branca (LED).

O Go!SCAN 3D tem uma resolução de 0,5 mm. oferece uma precisão de até 0,1 mm (0,004 pol) ou de 0,3 mm por 1 m (0,012 por 3 pés)

Stand-Off Distance 400mm

<http://www.3dscanco.com/products/3d-scanners/artec-scanners/artec/artec1.cfm>

uses a regular flash bulb instead of a laser

high resolution (up to 0.5 mm) and high accuracy (up to 0.1 mm).

Working distance 0.4 – 1 m

Kinect:

Dois exemplos: qualidade e limitações são as da Kinect.

<http://www.kscan3d.com/>

Objects that you wish to scan should be between 40 centimeters to 4 meters (16 inches to 13 feet) away from the sensor. Note that scan data quality is reduced as the distance from the sensor increases.

<http://reconstructme.net/blog/>

Multiple sensor handling: Have a look at our body-scanning example that uses 2 sensors that work together and is written in less than 100 lines of code including calibration.

Aplicações

A capacidade de capturar qualquer objeto físico e ter um modelo digital do mesmo em pouco tempo é útil e pode ser usado para diferentes funções como prototipagem rápida de ideias/conceitos, modelação de objetos, controlo de qualidade, preservação histórica, etc.

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

<http://en.wikipedia.org/wiki/3D_scanner#Applications>

<http://www.slideshare.net/dlanman/build-your-own-3d-scanner-introduction>

“Who can benefict from 3D Scannnig”

<http://www.3dscanco.com/about/3d-scanning/benefits.cfm>

Entretenimento e Consumo

* Import scultures into a 3D modeling/rendering pipeline
* Capture Geometric (and photometric) properties for relighting
* Fit clothes, track 3D interaction, free-viewpoint vídeo (3D TV), etc.

Medicina

* Medical imaging (x-ray, CT, MRI, etc) and surgical planning
* Measuring dimentions (dental impressions and hip replacement surgery)
* Tele-Surgery (AR, vídeo see-thorugh)

3-D images are being used across the medical field to replicate organic matter into perfectly fitted prosthetics

3D imaging for bite mark analysis [pdf]

Preservação Histórica

* Preserve/restore deteriorating Works and unite dispersed collections
* Facilitate academic study (tooling, lighting, pentimenti, revision history^)
* Replicate collections (souvenirs, retain repatrieted Works, etc.)

One common example of 3-D scanner use involves art preservation. For instance, a museum can keep a scanned copy of artifacts and paintings to recall later for repair or replication.

Indústria

* Manufacturing and process control (tolerances and alignment)
* Reverse engineering (repairing antiques and replicating designs)
* Remote inspection (inacessible or dangerous environment)

Determine whether manufactured parts are within tolerances

In manufacturing, a 3-D image of a machine may be recalled to hone in on a particular part that may be broken for an easier fix.

Robótica

* Motion planning (manipulation, gripping, pushing/pulling, etc)
* Simultaneous localization and mapping (SLAM)
* Autonomous navigation (DARPA Grand/Urban Challenge)

<http://www.crn.com/news/components-peripherals/240156453/3-d-scanning-technology-with-unlimited-application.htm>

Human limbs, military vehicles and giant murals painted by Leonardo da Vinci -- these are just a few applications of 3-D scanning technology that are already being used in industries from medical to manufacturing to art.

One common example of 3-D scanner use involves art preservation. For instance, a museum can keep a scanned copy of artifacts and paintings to recall later for repair or replication. In manufacturing, a 3-D image of a machine may be recalled to hone in on a particular part that may be broken for an easier fix. 3-D images are being used across the medical field to replicate organic matter into perfectly fitted prosthetics

­­­­­­­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>

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.

Gesture recognition: Enabling natural interactions with electronics: (introduction)

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

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[[6]](#footnote-6).

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 (descrito no setup?)

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.

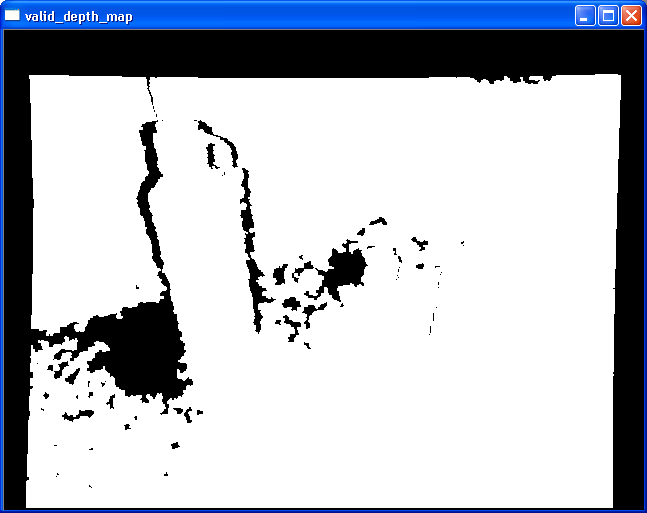


FIGURA X COM RUIDO DE DUAS KINECTS

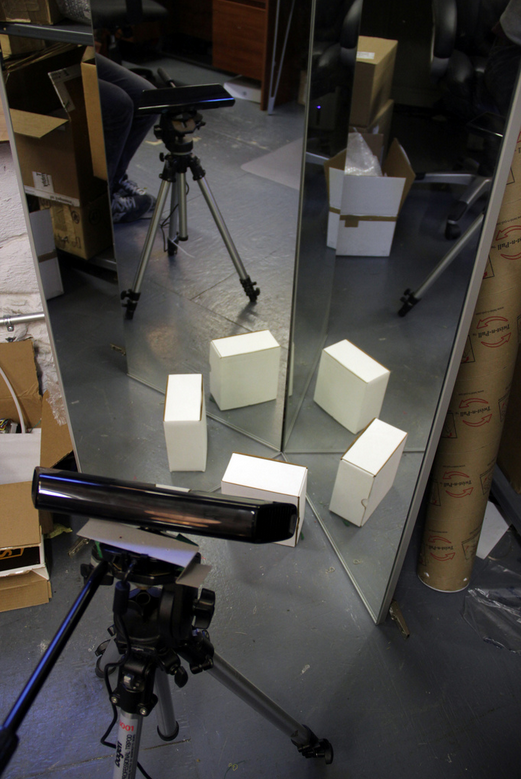
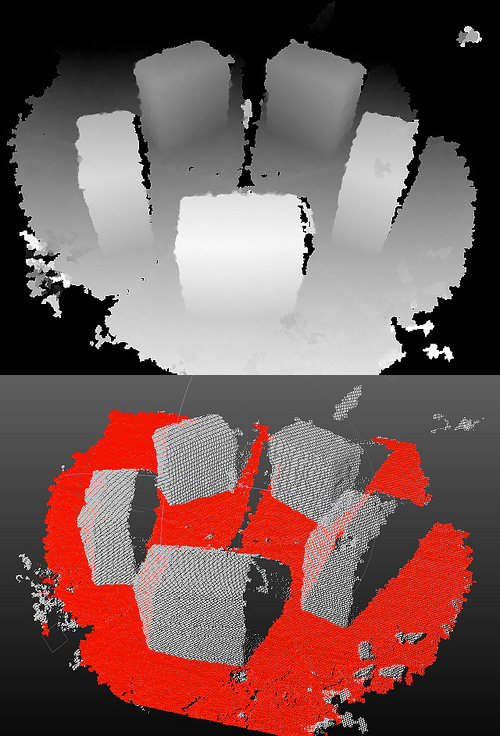
Verificou-se ainda que existe mais ruído quando as câmaras se encontram frente a frente.(?)

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)

Surround Structured Lighting for Full Object Scanning [pdf]

No livro Making Things See [referência], 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/>

Mirror Equations

Planar mirror reflection has been well understood since the early days of computer graphics. A 3D point (center of projection of a camera/projector) or direction (view/projection directions) reflected in a planar mirror is computed as

Equation

where *d* is the distance vector of the mirror supporting plane to the origin, and Am is the alignment (rotations) of the mirror local coordinate system so it lies in the XY - plane in the world coordinate system.

Exploiting Mirrors in Interactive Reconstruction with Structured Light [pdf]

Especificações

Decisão

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.

Kinect

Informação detalhada da Kinect (?)

Espelhos

Teoria sobre espelhos. Como funcionam, quais os erros que podem conter ().

Equação de reflecção. (?)

Descartar sobreposição da reflecção entre espelhos através de filtros na distância (depth map).

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.

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

Implementação

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

Arquitetura do Sistema

Dividir em física e lógica? Ou a lógica é o System Pipeline?

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.

Foram testadas duas abordagens diferentes á arquitetura do sistema

Top view + 4 espelhos – ARENA

O primeiro teste realizado envolveu a Kinect e quatro espelhos. A Kinect foi colocada por cima da área de ação, apontada para o chão de forma perpendicular ao mesmo a uma altura de XX cm. Desta forma, a imagem capturada pela câmara corresponde a uma área de XXcm por YYcm no entanto, com a colocação dos espelhos esta área de ação diminuiu para XXcm por YYcm. Como se consegue perceber, esta área está diretamente relacionada com a altura a que é colocada a câmara e como tal pode ser adaptada caso seja necessário cobrir uma superfície maior.

Este tipo de *setup* ficou com o nome de **arena** uma vez que a sua disposição se assemelha a essa construção.

A vantagem deste *setup* é a geração de cinco perspetivas diferentes em simultâneo. Como se pode ver na seguinte imagem, se o objeto a capturar fosse um cubo (fig X a) ) é possível ter capturas frontais de faces (apenas a que está no chão não poderá ser visualizada).

Perspetiva + 2 espelhos

O segundo teste realizado utilizou a Kinect e dois espelhos. A Kinect foi colocada a uma altura de aproximadamente XXcm apontada para o centro da área de ação. Os limites desta área são ditados pela distância a que os espelhos são colocados. Neste caso, a área reservada ao chão tem a forma de um pentágono com as dimensões ilustradas na figura seguinte.

Uma vantagem deste setup é a maior versatilidade do espaço disponível para captura.

System pipeline

O fluxo de execução (?) do sistema segue

Fluxo “normal” neste tipo de aplicações: aquisição -> processamento -> visualização

Fase0

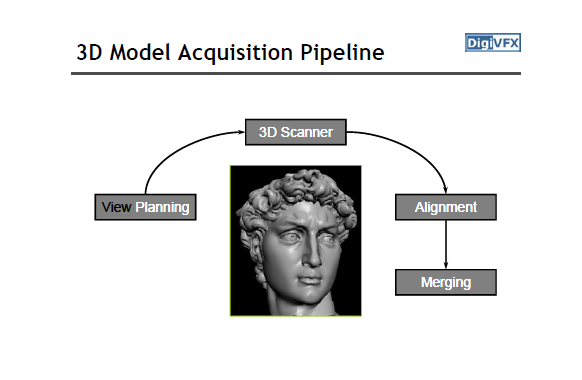
* Setup chão
* Setup espelhos

Fase1

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

Fase 2

* Remoção de outliers?
* Criação de 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.

Geração da Mesh

Algoritmo greedy PCL ? Ver artigo “On Fast Surface Reconstruction Methods for Large and Noisy Datasets (2009)”

Mais algoritmos?

Reconhecimento / matching de objetos?

Data Visualization / Recording´

Surface Reconstruction of Point Clouds Captured with Kinect [pdf]

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..

Modelos / Nuvens de pontos pontuais para ply ou obj?

Problemas e Soluções

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)

“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

Time Critical Isosurface Reﬁnement And Smoothing

Ruído

Pontos a mais que não conseguiram ser filtrados.

Tempo real

* Análise 2D mais extensiva para remover blobs pequenos?
* Remoção de outliers

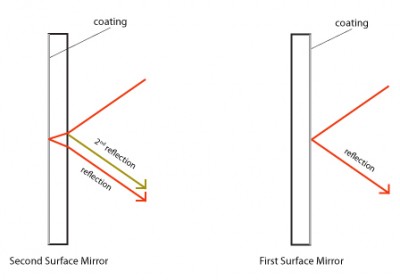
*MeshLab* pode ser usado para modelos fixos e o “lixo” pode ser retirado manualmente.

Buracos de Informação

Causas: tipo de material, interferências entre perspetivas “cruzadas”

SSD: Smoothing Signed Distance Surface Reconstruction. F. Calakli, G. Taubin, CGForum, 2011. – “Particulary good at extrapolating missing data” http://mesh.brown.edu/ssd

Informação Imprecisa



<http://forum.david-3d.com/viewtopic.php?p=9285>a

The setup above does not work well because I used regular mirrors that i had. To get it to work you need to get optical mirrors or "First Surface Mirrors" because of the 2nd reflection there is a ghosting effect

Não só porque a imagem direta da camara também contém imprecisões -> discretização da informação de profundidade -> Kinect usada para distâncias mais longas. Objetos a curtas distâncias nota-se mais as imprecisões da captura.

A resolução diminui com o aumento da distância e o trajeto feito através dos espelhos aumenta essa distância.

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. <http://www.matterform.net/scanner> [↑](#footnote-ref-1)
2. <http://store.makerbot.com/digitizer.html> [↑](#footnote-ref-2)
3. <http://www.primesense.com/solutions/3d-sensor/> [↑](#footnote-ref-3)
4. [http://structure.io](http://structure.io/) [↑](#footnote-ref-4)
5. <http://cad-scan.co.uk/> [↑](#footnote-ref-5)
6. Centro de Computação Gráfica (www.ccg.pt) [↑](#footnote-ref-6)