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

360º Fast 3D 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 dificultando o reconhecimento da correta geometria dos objetivos (FiguraXX).

No caso dos scanners 3D, essa informação já existe o que ajuda à análise da cena e dos objetos nela presente. 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. Este tipo de sensores normalmente é direcional, isto é, apenas conseguem adquirir aquilo que é visível ao sensor e não conseguem tratar das partes ocludidas e, como tal, não conseguem adquirir toda a geometria do objeto. Contrariamente a este tipo de captura, entende-se por aquisição 360 a captura de informação geométrica sobre a superfície de um objeto ao longo de todo o seu perímetro exterior, isto é, conforme vista de qualquer ponto de um círculo que contenha o objeto no seu interior. Neste tipo de capturas, a informação tem que ser adquirida de várias perspetivas de forma a gerar dados que permitam posteriormente uma visualização livre de perspetivas do modelo capturado. No entanto, os sistemas que permitem fazer este tipo de capturas 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 (no caso de um *setup* móvel), o que implica um maior tempo para a realização da 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, os resultados dessa aplicação passam a ser 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 com diferentes fins e como tal, sistemas que possam ajudar na construção dos mesmos são desejáveis. Outra possível utilização destes modelos é a replicação 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 de diferentes formas. O caso específico da captura 360º da informação de um determinado objeto, além de útil levanta um desafio interessante relativamente 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, pois 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 à captura de 3D de entidades em movimento. No entanto, este tipo de *setup* é por norma mais dispendioso uma vez que é frequente utilizar múltiplas câmaras para realizar a captura.

Desta forma e juntando o melhor dos dois *setups*, o desafio será 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 permitiria 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 a informação necessária para criar uma vista livre do objeto em foco.

Objetivos

Desta forma, pretende-se com esta dissertação desenvolver um sistema de baixo custo capaz de realizar a captura 360º de informação 3D em tempo real a partir de um *setup* estático e com 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. Como foi referido, 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.* Desta forma, 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………..

Estado da Arte

Os primeiros sistemas de captura de informação em 3D remontam à década de 1960 e estes usavam luzes, câmaras e projetores para realizar a tarefa. Era um processo moroso que exigia muito esforço e tempo para conseguir ter resultados satisfatórios. Durante vários anos esta tecnologia não sofreu grandes desenvolvimentos e tal pode também ser justificado, por exemplo, pelas limitações de largura de banda ou pela capacidade de armazenamento disponível. 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, com propósitos diferentes como a captura a longa ou curta distância, a aquisição de uma qualidade detalhada ou a preferência pela prototipagem rápida. O aperfeiçoamento e difusão destes instrumentos serviu também como 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, 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 modelos 3D de objetos de pequenas dimensões com grande qualidade (erros nas ordens dos milímetros) e em poucos minutos. A Kinect é outro exemplo que prima pela sua versatilidade e tanto consegue capturar a geometria dos objetos como também realizar a captura e criar o modelo de um cenário 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 por sua vez impulsionará avanços no campo da Realidade Aumentada.

Neste capítulo serão descritos diferentes métodos de captura de informação 3D, focando a análise nos processos óticos de Estereoscopia, Time-of-Flight e Luz Estruturadauma vez que são os mais comuns e os que mais se enquadram com a ideia proposta. Será também abordada a Kinect enquanto marco da captura de informação tridimensional. Depois disso serão mostrados alguns exemplos de aplicações que conseguem realizar o scan 360º e por fim serão listadas algumas das possíveis aplicações práticas dos scanners 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 têm ainda um custo associado que 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 nível de detalhe elevado. No entanto, o tempo de aquisição não será uma limitação. Por outro lado, se se tiver um sistema de videoconferência ou de interação em tempo real, o tempo de aquisição terá que ser mínimo enquanto a qualidade dos modelos passa a ser secundária e pode ser até aproximada a 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 taxonomia dos sensores de aquisição 3D.

A primeira grande divisão prende-se com o facto de a captura ser feita ou não através de contacto. Por norma, técnicas de captura por contacto são intrusivas e como tal podem 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 e a Ressonância Magnética.

As tecnologias reflexivas, tal como o nome indica, exploram a radiação refletida pelos objetos para inferir a posição dos pontos. Estes sistemas não são intrusivos e, dependendo do comprimento de onda utilizado pela radiação podem divididos em sistemas óticos ou não-óticos. Os sistemas não-óticos utilizam radiação eletromagnética, ondas sonoras ou ultrassons para realizar a captura e baseiam-se no princípio de time-of-flight para efetuar as medições.

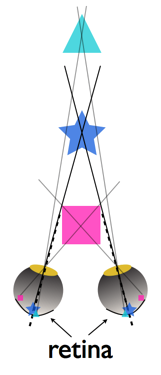
O caso das tecnologias óticas requerem um maior detalhe uma vez que são as mais comuns e as que mais se adaptam ao sistema proposto para esta dissertação. Usam radiação no espectro visível (100nm – 300um) e por norma tecnologias de baixo-custo que permitem realizar a captura de informação com boa qualidade, de forma rápida e em grande escala. No entanto, um problema deste tipo de sistemas é, por exemplo, a dificuldade em capturar a informação de superfícies brilhantes, reflexivas ou transparentes ou ainda as oclusões causadas pela própria geometria do objeto ou por diferentes objetos.

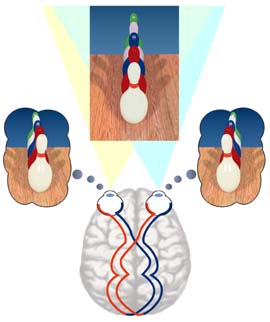
Dentro das tecnologias óticas, estas podem ser divididas em passivas ou ativas. Os sistemas passivos não emitem qualquer tipo de radiação e usam apenas aquela que é refletida pelo ambiente. Estas técnicas utilizam *setups* óticos e/ou perspetivas diferentes que depois são processadas por computador para gerar as coordenadas dos pontos. Estes são por norma sistemas baratos mas, em contrapartida, algo limitados e não produzem resultados com grande detalhe. Por outro lado, nos sistemas óticos ativos, já existe a emissão de radiação. A informação 3D é obtida a partir do processamento das características da radiação emitida e da informação capturada, resultado da interação entre essa radiação e a superfície do objeto. Estes sistemas são os mais comuns sendo que uma das razões é a facilidade em capturar não só a informação 3D mas também as cores dos objetos em simultâneo.

As divisões apresentadas até agora representam caraterísticas das tecnologias existentes no entanto há sistemas que usam mais que uma tecnologia. Estes sistemas híbridos são por norma mais robustos e apresentam resultados mais precisos à custa de uma maior complexidade e de um preço mais elevado.

De seguida serão abordados em mais detalhe os sistemas de captura de Estereoscopia, Time-of-Flight e Luz Estruturada uma vez que são os sistemas mais usados atualmente e os que mais se adaptam ao desafio presente.

Estereoscopia

De acordo com a taxonomia apresentada anteriormente, a estereoscopia é um método de captura de informação 3D que não utiliza contacto, é refletivo, ótico e passivo. Esta técnica baseia-se no sistema visual humano, isto é, um sistema binocular. Os olhos humanos estão separados por uma distância de aproximadamente 6.5cm o que faz com que tenham perspetivas ligeiramente diferentes da mesma cena. Estas duas imagens são processadas pelo cérebro e é a diferença entre elas que permite ter a noção de profundidade e saber assim a que distância é que um determinado objeto se encontra (Figura XX).

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

Os sistemas de captura estereoscópica baseiam-se no mesmo princípio para calcular o mapa de profundidades. São utilizadas imagens de uma determinada cena mas com perspetivas diferentes e a partir delas é construído um mapa de disparidades, isto é, um mapa com a representação das diferenças entre as duas imagens. Tendo essa informação juntamente com a informação relativa às camaras, é possível estimar a posição 3D dos pontos da cena através de triangulação. Numa fase prévia à captura, características como a posição relativa da câmara, orientação e os parâmetros internos (distância focal, centro ótico, parâmetros de distorção, etc.) são obtidas. A triangulação é feita para todos os pontos que têm correspondência nas duas imagens e para cada um é projetado um raio de acordo com as características recolhidas na fase de calibração. A interseção destes dois raios contém a representação 3D do ponto em questão. (Figura XX)



Imagem[[7]](#footnote-7)

O principal desafio da estereoscopia é o sistema de correspondência entre os pontos das duas imagens. Por norma, este método é apenas utilizado para a reconstrução de certos objetos com características fortes como cantos ou arestas bem definidas e em que as correspondências possam ser facilmente reconhecidas. Esta técnica requere um esforço computacional grande e a qualidade da captura está intimamente ligada à qualidade da fase de configuração/calibração e à qualidade dos sensores.

Atualmente já existem vários instrumentos capazes de fazer captura estereoscópica. Estes sistemas podem ter várias aplicações pelo que as que mais se distinguem são a realização de filmes em 3D, a inclusão em sistemas de robótica e a construção de sistemas de entretenimento 3D *low-cost* como consolas e telemóveis (FiguraXX)*.* O facto de serem uma tecnologia com um custo reduzido fez com que a estereoscopia não estagnasse na área industrial e com que a passagem para o mercado acontecesse com alguma facilidade. De salientar a inclusão destes sistemas em consolas de jogos portáteis como a Nintendo 3DS que, além de possibilitar a captura de fotografias em 3D também tira partida da visão estereoscópica para a criação de melhores sistemas de realidade aumentada.

Imagens[[8]](#footnote-8)

Time-Of-Flight

O princípio por de trás desta tecnologia baseia-se na medição da distância aos pontos das superfícies através do tempo que a radiação emitida demora a chegar aos objetos e voltar ao scanner. Sabendo esse tempo, a velocidade e a direção da emissão da radiação, é possível saber a distância a que determinada superfície se encontra e qual as coordenadas 3D dos pontos. Os sistemas Time-of-Flight (ToF) não utilizam contacto e são reflexivos e, dependendo da radiação usada, podem ser não-óticos como os radares (ondas eletromagnéticas ou de baixa frequência) ou os sonares (ondas acústicas), ou podem ser classificados como óticas ativas no caso dos radares óticos. Estes sistemas podem ainda ser referidos como LIDAR (LIght Detection and Ranging) ou LADAR (Laser Detection and Ranging) Ref[[9]](#footnote-9).

No caso dos sensores ToF ponto a ponto a distância de um determinado ponto na cena é calculado pelo princípio explicado em cima ou seja, a distância à câmara de cada ponto P é dada por

P = (CT)/2

Onde C é a velocidade da radiação e T o tempo medido correspondente á ida e volta da radiação.



Este processo é efetuado para todos os pontos que se pretende medir o que normalmente implica a movimentação do sensor tornando-o assim inapto para capturas dinâmicas, isto é, de cenas em que existe movimento. Por outro lado, no caso dos sensores ToF matriciais, a geometria é capturada num único instante por uma matriz de sensores. Cada elemento dessa matriz faz a medição de forma independente produzindo um mapa de profundidade a velocidades interativas. Ref[[10]](#footnote-10)

Uma vez que este tipo de tecnologia envolve a velocidade da luz (3x108m/s), o detalhe deste tipo de tecnologias fica limitada à velocidade a que os sensores conseguem fazer essas medições. Por exemplo, para se medir um detalhe com 1mm, a diferença entre medições é de 5ps o que exige a existência de um relógio capaz de medir passos dessa grandeza. A escolha de diferentes tipos de relógios leva a diferentes tipos de sensores ToF pelo que os mais comuns sãos a abordagem de modulação da intensidade de ondas contínuas, obturadores óticos e Single photon avalanche diodes, Outros fatores como a iluminação externa do ambiente ou interferências de outras camaras também causa perturbações na captura levando a medições imprecisas ou erradas.

Este tipo de sensores tem uma utilização bastante versátil uma vez que a amplitude do seu alcance é muito grande. Sistemas como o FARO 3D imager [[11]](#footnote-11)ou o Riegl VZ-6000[[12]](#footnote-12) foram desenvolvidos para longo alcance podendo efetuar medições a várias centenas de metros. Estes sistemas são ideais para fazer a medição de terrenos, tanto no chão como a partir de meios aéreos, de forma a produzir mapas topográficos. A arquitetura e construção civil também são áreas que podem beneficiar desta tecnologia através de medições e consequente validação em diferentes fases de projetos. No caso da área de preservação cultural e arqueologia estes sistemas são também utlizados para construir modelos 3D com grande detalhe e de forma não intrusiva.

Por outro lado, sensores como a CamCube[[13]](#footnote-13), a SR4000[[14]](#footnote-14) ou a Creative[[15]](#footnote-15) foram desenvolvidos para curto alcance (0.8-5m) atingindo velocidades de captura superiores aos 30fps. Estas características fazem com que os sensores consigam detetar e seguir gestos efetuados por humanos com facilidade permitindo a criação de interfaces naturais, úteis para áreas como o entretenimento, robótica ou até medicina. A possibilidade de aquisição de modelos também pode ser explorada podendo-se tirar partido da aquisição em tempo real para beneficiar a captura de ambientes voláteis ou em movimento.

Luz Estruturada

Os sistemas de luz estruturada enquadram-se na categoria de métodos de captura que não utilizam contacto, são refletivos e ativos. Este tipo de sistemas normalmente é composto por uma componente emissora e um ou mais sensores de captura. Para realizar este processo é utilizado um princípio similar àquele que é usado nos métodos estereoscópicos, a triangulação ativa onde, neste caso, a perspetiva da segunda camara é substituída pela entidade emissora. O emissor projeta um padrão na cena que depois é capturado pelo sensor. Essa informação juntamente com os dados recolhidos no processo de calibração é analisada e a partir das deslocações encontradas no padrão, são calculados os pontos 3D do mundo.

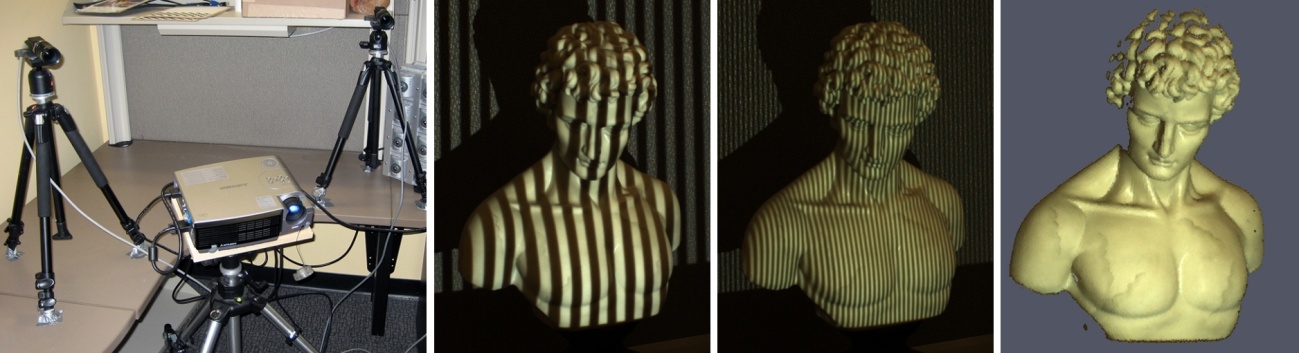
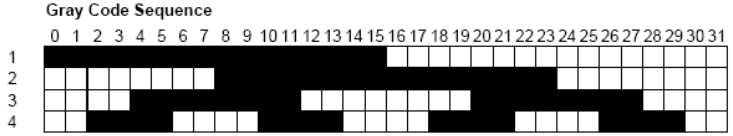


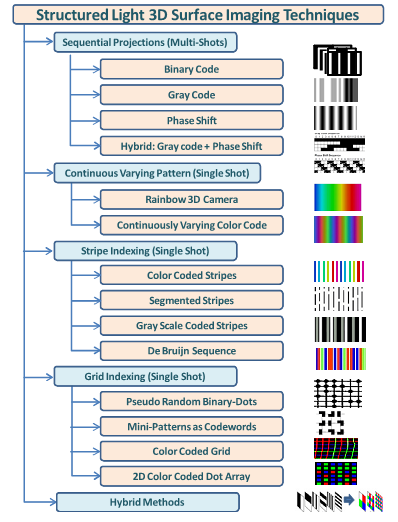
Imagem: http://mesh.brown.edu/3dpgp-2009/homework/hw2/hw2.html

Existem várias técnicas de luz estruturada que diferem essencialmente no tipo de padrão projetado o que, consequentemente, influencia a qualidade e o tempo da captura. Uma das técnicas usadas é o Gray Code, uma técnica sequencial de padrões binários compostos por tiras pretas e brancas seguindo a sequência de Gray (FiguraXX).

A análise dos diferentes padrões permite a identificação de zonas únicas com tiras únicas que, baseando-se no princípio da triangulação conseguem gerar pontos 3D da cena. Esta técnica é muito precisa e bastante tolerante às texturas das superfícies uma vez que apenas usa valores binários. Em contra partida, para se conseguir alcançar resoluções elevadas é necessário um grande número de padrões sequenciais e, como tal, a aquisição torna-se mais demorada. Isto faz com que este método seja inapto para capturas de cenas dinâmicas ou entidades vivas como o ser humano.



Outros métodos como a variação de padrões coloridos[[16]](#footnote-16), métodos Stripe Index[[17]](#footnote-17) ou padrões de grelhas 2D espaciais[[18]](#footnote-18) permitem a aquisição da informação 3D do objeto em apenas uma captura, conferindo-lhes as propriedades de necessária para suprir os problemas descritos dos códigos sequenciais. Na figuraXX são indicadas algumas dessas técnicas de acordo as metodologias utilizadas. Ref[[19]](#footnote-19)



Apesar de grande variedade de métodos e das diferentes vantagens de cada um deles, os sistemas de luz estruturada têm algumas desvantagens que são comuns a todos. A área capturada é limitada pelo alcance do projetor pela amplitude do sensor de captura o que normalmente se traduz em áreas de ação pequenas e como tal, o processo de captura de objetos fica limitado a objetos de pequena dimensão. Outro problema é a aquisição das características como a cor do objeto ou a luz ambiente uma vez que a sobreposição da luz projetada adultera essas propriedades. Neste caso, o problema pode ser ultrapassado usando padrões de luz que não pertençam ao espetro visível, tal como é feito com a Kinect e que será explicado no Capítulo XX.

Estes sistemas são utilizados para a aquisição de modelos com grande detalhe nas mais diversas áreas. Várias empresas, como a 4DDynamics[[20]](#footnote-20), a 3D-Shape[[21]](#footnote-21) ou a Createform[[22]](#footnote-22) oferecem produtos e serviços orientados para o scan de alta resolução do corpo-humano ou apenas da face. Isto pode ter aplicações não só em áreas como o cinema e o entretenimento mas também na saúde. Ainda nessa área, a Maestro3D[[23]](#footnote-23) e a Cynoprod[[24]](#footnote-24) especializaram-se na aquisição de modelos dentários conseguindo produzir modelos com grande detalhe de forma rápida. A FlashScan3D[[25]](#footnote-25), mais ligada às áreas das ciências forenses, construiu um scanner de que se destaca pela sua precisão para a captura de impressões digitais. Noutras áreas, como a indústria, estes sistemas também são usados por empresas especializadas como a Vitronic[[26]](#footnote-26) ou a Optimet[[27]](#footnote-27) que usam os scanners 3D para analisar peças complexas e inspecionar a sua qualidade. De salientar ainda a Kinect como um produto que chegou aos consumidores em grande escala e que tem visto a crescer o número e a variedade de aplicações que a usam, como será descrito na secção seguinte.

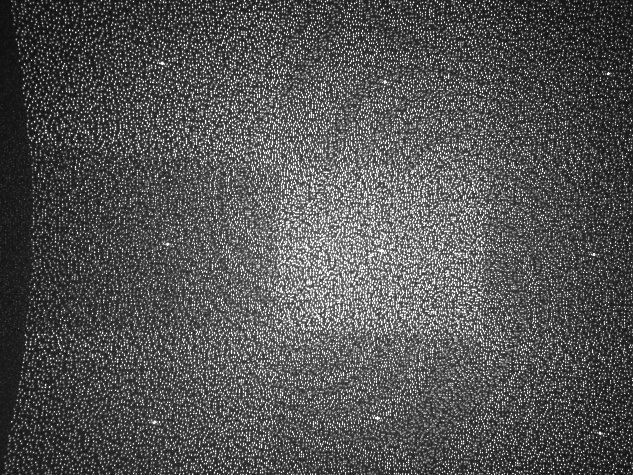
Microsoft Kinect

A Kinect, anteriormente conhecido como “Project Natal” (FiguraXX), é uma camara RGBD, ou seja, uma camara capaz de capturar simultaneamente imagem de cor (RGB) e de profundidade ((D)epth). Foi a primeira camara a sair para o mercado que juntou num só dispositivo esses sensores e também um microfone multi-array e um processador interno com software proprietário.



Imagem[[28]](#footnote-28)

A nível de visão, a Kinect vem equipada com um sensor de cor tem a capacidade de capturar imagens RGB 640x480 a 30fps ou com uma maior resolução, 1280x960 no entanto a 10fps. O sistema de captura de profundidade é composto por emissor e um sensor de infravermelhos. Apesar dos resultados serem semelhantes aos de sistemas de Time-of-Flight, a Kinect recolhe a informação de profundidade através da técnica de luz estruturada no espetro infravermelho. É emitido um padrão de pontos esparsos pelo emissor IR que depois de analisado produz o mapa de profundidade, com uma dimensão de 640x480 e um nível de detalhe de 11bits. Informação mais detalhada sobre o funcionamento e limitações da Kinect será abordada na secção XX. Ref[[29]](#footnote-29)



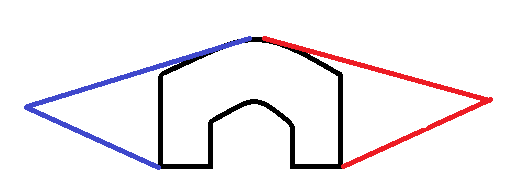
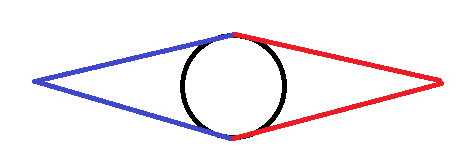
Apesar de não ser o único sensor RGBD com este tipo de características presente no mercado (existe também a Asus Xtion Pro Live[[30]](#footnote-30) e a PrimeSence Carmine 1.09[[31]](#footnote-31)), a Kinect foi a que teve uma maior aceitação. Este sensor foi desenvolvido e lançado pela Microsoft no final de 2010 enquanto sensor de movimento e como acessório da Xbox 360. Foi publicitado como uma nova forma de interação no mundo dos jogos pois não haveria a necessidade da utilização de um controlador físico, apenas faria uso de gestos e voz. Este dispositivo é um fenómeno de popularidade tendo atingido a marca de 1milhão de unidades vendidas em menos de duas semanas (estando este à venda apenas nos Estados Unidos) e a marca de 8 milhões nos primeiros 60 dias, o que lhe valeu a distinção de Fast Selling Gaming Peripheral no livro de recordes do Guinness[[32]](#footnote-32). No entanto, a popularidade deste dispositivo não se ficou pelos jogos da consola e também suscitou grande curiosidade no mercado de desenvolvimento de software devido às suas características. A junção da informação entre camara de cor e a de profundidade possibilitou a deteção de utilizadores na cena e inferir os seus esqueletos, com 48 pontos de controlo, em tempo real. A sua eficácia, precisão e baixo custo tornaram-na num objeto de eleição para projetos em áreas tão diferentes como robótica, jogos, interfaces naturais ou performances artísticas.

Quando saiu, a 4 de novembro de 2010, a Kinect foi lançada enquanto acessório de uma consola de jogos e não foi disponibilizado nenhum software para desenvolvimento. Nessa mesma data e de forma a contornar essa situação, a empresa Adafruit Industries ofereceu uma recompensa[[33]](#footnote-33) para quem desenvolvesse um driver opensource capaz de aceder à informação do sensor. Seis dias mais tarde e depois de a recompensa ter triplicado para os 3000$, foi lançada a primeira versão do *libfreenect*, um driver livre capaz de ler o stream de vídeo de cor e de profundidade da camara. Um mês mais tarde e em resposta ao crescente interesse e número de projetos da comunidade opensource, a PrimeSence lançou o seu próprio driver, também ele opensource, a framework OpenNI e ainda os binários para o middleware NiTE, capaz de detetar e fazer o tracking do esqueleto entre outras funcionalidades. A Microsoft apenas entrou neste mercado em Fevereiro do ano seguinte lançando um SDK para fins não comerciais.

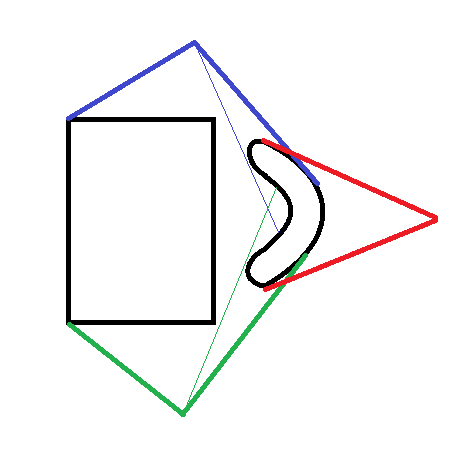
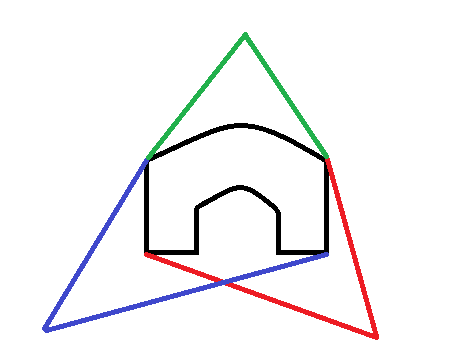
Desde então os softwares evoluíram oferecendo maior precisão e mais funcionalidades (tracking de mãos, deteção de gestos, reconhecimento de voz, etc) e o hardware também acompanhou. E Maio de 2012 foi lançada a Kinect for Windows, um sensor praticamente igual ao original e orientado ao mercado dos computadores pessoais. Este sensor tinha apenas como diferença a adição da funcionalidade de Near Mode que, respondendo ao seu novo propósito, lhe dá uma maior precisão a distâncias mais próximas da câmara. A próxima iteração na evolução neste hardware está prevista para o início de 2014 com o lançamento da Kinect2, juntamente com a nova geração da consola da Microsoft. Muitas especulações têm sido feitas sobre o novo dispositivo[[34]](#footnote-34) e entre elas, espera-se que o novo sensor tenha uma imagem de cor FullHD a 30fps, um maior campo de visão (70ºhorizontal, 60ºVertical) e uma stream de profundidade com uma maior resolução e uma qualidade superior em zonas de conflito como arestas e objetos que se encontrem mais próximos da câmara. (?)

Sistemas de Aquisição 360º

Como foi descrito na introdução, a aquisição 360º corresponde á captura de informação de várias perspetivas centradas no objeto a capturar. Para realizar este tipo de captura precisamos no mínimo de duas perspetivas diferentes do objeto. Como se pode ver na figura em baixo, se o objeto tiver uma geometria simples, como é o caso de uma esfera, tendo apenas duas perspetivas consegue-se capturar a informação de toda a superfície do objeto. No entanto, como se pode ver no segundo exemplo, caso o objeto a capturar seja um pouco mais complexo, tal já não é possível e, independentemente da posição em que se irá capturar a informação, haverá zonas do objeto em que não será possível capturar informação.



Neste caso específico, o problema poderia ser ultrapassado com a adição de mais uma fonte captura (FiguraXX) e desta forma, toda a superfície do objeto voltaria a estar coberta. Contudo, com o aumentar da complexidade do objeto ou até com a existência de múltiplos objetos em cena, esta dificuldade voltará a surgir e haverá casos em que a adição de novas perspetivas para aquisição já não conseguirão resolver a situação. Pode haver partes dos objetos que não é possíveis capturar e este problema apenas poderá ser ultrapassado com outro tipo de sistemas de captura e não apenas os óticos.

. 

Aquisição Estática

Os pressupostos da aquisição estática de um objeto são que, tanto o sistema de aquisição como o objeto a capturar não necessitam de se mover para a realização da captura 360º. Esta limitação requere a introdução de múltiplas fontes de captura que, por norma, implicam a aquisição de mais camaras. Isto, além de introduzir mais complexidade e carga computacional, torna o custo do sistema mais elevado. Por outro lado, com um setup estático tem-se como uma das vantagens a possibilidade de aquisição de informação das diferentes perspetivas em simultâneo. Esta característica beneficia as capturas de entidades que estejam em movimento uma vez que permite tirar um “snapshot” ao espaço num determinado momento ou até a gravação contínua de vídeo 3D.

Já existem alguns produtos e serviços no mercado que oferecem este tipo de aquisição:

A SpaceVision[[35]](#footnote-35) tem um produto de nome Cartesia Series Portable 3D Body Scanner que afirma ser o primeiro sistema de scan corporal portátil do mundo. É constituído por nove camaras distribuídas igualmente por três torres e usa laser scanning para realizar a aquisição de informação. A captura demora aproximadamente dois segundos a ser feita e é gerada uma nuvem de pontos com cerca de 1milhão de pontos e com um erro de medição médio inferior a 3mm.

A 4DDynamics[[36]](#footnote-36) comercializa um sistema para fazer a captura 3D do corpo de uma pessoa e como tal está orientado para a indústria médica ou cinematográfica. O setup é constituído por 2 a 8 scanners (camaras) e é usado um sistema de projeção de padrões de luz estruturada. A captura não é feita instantaneamente (demora 1 a 2 segundos dependendo da configuração) no entanto foi desenvolvida uma técnica “steady Scan” para compensar pequenos movimentos como aqueles resultantes do batimento cardíaco ou da respiração.



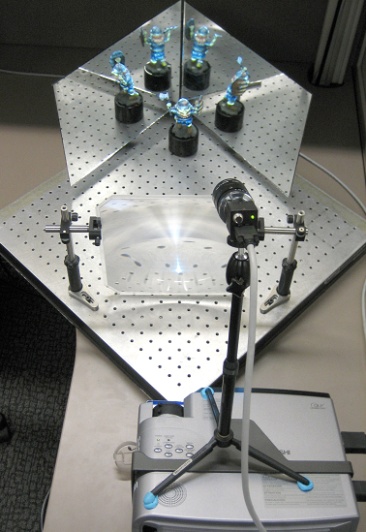
A IR-Entertainment[[37]](#footnote-37) utiliza um sistema em forma de anel com câmaras Canon DSLR de 18mp (FiguraXX). A geração dos modelos é conseguida através de técnicas de fotometria e luz estruturadas dando origem a modelos de qualidade muito elevada. O número de camaras utilizadas depende do fim pretendido variando entre as 52 para a realização de captura facial e as 115 para a captura corporal. Em ambos os casos são produzidos modelos 360º Gigapixel. Este sistema consegue realizar a captura instantânea de poses estáticas ou em movimento no entanto não permite captura contínua de informação.

A TC2[[38]](#footnote-38) tem também um sistema orientado à captura 3D do corpo humano mas esta recorre à utilização de Kinects para o fazer. São utilizados 16 sensores de profundidade espalhados equitativamente por quatro colunas. A aquisição demora cerca de 7segundos a ser realizada e consegue atingir uma precisão de 3mm. Este sistema foi construído a pensar principalmente no mercado têxtil podendo ser usada como cabine para a prova virtual de vestuário ou para a extração das medidas do corpo e posterior customização de sugestão de tamanhos de roupa.



Existem ainda outros sistemas, estes numa vertente mais académica, capazes de realizar o mesmo tipo de captura. Em ArtigoX[[39]](#footnote-39) foram usadas técnicas muito similares às usadas pela TC2. São utilizadas 8 Asus Xtion Pro Live para realizar a captura que estão também elas espalhadas por 4 barras verticais. Os resultados mostraram um erro máximo inferior a 20mm. Em ArtigoY[[40]](#footnote-40) a técnica usada também foi muito semelhante pelo que neste sistema forma utilizadas 5 Kinects, uma a cerca de 1.30m do chão apontada para cima de forma a capturar com mais detalhe a parte superior da pessoa, e as outras 4 em forma de círculo com 3.60m de diâmetro e a 1.80 m de altura de forma a conseguir detetar todo o corpo. O detalhe deste sistema não foi especificado nos resultados.

No ArtigoZ[[41]](#footnote-41), foi desenvolvido um sistema capaz de realizar a captura facial usando apenas uma camara e dois espelhos de forma a gerar mais perspetivas sobre o indivíduo. A partir das três perspetivas adquiridas o mapa de profundidade é construído por um sistema baseado na estereoscopia e na minimização de energia atingindo um detalhe inferior a 4% do tamanho do modelo capturado. Já em ArtigoW[[42]](#footnote-42), o sistema construído, apesar de usar igualmente uma câmara e dois espelhos, estes foram colocados de forma a conseguir-se extrair numa só imagem 5 perspetivas diferentes do objeto a capturar: uma perspetiva correspondente à visão da camara, duas outras vindas da reflecção direta dos espelhos e ainda mais duas resultantes da reflecção entre espelhos como se pode ver na imagem seguinte.

Este sistema utiliza o código de Gray como padrão de luz estruturada para fazer a aquisição 3D em tempo real. Não fomam fornecidos dados sobre o tempo específico da captura nem da qualidade da mesma além de imagens representativas.

Aquisição Móvel

Este tipo de aquisição tem mais liberdade que a aquisição estática uma vez que tanto o sistema de aquisição como o objeto deixam de ter a restrição de imobilidade.

O facto de se poder mover o sistema de captura ou o objeto em questão faz com que se consiga obter informação de todas as perspetivas do objeto usando apenas um sensor, limitando assim o custo do sistema. Para realizar esta tarefa basta recolher amostras das várias perspetivas movendo para isso a camara em torno do objeto ou rodando o objeto à frente da camara. A junção das nuvens de pontos das várias perspetivas pode ser feita após as capturas ou, como em X e Y, à medida que esta feita e através de SLAM (Simultaneous Location and Mapping) Ref para SLAM??.

No caso de objetos estáticos, apesar de levar mais tempo a realizar a captura, isto não constitui um problema pois o objeto mantém-se inalterado. No entanto, se se quiser capturar a geometria de uma pessoa, este procedimento torna-se mais complicado uma vez que a pessoa pode mexer-se durante o processo da captura introduzindo erro uma vez que o modelo é estático.

Já existem alguns produtos e serviços no mercado que oferecem este tipo de aquisição:

A Makerbot lançou o Digitalizer[[43]](#footnote-43), um scanner 3D orientado para objetos de pequenas dimensões (20.3cm de diâmetro por 20,3cm de altura). Este sistema utiliza dois lasers e uma camara de 1.3mp para realizar a captura conseguindo produzir modelos com uma resolução de 0,5mm e um erro aproximado de 2mm. O modelo é colocado no centro do sistema que está equipado com uma plataforma rotativa que faz rodar o objeto de forma a realizar a captura 360º. Devido a este processo, o tempo de captura é de aproximadamente 12minutos. Um produto muito similar é o scanner 3D da Matterform[[44]](#footnote-44) que é capaz de capturar a geometria de objetos com dimensões até 18cm de diâmetro por 25cm de altura. Vem igualmente equipado por 2 lasers e uma camara HD e o detalhe dos modelos capturados é o mesmo. A velocidade de captura pode variar entre os 5 e os 10 minutos consoante o nível de detalhe desejado: modelos mais detalhados, mais tempo de aquisição.

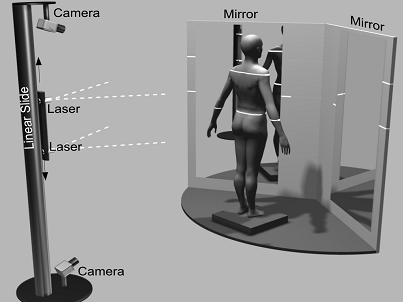


Um tipo de sistemas diferentes são os scanners de mão ou portáteis. Estes sistemas normalmente são pequenos e de fácil transporte e manuseamento. A captura é feita através da movimentação do scanner à volta do objeto de forma a captura-lo de todas as perspetivas necessárias. A tecnologia utilizada por estes sistemas varia sendo que a mais utilizada é a luz estruturada. A Go!SCAN 3D[[45]](#footnote-45) e a Artec MTH[[46]](#footnote-46), por exemplo, utilizam um padrão de luz branca codificado através de uma LED e apenas uma luz de um flash normal respetivamente. Ambos os sistemas são utilizados a distâncias curtas (0,4m – 1m) e oferecem resultados muito precisos com uma resolução de 0,5mm e detalhes de 0,1mm.



Outro produto que pode ser enquadrado nesta categoria é a Kinect. Apesar de este sensor poder ser utilizado de forma estacionária, para a realização de uma captura 360 uma das abordagens mais comuns é a movimentação da camara em torno do objeto de forma a realizar toda a captura. Como já foi descrito anteriormente, este sensor utiliza um padrão de luz infravermelha estruturada e permite a captura simultânea da imagem de cor e de profundidade. Apesar de não ser tão precisa como outros produtos, consegue produzir resultados com uma resolução espacial de 0,75mm e detalhes de 1.5mm a um baixo custo. Várias empresas desenvolveram software que utiliza a Kinect como sensor de aquisição. Alguns exemplos são a KScan3D[[47]](#footnote-47) ou a ReconstructMe[[48]](#footnote-48) que, utilizando um ou mais sensores, permitem a aquisição de geometria de objetos singulares ou até a reconstrução de espaços completos como uma sala de estar ou até toda uma casa.

Já no ArtigoY[[49]](#footnote-49) um outro tipo de sistema foi implementado utilizado espelhos de forma a aumentar o número de perspetivas a captura. Neste caso foi construída uma cabine utilizando dois espelhos planos e para a aquisição é usada pelo menos uma camara dando assim ao sistema três perspetivas diferentes do indivíduo para a geração do mapa de profundidade.



A técnica usada para a geração do mapa de profundidade foi a de laser scanning com um módulo móvel de lasers (o número de lasers pode variar) que se movimenta linearmente na vertical. O tempo de geração do modelo não é especificado no entanto como o sistema é móvel, a captura não pode ser feita em tempo real. O erro medido neste sistema é inferior a 4mm.

Aplicações

A aquisição de informação em 360º permite capturar a geometria de um objeto de forma semiautomática, o que leva à poupança de tempo, dinheiro e materiais. Esta característica possibilitou que a criação de modelos 3D deixasse de ser exclusiva de modeladores e artistas 3D e passasse a estar disponível para a maior parte dos públicos. Desta forma, o número de aplicações que começaram a usar este tipo de informação e a variedade das áreas aplicáveis aumentou, como já se foi mostrando nas secções anteriores.

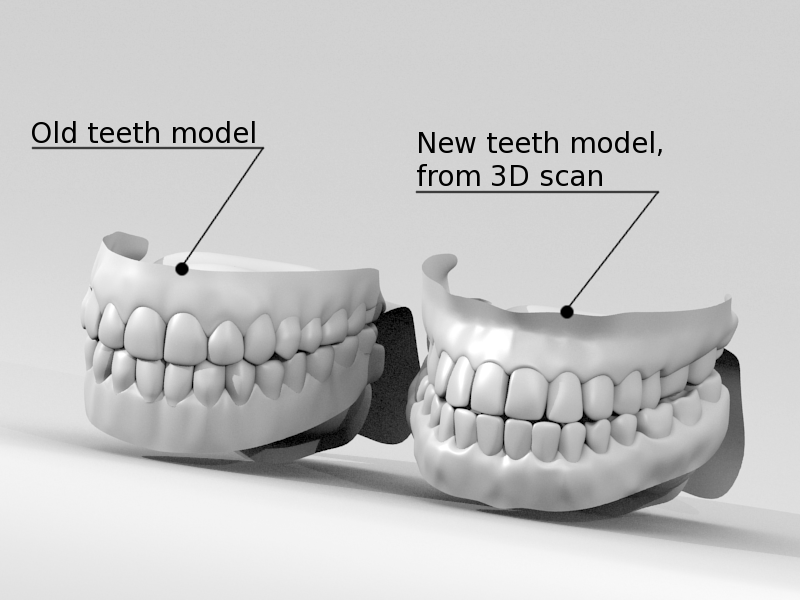
Nesta secção serão mostradas algumas das possíveis aplicações dos resultados destes *scanners*, dividindo-os nas três áreas que atualmente têm mais utilização: objetos, corpo humano e o cenário completo.

Objetos

A aquisição da geometria de objetos por scanners 3D permite que o modelo dos mesmos seja adquirido muito mais facilmente e de forma mais rápida. Dependendo dos materiais dos objetos e da técnica usada, consegue-se chegar a resoluções e precisões inferiores a milímetros. Esta informação pode ser usada de várias formas.

Em áreas como a indústria de videojogos e a indústria cinematográfica, estes modelos podem ser introduzidos diretamente nos processos de modelação ou até de render poupando assim tempo na criação dos mesmos. Além das propriedades geométricas, também podem ser recolhidas informações sobre materiais e textura permitindo posteriormente a correta iluminação dos modelos. Mais orientado á área do consumo e aliado ao aparecimento das impressoras 3D, a aquisição de modelos e posterior replicação é também algo a ter em atenção. Isto pode facilitar processos de reparação/substituição de peças danificadas ou simplesmente pode ser usado para criação de objetos em miniatura de entidades reais. Apesar deste tipo de impressoras estar onde estavam as impressoras normais há 25-30 anos, as melhorias na qualidade de impressão, aumentos de desempenho e a redução de custos são vistos com frequência sugerindo assim bons indicadores para o futuro destes dispositivos.

Noutras áreas como a preservação histórica, a existência dos modelos 3D de objetos históricos é também uma mais-valia, permitindo uma análise detalhada dos mesmos sem restrições espaciais nem com o perigo de danificar o objeto. Por outro lado no caso de objetos já danificados, a aquisição 3D dos fragmentos pode ser usada também para a reconstrução digital do objeto original sem o perigo de incorrer em mais danos. Para além do estudo e visualização virtual da informação, estes modelos também podem ser usados para produzir réplicas, sejam elas miniaturas para souvenires como peças idênticas para, por exemplo, exposição temporária enquanto o original é restaurado. Já existem vários casos nesta área como, por exemplo, a estátua de Michelangelo (FiguraXX[[50]](#footnote-50)) que foi captura com um detalhe de 0,29mm resultando numa malha crua de 2 biliões de polígonos.

Outras áreas onde o detalhe é também importante são a indústria e a medicina. No caso da indústria, estes modelos são usados tanto para controlo de qualidade como para processos de *reverse engineering*. Além da automatização do processo do controlo, o facto de estas atividade poder ser realizada por máquinas permite ainda que estas possam ser efetuadas em zonas inacessíveis ou perigosas para o homem. Já na medicina, os scanners 3D são usados para digitalizar, por exemplo, estruturas dentárias (FiguraXX) ou até dos ossos que depois são usadas para a geração de modelos á medida de implantes ou próteses.

Outra utilização dos modelos 3D de objetos que, de certa forma, é transversal a todas as áreas, é a prototipagem rápida. Com estes instrumentos é possível digitalizar de forma rápida informação tridimensional de protótipos construídos à mão o que permite depois ilustrar ideias com a adição de informação digital. Esta solução é cada vez mais usada por designers e pode ser útil para a criação e disseminação de ideias de novos produtos.

Corpo Humano

A aquisição do corpo humano é uma das vertentes mais em popular dos scanners 3D. Além da extração do modelo de pessoas, a utilização de informação 3D para produzir informação lógica sobre o corpo humano abriu novas portas à área de interação homem-máquina.

No campo dos jogos, o campo da interação natural tem como exemplo de maior sucesso a Kinect para a Xbox360. No entanto esta não é a única área onde este tipo de interação natural é usado. A utilização de gestos é usada para navegar em menus aplicados às mais diversas áreas e, por exemplo, nas áreas de robótica, indústria ou medicina, pode ser usado para controlar os movimentos de máquinas, o que pode ser útil em ambientes em que seja impossível ou perigosa a presença humana. Neste último caso, na área da saúde, o controlo de materiais médicos sem a necessidade do contacto direto permite inclusive a realização de cirurgias à distância com a adição de tecnologia que permita a visualização em tempo real. Ainda na área da saúde, outra utilização do reconhecimento de gestos é a interpretação de linguagem gestual. Outra área que também utiliza o reconhecimento de gestos e também o volume a presença de pessoas é a arte digital havendo atualmente várias performances artísticas e peças de teatro a utilizarem esta tecnologia.

Por outro lado, a aquisição da geometria do corpo humano e criação do respetivo modelo tem também várias aplicações úteis. Uma das mais comuns é na área do cinema e dos videojogos para a criação e animação de personagens virtuais de forma realista. No caso específico da aquisição facial, a produção de informação altamente detalhada permite animar e dar expressões quase naturais às personagens, como foi usado em filmes como Timtim e Avatar (FiguraXX), ou em jogos como Mass Effect ou Dead Island. Outra área onde a geometria do corpo é também usada com sucesso é no retalho, por exemplo, com a construção de provadores virtuais. Através da captura de informação 3D são extraídas e usadas as medidas e volumetria das pessoas. A partir desta informação os sistemas conseguem calcular quais os tamanhos reais das pessoas que posteriormente, através de Realidade Aumentada, podem ver como as peças de vestuário ficariam, mesmo que os artigos não estejam disponíveis no local. Como foi referido em relação à interação, a utilização apenas da geometria do corpo é também usada em peças de arte digital para interação com elementos virtuais.

Os campos da saúde e da medicina são talvez as áreas onde a utilização da geometria do corpo humano é mais útil e versátil. A aquisição da informação 3D do corpo é utilizada para monotorização e para a procura e visualização de anomalias como malformações ou resultados de acidentes. Os exemplos mais comuns são os raio-x, a tomografia computorizada e as ressonâncias magnéticas. No caso de sistemas que conseguem fazer este tipo de observações em tempo real é ainda possível a deteção de movimentos em escalas milimétricas, por vezes impercetíveis ao olho humano, que são analisados de forma a detetar irregularidades. Ainda relacionado com esse tipo de movimentos, no campo das ciências forenses, há sistemas de aquisição facial de alta precisão que permitem a deteção de micro-expressões faciais que depois são utilizadas, por exemplo, na análise de depoimentos para o apuramento da veracidade das afirmações. Outra utilização destas tecnologias na saúde é na criação de modelos a partir de partes do corpo da pessoa. Estes modelos podem ser usados para a criação de próteses ou implantes que são usados nas próprias pessoas, aumentando assim o realismo dessas peças e semelhanças com as partes originais. Esta técnica também é usada para planeamento de cirurgias plásticas. A nível do ensino, o scan do corpo humano, além de permitir o estudo virtual minucioso, possibilita a criação de modelos ainda mais detalhados que são usados para, por exemplo, treinar cirurgias sem colocar em perigo nenhum paciente.

Cenário

Os scanners 3D também podem ser usados para fazer a captura de todo a estrutura de um determinado ambiente e não apenas partes dele. Estas capturas são, por norma, dinâmicas e evolvem algoritmos simultâneos de localização e mapeamento (SLAM).

Na área do entretenimento este tipo de captura permite a criação de ambientes virtuais semelhantes aos reais de forma fácil e relativamente rápida. Isto pode aumentar a imersividade no campo dos jogos uma vez que a navegação passa a ser feita num cenário virtual muito semelhante ao real. Esta característica também pode ser usada na área da cultura possibilitando a exploração e visitas virtuais realistas de museus ou espaços que estão geograficamente espalhados pelo mundo. Já no cinema, a construção destes mundos virtuais permite a realização de “free view-point videos”, isto é, peliculas em que o utilizador pode navegar pelo espaço (virtual) enquanto a narrativa ou as animações decorrem. Outro tipo de aquisição que atualmente é usada no cinema é a estereoscopia produzindo a noção de profundidade que é mostrada nos filmes 3D.

Por outro lado e numa vertente menos interativa, a área de preservação histórica também pode beneficiar da aquisição 3D de espaços histórico. A existência dos modelos 3D destes espaços possibilita que informação detalhada possa ser difundida digitalmente e desta forma, chegar a muitas mais pessoas. Além disso, permite ainda a visualização do espaço quando estes estão inacessíveis e potenciando assim a preservação dos mesmos.

Outra área que também tira partido dos scanners 3D, principalmente daqueles aptos para a captura em tempo real, é a robótica. Atualmente existem vários projetos que integram estas duas componentes adicionando a informação tridimensional do espaço à máquina. Por um lado, isto pode ser usado para o robot conseguir interagir com o cenário que o rodeia, mesmo que estes sejam dinâmicos, dando-lhe a capacidade de, por exemplo, manipular objetos ou interagir com outras máquinas. Por outro e no caso dos robots com mobilidade, a noção da tridimensionalidade do espaço facilita a navegação no mesmo através de SLAM (FiguraXX). Além da possibilidade de fazer a construção do modelo desses espaços autonomamente e em tempo real, isto abre portas á navegação autónoma. Esta área tem tido vários desenvolvimentos nos últimos anos e é aliciante sobretudo quando realizado ao ar livre, possibilitando assim avanços no campo dos automóveis autónomos.

Sumário

A captura de informação 3D está em rápida evolução e os sensores aptos para o fazerem estão finalmente a chegar aos consumidores com preços mais acessíveis, sendo que um dos sensores em maior destaque a Microsoft Kinect pela sua velocidade, precisão e baixo custo. Dependendo dos objetivos pretendidos, existem várias tecnologias capazes de realizar a captura de informação em 3D com características diferentes pelo que as que mais se destacam são a Estereoscopia, Time-of-Flight e Luz Estruturada. No caso dos sistemas de aquisição de 360º, existem vários mas com características diferentes que oscilam entre a qualidade e velocidade da aquisição. As suas aplicações são variadas abrangendo áreas tão diferentes como os videojogos, indústria, medicina ou artes.

Visão do Sistema (?)

Como foi descrito na introdução, o objetivo desta tese consiste em criar um sistema de baixo custo capaz de fazer a aquisição 360º de informação 3D de um objeto em tempo real, a partir de um *setup* estático e utilizando apenas uma câmara. Estas características geralmente contrapõe-se e como tal é um desafio conseguir concilia-las no mesmo sistema.

Com isto em mente, neste capítulo pretende mostrar-se como foi construída a visão deste sistema a partir dos requisitos enunciados. Inicialmente serão descritos em detalhe os objetivos e as características pretendidas para o sistema e de seguida serão expostas e justificadas as decisões tomadas, tanto a nível de material utilizado para a realização da captura como para a construção do setup físico. Isto determinará a forma como o sistema será construído e quais as suas características técnicas do mesmo. Este capítulo será concluído com a apresentação de algumas das possíveis aplicações deste sistema.

Descrição e Funcionalidades

Características como baixo-custo, aquisição 360º 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 360º de um objeto é útil para obter a geometria quase total do mesmo, apenas as partes ocludidas às perspetivas capturadas é não são adquiridas. Este tipo de informação pode ser utilizada para diferentes propósitos como será detalhado na secção CASOS DE USO. Esta aquisição pode ser feita de forma faseada, isto é, uma perspetiva de cada vez, ou então em simultâneo, capturando todas as perspetivas de uma só vez. No entanto, como uma das características da captura é que esta seja em tempo real, os métodos faseados não podem ser considerados. Dependendo da tecnologia utilizada, este tipo de captura pode levar alguns segundos ou até minutos a ser realizada e desta forma, a captura de objetos dinâmicos ou de entidades em movimento torna-se inviável: a informação das diferentes perspetivas com tempos também diferentes resultaria em aquisições erradas. Desta forma, a realização da captura das várias perspetivas em simultâneo traz vantagens a nível do tempo de captura de objetos estáticos e permite também a aquisição de cenas dinâmicas. Da mesma forma, a utilização de um setup estático torna-se também obrigatória uma vez que os setups móveis requerem obrigatoriamente mais tempo para realizar a captura 360º.

Por outro lado, o requisito de utilização de uma única câmara está apenas relacionada com o fato de se querer que o sistema também seja de baixo custo. Apesar de apresentar uma dificuldade adicional na conceção do sistema, este é um fator a ter em conta uma vez que os sensores de aquisição não são por norma acessíveis. Desta forma, a utilização de múltiplos sensores é descartada uma vez que, além de aumentarem a carga computacional do sistema, também aumentam o custo do mesmo.

Já existem algumas soluções que permitem fazer a aquisição 360º de informação 3D, no entanto, estas têm algumas características diferentes das propostas nesta dissertação. As soluções da 4DDynamics[[51]](#footnote-51) e a IR-Entertainment[[52]](#footnote-52) funcionando como estúdios e estão orientadas a modelos com grandes dimensões e à aquisição de modelos estáticos. Isto faz com que estas soluções tenham um nível de detalhe muito elevado mas em contrapartida requerem instalações de grande dimensão que, devido à qualidade e quantidade de material utilizado, são também soluções dispendiosas. A KScan3D[[53]](#footnote-53) por sua vez oferece uma solução mais acessível, utilizando Kinects como instrumento de captura, e permite realizar a aquisição 360º de objetos estáticos mas de forma faseada. Os modelos obtidos têm também menos pormenor que as soluções anteriores devido às capacidades do hardware usado.

Outras soluções, como o Digitalizer[[54]](#footnote-54) ouo *scanner* 3D da Matterform[[55]](#footnote-55), têm objetivos diferentes e estão orientadas à portabilidade e à comercialização para o público geral. Estes *scanners* conseguem fazer a aquisição 3D a 360º de objetos estáticos de pequena dimensão. Estão pensados para a criação de modelos completos e detalhados e, fazendo a ponte com as impressoras 3D, a sua replicação.

Decisões

No que toca à perceção do conteúdo de uma determinada cena, os computadores têm algumas limitações uma vez que esta é, por norma, representada com informação em duas dimensões. Isto gera alguns problemas em campos como a segmentação da cena e a deteção e reconhecimento de objetos. A introdução da tridimensionalidade pode ajudara a resolver algumas dessas questões. De acordo com as características pretendidas e os objetivos traçados, os principais pontos de decisão prendem-se com o material utilizado para a realizar a captura de informação e o *setup* a usar de forma a conseguir realizar a captura 360º em tempo real.

Captura

A captura de informação consiste na conversão de informação do mundo real para o formato digital. No caso do vídeo e da aquisição 3D, a informação a capturar consiste na imagem e geometria das cenas. No capítulo anterior foram referidas várias abordagens para a captura de informação em 3D no entanto, nem todas servem para o propósito deste projeto. Além da qualidade da informação adquirida, questões como o tempo de captura, preço e o fato de se querer um sistema de captura 360º são os principais fatores de decisão.

Os sistemas baseados em estereoscopia são bastante usados na indústria cinematográfica para a realização de filmes em 3D. A qualidade dos resultados produzidos encontra-se apenas na ordem dos centímetros no entanto é o suficiente para o tipo de aplicações que os usam. Normalmente são utilizadas duas camaras simples para efetuar a captura e produzir o efeito 3D o que faz com que estes sistemas sejam de baixo custo e de fácil integração. Apesar da obtenção da informação em 3D exigir uma grande complexidade computacional a nível de processamento de imagem, esta pode ser processada em tempo real com os processadores mais recentes. Em contrapartida, a informação produzida está dependente do ponto de vista a partir do qual foi capturada pelo que, para aquisição 360º torna esta técnica pouco viável. A aquisição teria de ser efetuada de forma faseada, perdendo a característica de tempo real, ou utilizando múltiplas fontes e captura dispersas pelas várias perspetivas, o que faria com o que o custo da aplicação aumente.

Por outro lado, os sistemas Time-of-Flight são mais versáteis e podem ser usados para medições a curto ou longo alcance nas mais diversas áreas. Estes sistemas são refletivos e ativos, isto é, emitem radiação para conseguir fazer as medições, através do reflexo das mesmas. Apesar de ser uma tecnologia que exige uma carga computacional elevada, os sensores começam agora a ter a capacidade de fazer esse processamento localmente diminuindo assim a complexidade dos softwares que a usam. A qualidade dos resultados pode chegar aos milímetros, dependendo das condições, e têm uma taxa de atualização elevada, tornando-os úteis principalmente para sistemas interativos. Da mesma forma, estes sensores geram uma quantidade de tráfego elevada, exigindo assim interfaces capazes de o suportar. Por esta razão e uma vez que a tecnologia também necessita de utilizar material emissivo como LEDs ou lasers, estes sistemas têm por norma um custo mais elevado.

Os sistemas de Luz Estruturada primam por serem sistemas capazes de produzir resultados de elevada qualidade chegando a poder atingir detalhes de micrómetros. Uma vez que utilizam a emissão e análise de padrões, esta tecnologia está mais orientada a ambientes indoor e controlados e é ideal para a captura da geometria de objetos e criação de modelos 3D. Apesar ser uma tecnologia computacionalmente exigente pode atingir-se taxas de atualização interativas no entanto, nem todas as variantes desta tecnologia o permitem. Os padrões sequenciais, como o código de Gray, necessitam a emissão de várias imagens com padrões diferentes diferentes para a aquisição de uma única cena o que faz com que o processo demore mais tempo a ser executado. Por outro lado, o tipo de radiação emitida também pode influenciar a captura se a obtenção da informação sobre as cores do objeto for um dos objetivos pretendidos. Se o emissor utilizar padrões de luz do espetro visível, estas podem adulterar a aquisição dessa informação. Uma forma de contornar esses casos pode passar por usar padrões do espetro infravermelho, tal como é feito com as câmaras de profundidade como a Kinect. Uma vez que os sistemas de luz estruturada são sistemas ativos e necessitam também de um sistema de projeção, o seu preço costuma ser também mais elevado.

Das diferentes metodologias analisadas, aquela que mais se adapta às características do sistema proposto é a de Luz Estruturada devido ao detalhe das capturas e ao tempo necessário para as efetuar. Dos vários sistemas existentes que utilizam esta tecnologia, o melhor compromisso encontrado entre qualidade, performance e custo foi nas câmaras de profundidade. Estes sensores têm como principal vantagem o acesso direto à informação 3D com boa qualidade e de forma rápida. Além disso, o baixo custo e a portabilidade são também fatores que os valorizam. Em contrapartida, os sensores de profundidade apresentam desvantagens como o alcance limitado, tanto a distâncias muito curtas (0-50cm) como longas (superiores a 5 metros), e dificuldade na aquisição de informação em condições menos favoráveis como a exposição direta de luz solar nas superfícies ou a aquisição de materiais brilhantes/refletores ou transparentes.

Dentro desta gama de sensores, foram estudadas as características de quatro dispositivos presentes no mercado: a Kinect, a Xtion PRO Live e as Carmine 1.09 e 1.08, como se pode ver no seguinte quadro.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Kinect/Kinect4Windows | Xtion PRO LIVE | Carmine 1.08 | Carmine 1.09 |
| Distance min (m) | 0.4m/0,8m\* | 0,8m\* | 0.8m | 0.35m |
| Distance max (m) | 3,5m\*\* | 3,5m\*\* | 3.5m | 1.40m |
| FOV vertical | 43º | 45º | 45º | 45º |
| FOV horizontal | 57º | 58º | 57.5º | 57.5º |
| Depth Size | 640x480 (30FPS) | 640x480 (30FPS) | 640x480 (60FPS) | 640x480 (60FPS) |
| Color Size | 1280x960 (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 50cm 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

Todas elas apresentam características muito semelhantes em relação às suas especificações e custos de aquisição. Apenas a Carmine 1.09 se diferencia um pouco uma vez que está orientada apenas para aquisições de curto alcance. Entre as quatro camaras, decidiu-se utilizar a Kinect uma vez que é o dispositivo mais popular e com mais suporte por parte da comunidade de desenvolvedores/programadores. Das duas soluções que a Microsoft apresenta, a escolha recairia pela *Kinect* para a *Xbox* uma vez que é mais económica e as características são muito semelhantes. No entanto, e para os testes realizados, o sensor utilizado foi a *Kinect for Windows*, cedida pelo CCG[[56]](#footnote-56), tendo como única vantagem a maior precisão para capturas efetuadas a distâncias inferiores a 80cm da camara.

Setup (Aquisição 360º)

A captura 360º de informação pressupõe que esta seja adquirida de várias perspetivas. Uma das formas de o fazer é através da movimentação do sensor à volta do objeto ou pela movimentação do objeto em frente do sensor, tal como é utilizado nos sistemas da Go!SCAN 3D[[57]](#footnote-57) ou no Digitalizer[[58]](#footnote-58) respetivamente. No entanto, devido à característica pretendida de tempo real, a captura faseada destas diferentes perspetivas não é viável. Outra das formas de efetuar este tipo de captura é através da utilização de várias fontes de captura. Esta abordagem é usada em vários sistemas como o da IR-Entertainment[[59]](#footnote-59) ou no sistema referido em Y[[60]](#footnote-60), no entanto, tem como desvantagens um maior custo em *hardware* e também o aumento da carga computacional.

No caso específico da utilização de várias *Kinects*, além dessas questões existe ainda outro problema relacionado com a interferência causada entre estas. Quando, por exemplo, duas *Kinects* estão direcionadas para uma determinada zona existem áreas de sobreposição, isto é, zonas cuja aquisição é feita por ambos os sensores. Nestas regiões há a formação de ruído devido à sobreposição dos padrões de IR que impossibilitam a correta estimativa da profundidade dessas áreas (FIGURA XX). Nos teste realizados verificou-se ainda que existe mais ruído quando as câmaras se encontram em posições frontais.

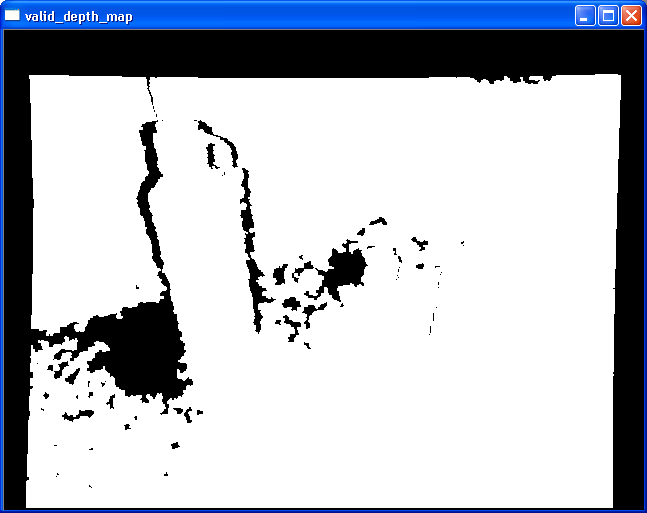
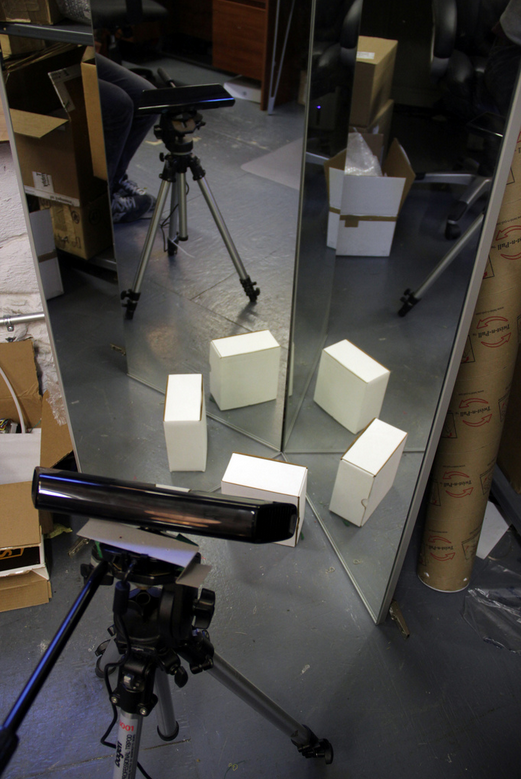
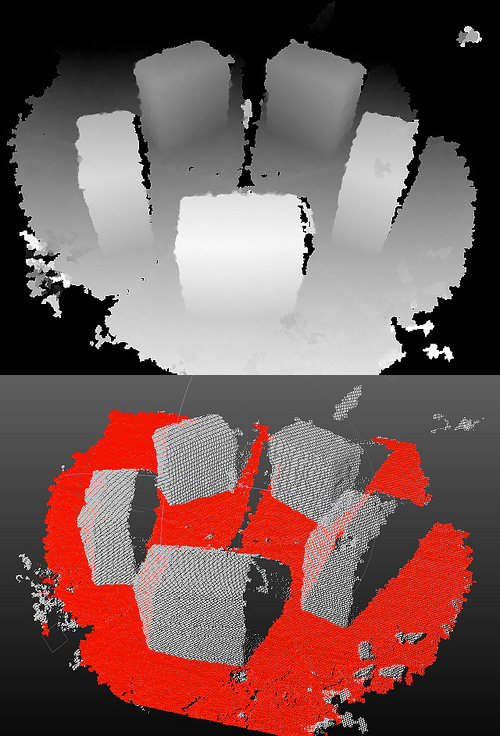


FIGURA X COM RUIDO DE DUAS KINECTS

Outra forma de se conseguir capturar informação de várias perspetivas de um objetos é com da utilização de espelhos. Devido às suas propriedades refletoras, este material permite a visualização de informação de diferentes perspetivas a partir de um único ponto de vista. A utilização de espelhos para a realização de captura em 3D já foi concretizada noutros sistemas de aquisição como em ArtigoX[[61]](#footnote-61) ou em ArtigoY[[62]](#footnote-62). A colocação do espelho em frente ao sensor e por trás do objeto permite que seja capturada informação de várias perspetivas com apenas um sensor. Dependendo do número de espelhos utilizados e da posição relativa entre eles é ainda possível aumentar o número de perspetivas diferentes, tal como foi demonstrado em ArtigoW[[63]](#footnote-63).

No entanto, a utilização de espelhos com a *Kinect* não se encontra documentada pelo que a única referência encontrada até à data da escrita desta dissertação encontra-se no livro Making Things See [referência]. Aqui é mencionado a utilização de espelhos pelo artista e investigador Kyle McDonald como um exemplo da versatilidade da Kinect (FIGURA XX) e para mostrar como é que esta se comporta na presença de espelhos. No entanto, não foram encontradas evidências de este trabalho ter sido continuado.



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

De acordo com as características necessárias para a construção do sistema e as características das soluções analisadas, o melhor compromisso encontrado foi a utilização de um sistema de espelhos. O posicionamento destes materiais na periferia da cena permite recolher informação sobre o objeto de outras perspetivas como se tratassem de pontos de aquisição virtuais. Estes são capturados também pela câmara, assim como uma visão direta do objeto, maximizando desta forma a quantidade de informação recolhida de uma só vez. Apesar de este processo continuar a exigir uma carga computacional acrescida a nível de análise de imagem, esta carga é igual ou menor à existente no caso de múltiplas camaras. Além disso, o custo global do sistema diminui uma vez que os espelhos são menos dispendiosos que os sensores: um espelho de X por X consegue ser adquirido por Y€.

Casos de Uso

Um sistema de captura de informação em 3D consegue, de forma mais ou menos eficiente, adquirir a informação geométrica de objetos ou outras entidades no formato de uma nuvem de pontos. Esta nuvem representa a entidade capturada e pode facilmente ser visualizada através de software com o MeshLab[[64]](#footnote-64). Uma vez criadas mais que uma nuvem de pontos, estas podem também ser unidas de forma a criar uma representação única do objeto, com mais qualidade. O sistema proposto tem como objetivo a aquisição a 360º da informação geométrica de um objeto (ou outras entidades) com taxas de atualização interativas, como tal, será capaz de capturar e unir a informação proveniente das nuvens de pontos das várias perspetivas em simultâneo.

Estas características fazem com que seja possível fazer a aquisição 360 da geometria de um modelo em tempo real o que abre portas a outro tipo de aplicações para além da simples captura de modelos. No entanto e devido às restrições dos materiais e setup usados, o sistema tem várias limitações: a área de ação é limitada pelo alcance da Kinect e pelo posicionamento dos espelhos e a qualidade da aquisição é também influenciada pela distância a que estes se encontram do objeto a capturar.

Modelos 3D de Objetos

Como já foi explicado anteriormente, a criação de modelos 3D de objetos permite gerar uma representação digital da geometria de um dado objeto físico e também capturar outras propriedades como a cor ou texturas. Além da nuvem de pontos, para este fim é também útil a geração de *mesh* do objeto, isto é, criar as relações entre os pontos da nuvem de forma a gerar uma malha poligonal representativa do objeto. Existem vários métodos para atingir este objetivo, no entanto, este processo é por norma computacionalmente exigente e como tal, difícil de atingir em tempo real.

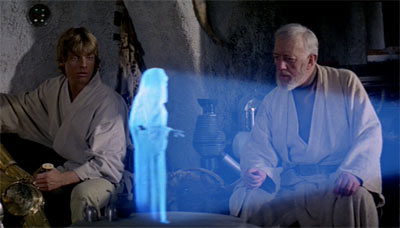
Utilizando o sistema proposto, a geração de modelos em 3D pode ser conseguida de forma simples. A nuvem de pontos de um determinado objeto pode ser criada a partir de uma única captura, isto é, a partir da informação de uma única frame, no entanto esta contém por norma algum ruído proveniente da Kinect. Uma vez que os modelos 3D exigem uma qualidade elevada, uma forma de contornar este problema pode passar por uma captura prolongada através da utilização de várias *frames* gerando assim a nuvem de pontos do objeto de forma incremental. Desta forma, esta exposição prolongada do objeto ao processo de captura permite que seja recolhida mais informação e desta forma consegue-se (através de software) minimizar o ruido inerente à Kinect. A geração de malha poligonal pode ser feita posteriormente, depois de capturada a informação da nuvem de pontos.

Estes modelos podem ser usados para vários fins pelo que a aplicação mais direta é a integração dos mesmos em mundos virtuais. Áreas como os jogos de vídeo, cinema e realidade aumentada podem beneficiar deste sistema através da integração de modelos de objetos do dia-a-dia nos seus cenários. Outra aplicação que pode ter interesse inclusive a nível doméstico é a criação de modelos de objetos tendo como objetivo a sua replicação através de impressoras 3D. Esta vertente revela-se especialmente útil no que toca à substituição de peças ou componentes danificados ou perdidos.

Vídeo 3D Real e Holografia

A geração 360 de informação 3D em tempo real faz com que seja possível visualizar a cena a partir de qualquer ponto de vista circundante. A gravação desta informação possibilita que se consiga obter vídeo em 3D real independente da perspetiva de aquisição. Isto permite também que possam ser geradas representações 3D de objetos animados ou mesmo seres vivos em movimento em tempo real.

As prioridades de uma aquisição em movimento e em tempo real passam por uma elevada taxa de atualização em detrimento da qualidade da captura. Neste caso, uma vez que a informação se encontra em constante atualização, o ruído originado pela Kinect durante a captura é pouco percetível e como tal será mais tolerável à visão humana. Uma forma de contornar este problema é a aproximação da nuvem de pontos a modelos predefinidos das entidades a capturar. Por exemplo, se se estiver a realizar a aquisição do modelo de uma pessoa, esta pode ser aproximada ao modelo de um esqueleto humano de forma a reduzir o ruido e tentar colmatar possíveis falhas no processo da captura.

Na vertente de vídeo, este tipo de aquisição pode dar origem a vídeos em 3D real, independentes do ponto de vista, e que por essa mesma razão permitem a navegação no espaço durante uma reprodução de uma gravação. Esta característica pode ser interessante para novas abordagens na área de storytelling uma vez que pode permitir que se desenrolem de várias histórias ao mesmo tempo mas em espaços diferentes. Já na vertente de aquisição e visualização tempo real e aliado a uma ligação de dados de alta velocidade, esta tecnologia pode permitir a comunicação entre pessoas em vídeo 3D real. Partindo deste mesmo princípio mas aplicando-o a uma área mais ambiciosa, a partir de um sistema deste género estaria também disponível a informação necessária para criação de hologramas e a comunicação entre pessoas através dos mesmos (Figura XX).

<http://www.trickedbythelight.com/tbtl/images/HologramStarwars.jpg>

Análise interativa de Modelos

A geração de modelos 3D de objetos a partir de uma nuvem de pontos é útil para a integração dos mesmos em diversas áreas. No entanto, outra forma de utilizar a informação gerada prende-se com a análise e extração de características sobre essa nuvem de pontos que permita inferir outras informações que não apenas a sua estrutura geométrica. Exemplos disso são a segmentação de informação para, por exemplo, separação de diferentes componentes ou a extração do esqueleto das estruturas das nuvens de pontos cruas, isto é, o conjunto de linhas-guia das partes mais representativas dessa nuvem.

O sistema proposto consegue gerar nuvens de pontos de um determinado objeto ou entidade em tempo real. Para gerar informações como o esqueleto da nuvem e uma vez que elas se tratam de aproximações, a extração de nuvens pode ser feita diretamente sem que para tal seja necessário processamento prévio dos dados da aquisição. Desta forma é possível analisar e gerar informação adicional de forma interativa e aplica-la assim a vários casos de uso.

Este tipo de informação pode ser bastante útil em sistemas de realidade aumentada. Imaginando o caso de um provador de roupa virtual, a extração deste esqueleto e de outras propriedades como a altura da pessoa e a espessura dos seus membros permite com que seja possível produzir uma experiência de realidade aumentada mais rica. Noutra área, mais ligada á realidade virtual, a geração deste esqueleto também pode ser usado para a animação de personagens e como a aquisição é 360º tem mais resistência às oclusões devido às movimentações da pessoa. A segmentação também pode ser utilizada em sistemas de AR de forma a separar os objetos uns dos outros e estes do cenário. Esta compreensão do cenário faz com que a sobreposição de informação seja mais fácil de ser conseguida.

Sumário

Os objetivos desta tese passam por criar um sistema de baixo custo capaz de fazer a aquisição 360º de informação 3D de um objeto em tempo real. Estas características nem sempre são compatíveis e como tal foram analisados os melhores métodos e tecnologias para o fazer e tiveram que ser tomadas decisões em duas áreas principais: o sensor responsável pela captura de informação e o setup usado para realizar a captura 360.

Em relação à captura foi decidido utilizar-se o sensor da Microsoft, a Kinect. Este dispositivo tem um custo relativamente reduzido e tem como principais vantagens o acesso rápido e de boa qualidade a informação de profundidade da cena. Quanto ao setup utilizado, a escolha recaiu pela utilização de uma única camara e de espelhos para capturar as diferentes perspetivas. Desta forma o custo do sistema é o mais reduzido (em comparação com as outras alternativas) e é feita uma maximização da utilização dos recursos disponíveis.

Foram ainda mostradas algumas das possíveis aplicações deste sistema que recaem em áreas como a construção de modelos 3D, geração de vídeo em 3D real e transmissão do mesmo em tempo real e a análise e geração de informação adicional aos modelos, como a segmentação da cena e extração de esqueletos, de forma interativa.

Implementação

Neste capítulo começaremos por mostrar como é que o sistema foi construído.

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

Conceção do sistema

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.

Características do Sistema

Coisas

Kinect

Informação detalhada da 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>

<http://eng.au.dk/fileadmin/DJF/ENG/PDF-filer/Tekniske_rapporter/Technical_Report_ECE-TR-6-samlet.pdf> - pag 3

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

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).

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]

Arquitetura Geral do Sistema

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

Casos de Estudo

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 objeto e dos espelhos em relação ao objeto e à camara.

Foram testadas duas abordagens diferentes á arquitetura do sistema

Cenário 1 - 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.

Fluxo de Execução

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

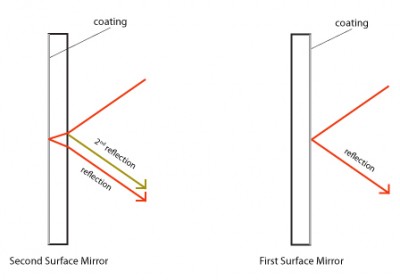
*MeshLab* pode ser usado para modelos fixos e o “lixo” pode ser retirado manualmente.

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

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. <http://www.vision3d.com/stereo.html>

   <http://www.dashwood3d.com/blog/beginners-guide-to-shooting-stereoscopic-3d/> [↑](#footnote-ref-6)
7. Depth Map and 3D Imaging Applications - Algorithms and Technologies (2012) [↑](#footnote-ref-7)
8. <http://reviews.cnet.co.uk/portable-gaming/nintendo-3ds-review-50000079/>

   <http://www.olivieris.toile-libre.org/index.php?pg=18&id=18>

   <http://techcrunch.com/2009/07/22/the-fujifilm-finepix-real-3d-w1-becomes-official-includes-3d-movie-mode/> [↑](#footnote-ref-8)
9. 3D Scanner, State of the Art [pdf] [↑](#footnote-ref-9)
10. Time-of-Flight Cameras and Microsoft Kinect [book][pdf] [↑](#footnote-ref-10)
11. <http://www.faro.com/en-us/products/metrology/faro-3d-imager/overview> [↑](#footnote-ref-11)
12. <http://www.riegl.com/nc/products/terrestrial-scanning/produktdetail/product/scanner/33/> [↑](#footnote-ref-12)
13. <http://www.pmdtec.com/> [↑](#footnote-ref-13)
14. <http://www.mesa-imaging.ch/> [↑](#footnote-ref-14)
15. <http://click.intel.com/intelsdk/Default.aspx> [↑](#footnote-ref-15)
16. Rainbow three-dimensional camera: new concept of high-speed three-dimensional vision systems [pdf] [↑](#footnote-ref-16)
17. Color-encoded structured light for rapid active ranging [pdf] [↑](#footnote-ref-17)
18. Structured light stereoscopic imaging with dynamic pseudo-random patterns [↑](#footnote-ref-18)
19. Structured-light 3D surface imaging: a tutorial [pdf] [↑](#footnote-ref-19)
20. http://www.4ddynamics.com/3ddynamics/ [↑](#footnote-ref-20)
21. http://www.3d-shape.com/home/home\_d.php [↑](#footnote-ref-21)
22. http://www.creaform3d.com/pt/solucoes-para-area-de-saude [↑](#footnote-ref-22)
23. http://www.maestro3d.com/ [↑](#footnote-ref-23)
24. http://www.cynoprod.com/ [↑](#footnote-ref-24)
25. <http://www.flashscan3d.com/> [↑](#footnote-ref-25)
26. http://www.vitronic.de/en [↑](#footnote-ref-26)
27. http://www.optimet.com/ [↑](#footnote-ref-27)
28. <http://msdn.microsoft.com/en-us/library/jj131033.aspx>

    <http://www.engadget.com/2010/06/13/microsoft-kinect-gets-official/> [↑](#footnote-ref-28)
29. Time-of-Flight Cameras and Microsoft Kinect [Book][pdf] [↑](#footnote-ref-29)
30. <http://www.asus.com/Multimedia/Xtion_PRO_LIVE/#specifications> [↑](#footnote-ref-30)
31. <http://www.primesense.com/get-your-sensor2/> [↑](#footnote-ref-31)
32. http://www.guinnessworldrecords.com/records-9000/fastest-selling-gaming-peripheral/ [↑](#footnote-ref-32)
33. http://www.adafruit.com/blog/2010/11/04/the-open-kinect-project-the-ok-prize-get-1000-bounty-for-kinect-for-xbox-360-open-source-drivers/ [↑](#footnote-ref-33)
34. <http://thenextweb.com/microsoft/2013/02/20/microsofts-kinect-2-0-specifications-allegedly-leak-1920x1080-color-stream-60ms-latency-and-usb-3-0/> [↑](#footnote-ref-34)
35. <http://www.space-vision.jp/EP-Body_Scanner.html> [↑](#footnote-ref-35)
36. <http://www.4ddynamics.com/3d-scanners/bodyscanner/> [↑](#footnote-ref-36)
37. <http://ir-ltd.net/> [↑](#footnote-ref-37)
38. <http://tc2.com/index_3dbodyscan.html>

    <http://www.tc2.com/pdf/kx16.pdf> [↑](#footnote-ref-38)
39. <http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XXXIX-B3/531/2012/isprsarchives-XXXIX-B3-531-2012.pdf>

    NATURAL USER INTERFACE SENSORS FOR HUMAN BODY MEASUREMENT (2012) [pdf]] [↑](#footnote-ref-39)
40. <http://www.iti.gr/iti/files/document/publications/manuscript6camready_.pdf> [↑](#footnote-ref-40)
41. 3D Face Reconstruction from a Single Camera Using a Multi-Mirror Set-up [pdf] [↑](#footnote-ref-41)
42. Surround Structured Lighting for Full Object Scanning [pdf] [↑](#footnote-ref-42)
43. http://store.makerbot.com/digitizer.html [↑](#footnote-ref-43)
44. http://www.matterform.net/scanner [↑](#footnote-ref-44)
45. http://www.goscan3d.com/ [↑](#footnote-ref-45)
46. http://www.3dscanco.com/ [↑](#footnote-ref-46)
47. <http://www.kscan3d.com/> [↑](#footnote-ref-47)
48. http://reconstructme.net/ [↑](#footnote-ref-48)
49. 3D Body Scanning in a Mirror Cabinet [pdf] [↑](#footnote-ref-49)
50. <https://graphics.stanford.edu/papers/dmich-sig00/dmich-sig00-nogamma-comp-high.pdf> [↑](#footnote-ref-50)
51. <http://www.4ddynamics.com/3d-scanners/bodyscanner/> [↑](#footnote-ref-51)
52. <http://ir-ltd.net/> [↑](#footnote-ref-52)
53. <http://www.kscan3d.com/> [↑](#footnote-ref-53)
54. <http://store.makerbot.com/digitizer.html> [↑](#footnote-ref-54)
55. <http://www.matterform.net/scanner> [↑](#footnote-ref-55)
56. Centro de Computação Gráfica (www.ccg.pt) [↑](#footnote-ref-56)
57. http://www.goscan3d.com/ [↑](#footnote-ref-57)
58. http://store.makerbot.com/digitizer.html [↑](#footnote-ref-58)
59. <http://ir-ltd.net/> [↑](#footnote-ref-59)
60. <http://www.iti.gr/iti/files/document/publications/manuscript6camready_.pdf> [↑](#footnote-ref-60)
61. Multiple-view 3D Reconstruction Using a Mirror [↑](#footnote-ref-61)
62. 3D Body Scanning in a Mirror Cabinet [pdf] [↑](#footnote-ref-62)
63. Surround Structured Lighting for Full Object Scanning [pdf] [↑](#footnote-ref-63)
64. <http://meshlab.sourceforge.net/> [↑](#footnote-ref-64)