Título

Kinectulum

3D Object reconstruction using Kinect and mirrors

3D Object reconstruction using a single Kinect and mirrors

3D Image Acquisition using a Static Setup

3D Information Acquisition of Small Daily Objects

3D Full Data Acquisition a Static Setup

3D Panoramic Data Acquisition using a Static Setup

3D Concentric Data Acquisition using a Static Setup

3D 360 Data Acquisition using a Static Setup

Head

Thanks

You’re welcome.

Abstract

Things

What did you do?

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

How did you do it? State methods.

What did you learn? State major results.

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

Resumo

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

­Introdução

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

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

???

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

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

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

Motivação

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

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

Desta forma e juntando o melhor dos dois *setups*, seria um desafio conceber um sistema capaz de fazer uma aquisição 360º em tempo real utilizando apenas uma câmara e um *setup* estático. Isto permitiria a aquisição de mais informação em menos tempo e de uma forma económica. Um sistema deste género poderia permitir a geração de vídeo em 3D real (não apenas com a noção de profundidade estereoscópica) que posteriormente, no prisma de um espectador, possibilitaria a visualização desse vídeo de forma dinâmica em várias perspetivas. A geração de dados com estas características poderá também ser usada mais tarde em sistemas holográficos uma vez que existe informação suficiente para criar uma vista concêntrica do objeto em foco.

Objetivos

Neste contexto, pretende-se com esta dissertação desenvolver um sistema de baixo custo capaz de realizar a captura 360º de informação 3D em tempo real de entidades de pequenas/médias dimensões a partir de um *setup* estático e de apenas uma câmara. Inicialmente será necessário compreender o processo de captura de informação 3D e como a partir dessa informação poderemos gerar o objeto em questão.

A captura 360º de informação 3D de um objeto requere a aquisição de informação de várias perspetivas. Para o conseguir fazer de forma instantânea (todas as perspetivas ao mesmo tempo) é comum usar-se vários sensores/câmaras, no entanto, além do aumento da carga computacional, isso também envolve um maior custo em *hardware.* Como tal, um dos desafios será conceber a arquitetura de um sistema que consiga obter a informação de todas essas perspetivas em simultâneo e de forma eficiente utilizando apenas um sensor.

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

Em suma, os objetivos para esta dissertação são:

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

Estrutura do Documento

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

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

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

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

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

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

3D Scan – Related Work

Os primeiros sistemas de captura de informação em 3D remontam à década de 1960 e estes usavam luzes, câmaras e projetores para realizar a tarefa. Era um processo moroso e que exigia muito esforço e tempo para conseguir ter resultados satisfatórios. Durante vários anos esta tecnologia não sofreu grandes desenvolvimentos e tal pode 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, atualmente também já existem sistemas que permitem fazer esse tipo de aquisições em casa. Produtos como o matterform[[1]](#footnote-1) e o digitalizer[[2]](#footnote-2) permitem capturar o modelo 3D de um objeto 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 compartimento completo. A vertente móvel deste tipo de sensores tem-se tornado cada vez mais apelativa. Soluções como o Capri da primesense[[3]](#footnote-3) ou projetos inovadores como o Structure Sensor[[4]](#footnote-4) ou o CADScan[[5]](#footnote-5) poderão levar à proliferação destes sensores em ambiente de mobilidade o que poderá levar a avanços significativos no campo da Realidade Aumentada.

Neste capítulo serão descritos diferentes métodos de captura de informação 3D, focando a análise nos processos óticos **de Estereoscopia, Time-of-Flight** **e Luz Estruturada** uma 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 destes sistemas.

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 neste caso. Por outro lado, se se tiver um sistema de videoconferência, o tempo de aquisição terá que ser mínimo enquanto a qualidade dos modelos passa a ser secundária e pode ser até aproximada a modelos faciais 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.

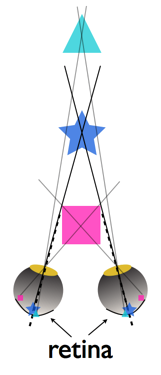
O caso das tecnologias óticas requere um maior detalhe uma vez que são as mais comuns e as que mais se adaptam ao sistemas proposto para esta dissertação. Usam radiação no espectro visível (100nm – 300um) e por norma tecnologias de baixo-custo que permitem realizar a captura de informação com boa qualidade, 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.

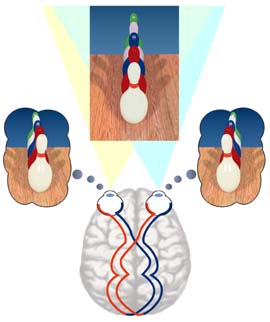
Dentro das tecnologias óticas, estas podem ser divididas em passivas ou ativas. Os sistemas passivos não emitem qualquer tipo de radiação e usam apenas aquela que é refletida pelo ambiente. Estas técnicas utilizam *setups* óticos e/ou perspetivas diferentes que depois são processadas por computador para 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 inferir 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*.* 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 fissão 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 na ordem dos 40fps. Estas características fazem com que estes 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 com estes sensores 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 e são refletivos, óticos e ativos. Este tipo de sistemas normalmente é composto por uma componente emissora e um ou mais sensores de captura. O emissor tem como função projetar um padrão de luz na superfície a digitalizar que o sensor captura para depois ser analisado e gerada a informação em 3 dimensões. Para realizar este processo é utilizado um princípio similar àquele que é usado nos métodos estereoscópicos, a triangulação ativa, onde a perspetiva da segunda camara é substituída por uma entidade emissora. O padrão projetado na cena é depois capturado e analisado juntamente com os dados recolhidos no processo de calibração e a partir das deslocações encontradas são calculados os pontos 3D do mundo.

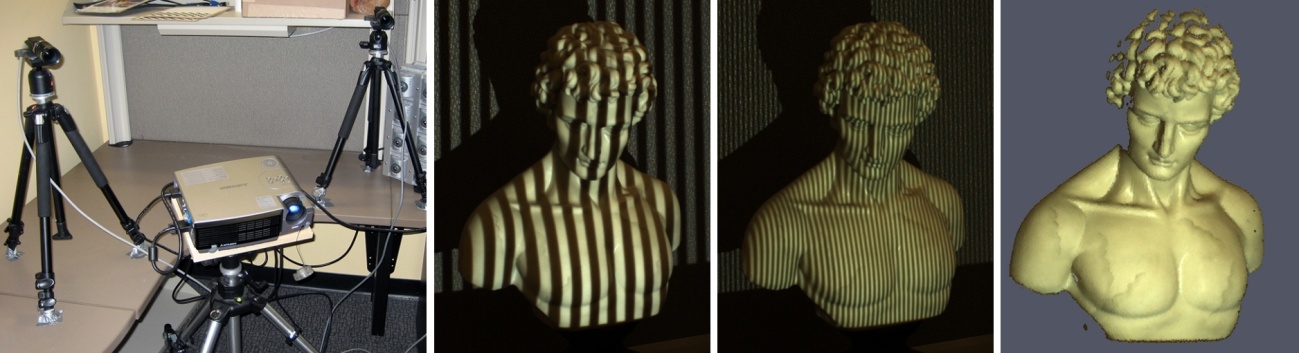
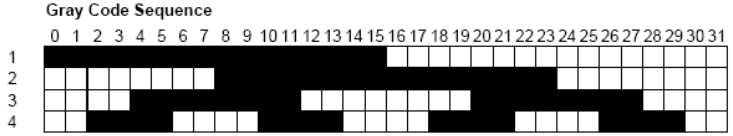


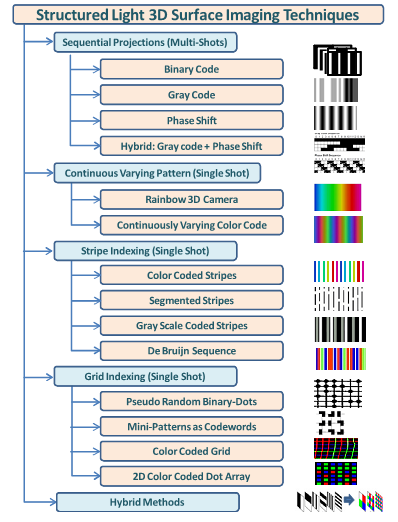
Imagem: http://mesh.brown.edu/3dpgp-2009/homework/hw2/hw2.html

Existem várias técnicas de luz estruturada que essencialmente diferem no tipo de padrão projetado e consequentemente influenciam o tempo e a qualidade da captura. Uma das técnicas usadas é o Gray Code, uma sequência 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 mais 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 fazendo 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 frame, conferindo-lhes a capacidade 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 captura o que normalmente se traduz em áreas de ação pequenas. Consequentemente, 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. Este problema em específico pode ser ultrapassado usando padrões de luz sem ser do espetro visível, como é feito com a Kinect, que será explicado em 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 do corpo-humano ou apenas da face. Isto pode ter aplicações diversas em áreas como o cinema e entretenimento e também para 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 enquanto a FlashScan3D, mais ligada às áreas das ciências forenses, construiu um scanner de grande precisão para a captura de impressões digitais. Noutras áreas, como a indústria, estes sistemas também são usados por empresas especializadas como a Vitronic[[25]](#footnote-25) ou a Optimet[[26]](#footnote-26) 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.

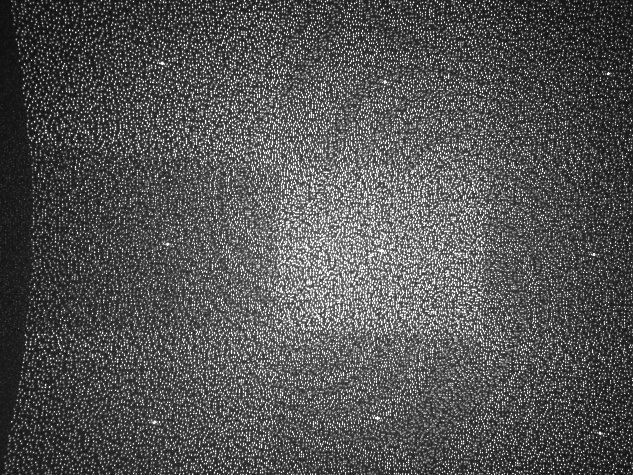
Microsoft Kinect

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[[27]](#footnote-27)

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 uma câmara 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. 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[[28]](#footnote-28)



Apesar de não ser o único sensor RGBD com este tipo de características presente no mercado (existe também a Asus Xtion Pro Live[[29]](#footnote-29) e a PrimeSence Carmine 1.09[[30]](#footnote-30)), 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 se utilizaria 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[[31]](#footnote-31). 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 humanos na cena e inferir os seus esqueletos, com 48 pontos de controlo, em tempo real e 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.

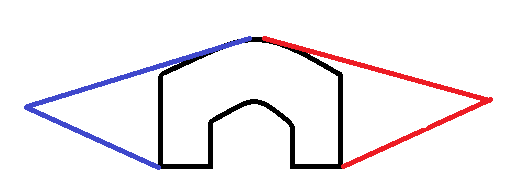
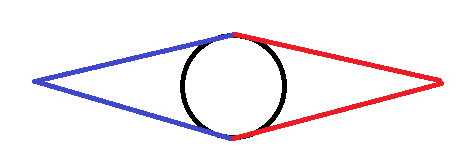
Quando saiu, 4 de novembro de 2010, a Kinect foi lançada enquanto acessório de uma consola de jogos e não foi disponibilizado nenhum software para desenvolvimento. Nessa mesma data, de forma a contornar essa situação, a empresa Adafruit Industries ofereceu uma recompensa[[32]](#footnote-32) para quem desenvolvesse um driver open-source capaz de aceder à informação do sensor. Seis dias mais tarde e depois da recompensa ter triplicado para os 3000$, foi libertada lançada a primeira versão do *libfreenect*, um driver opensource 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 opensource e 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 o seu SDK não comercial.

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 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[[33]](#footnote-33) 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. (?)

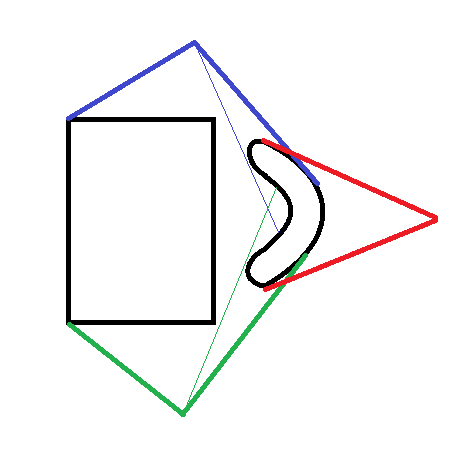
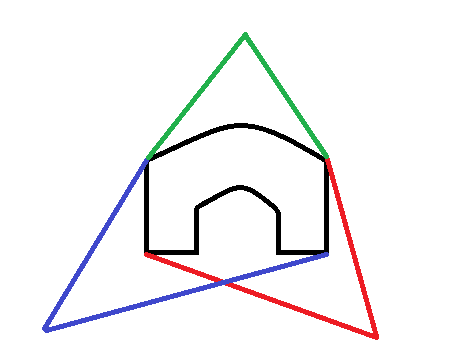
Sistemas de Aquisição Concêntrica/360

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 são possíveis de capturar e este problema apenas poderia 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 desta captura. Esta limitação requere a introdução de múltiplas fontes de captura que, por norma, implicam a aquisição de mais camaras o que, além de introduzir mais complexidade e carga computacional, tornam o seu custo do sistema mais elevado. Por outro lado, com um setup estático tem-se como principal vantagem 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[[34]](#footnote-34) tem um produto de nome Cartesia Series Portable 3D Body Scanner que afirma se o primeiro sistema de scan corporal portátil do mundo. É constituído por nove camaras distribuídas igualmente por três torres e usa laser scanning realizar a aquisição de informação. A captura demora 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[[35]](#footnote-35) tem um sistema para fazer a captura 3D do corpo de uma pessoa e 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. 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 os resultantes do batimento cardíaco ou da respiração. 

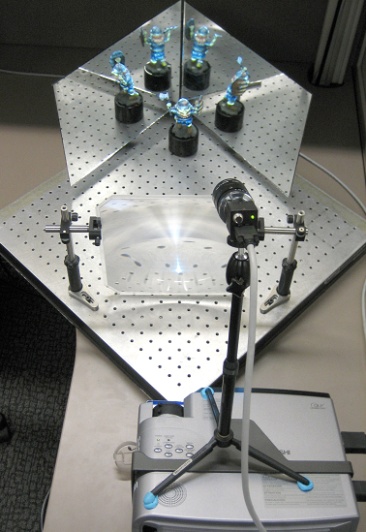
A IR-Entertainment[[36]](#footnote-36) utiliza um sistema em forma de anel com câmaras Canon DSLR de 18mp (FiguraXX). A geração dos modelos é conseguida através de técnicas de fotometria e luz estruturadas dando origem a modelos de qualidade muito elevada. O número de camaras utilizadas depende do fim 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[[37]](#footnote-37) tem também um sistema para captura 3D do corpo mas esta recorre à utilização de Kinects para o fazer. São utilizados 16 sensores de profundidade espalhados igualmente por quatro colunas. A aquisição demora cerca de 7segundos a ser feita e consegue atingir uma precisão de 3mm. Este sistema foi construído a pensar principalmente no mercado têxtil podendo ser usada como cabine virtual para a prova de roupa virtual 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 [[38]](#footnote-38)foram usadas técnicas muito similares às usadas pela TC2 e foram usadas 8 Asus Xtion Pro Live para realizar a captura espalhadas por 4 barras verticais. Os resultados mostraram um erro máximo inferior a 20mm. Em ArtigoY[[39]](#footnote-39) 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 a parte superior da pessoa, e as outras 4 em forma de um 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[[40]](#footnote-40), foi desenvolvido um sistema capaz de realizar a captura facial usando apenas uma camara e dois espelhos de forma a gerar mais perspetivas do indivíduo a capturar. 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[[41]](#footnote-41), 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 forma fornecidos dados sobre o tempo específico da captura nem da qualidade da mesma.

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 em frente da camara. A junção das nuvens de pontos das várias perspetivas pode ser feito após as capturas ou, como em X e Y, pode ser feita à medida que fé feita a aquisição através de SLAM (Simultaneous Location and Mapping) Ref.

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 em Setembro de 2013 o Digitalizer[[42]](#footnote-42), 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. A Matterform[[43]](#footnote-43) tem também um scanner 3D muito similar ao da Makerbot 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.





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



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

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

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

Stand-Off Distance 400mm

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

uses a regular flash bulb instead of a laser

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

Working distance 0.4 – 1 m

Kinect:

Dois exemplos: qualidade e limitações são as da Kinect.

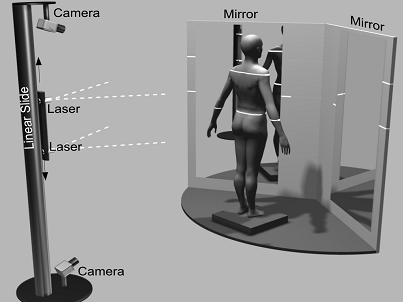
<http://www.kscan3d.com/>

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

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

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

Já no ArtigoY [[44]](#footnote-44)foi utilizado um sistema de 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 foi 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.

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

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

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

“Who can benefict from 3D Scannnig”

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

Entretenimento e Consumo

Modelos 3d para e jogos

Filmes 3d, free-viewpoint vídeo (3D TV),

Impressoras 3D / Replicação de objetos domésticos. “3-D scanning and printing is now where regular printing on paper was 20-25 years ago. Speed and capability is going to go up and the cost is going to rapidly decline.”

Modelo 3D do corpo humano para moda / Virtual Try On

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

Preservação Histórica

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

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

Multi-Sensors and Multiray Reconstruction for Digital Perservation [pdf]

Michael Angelo project <https://graphics.stanford.edu/papers/dmich-sig00/>

Indústria

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

Determine whether manufactured parts are within tolerances

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

Use physical objects to conceptualize the idea; this is typically done by an industrial designer in clay lead design or foam lead design. 3D scanning can also be applied at the idea concept phase by digitizing objects, then using them as renderings in concept illustrations.

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

Robótica

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

Medicina

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

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

3D imaging for bite mark analysis [pdf]

Utilização dos modelos 3D para treinar cirurgias.

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

­­­­­­­Summary

Things

Concept Design (?)

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

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

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

Descrição e Funcionalidades (?)

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

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

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

X – <http://ir-ltd.net/>

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

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

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

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

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

Decisões

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

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

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

Captura

Vários dos métodos referidos no capítulo anterior foram testados ….

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

Utilização de câmaras + estereoscopia ? descartar

…

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

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

QUADRO COM CARACTERÍSTICAS DA KINECT E OUTRAS CAMARAS

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

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

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

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

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

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

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

Apesar das câmaras X, Y e a Kinect serem bastante similares em termos de características e preços, decidiu-se utilizar a Kinect como dispositivo de captura para este projeto devido à sua popularidade em ambientes de desenvolvimento. Entre as duas Kinects apresentadas, a escolha recairia pela Kinect base uma vez que é mais económica. No entanto e para os testes realizados foi usada uma Kinect for Windows cedida pelo CCG[[45]](#footnote-45).

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

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

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

Foram ainda realizados testes com a utilização de duas Kinects em simultâneo (descrito no setup?)

Setup (Aquisição Concêntrica)

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

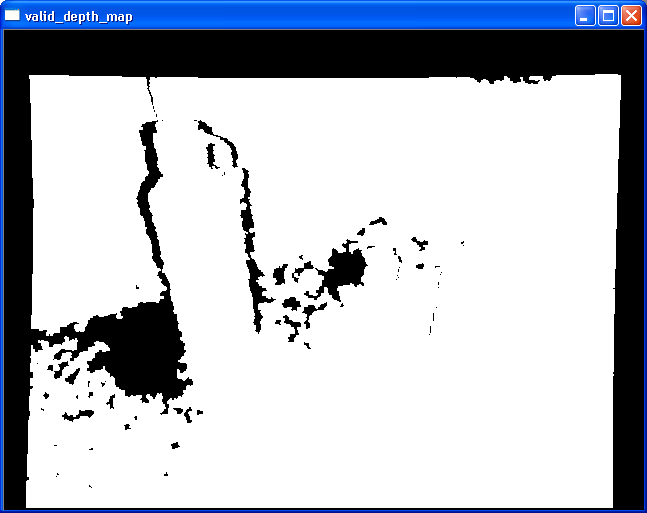


FIGURA X COM RUIDO DE DUAS KINECTS

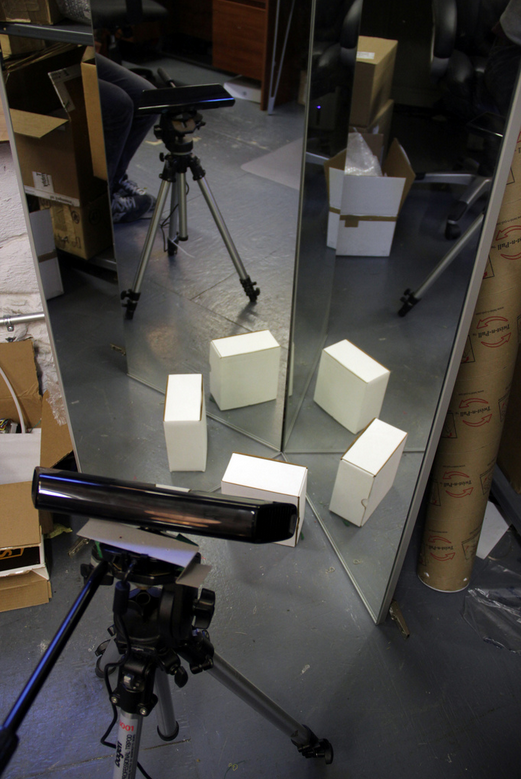
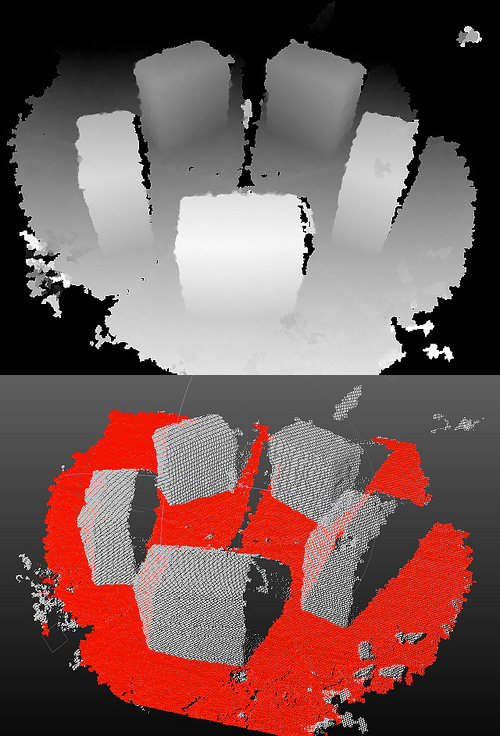
Verificou-se ainda que existe mais ruído quando as câmaras se encontram frente a frente.(?)

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

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

Surround Structured Lighting for Full Object Scanning [pdf]

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



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

Mirror Equations

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

Equation

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

Exploiting Mirrors in Interactive Reconstruction with Structured Light [pdf]

Especificações

Decisão

System Overview

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

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

FIGURA COM DOIS SETUPS: ARENA + SAPATOS

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

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

Kinect

Informação detalhada da Kinect (?)

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>

Espelhos

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

Equação de reflecção. (?)

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

Casos de Uso

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

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

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

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

Geração e transmissão de 3D em tempo real.

Holografia

Caso de uso 1

Caso de uso 2

Implementação

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

Arquitetura do Sistema

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

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

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

Foram testadas duas abordagens diferentes á arquitetura do sistema

Top view + 4 espelhos – ARENA

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

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

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

Perspetiva + 2 espelhos

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

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

System pipeline

O fluxo de execução (?) do sistema segue

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

Fase0

* Setup chão
* Setup espelhos

Fase1

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

Fase 2

* Remoção de outliers?
* Criação de mesh



Setup

*Setup “Manual” referred to Appendix?*

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

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

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

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

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

IMAGEM DEMONSTRATIVA [Imagem com selecção dos pontos do chão]

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

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

Aquisição da Nuvem de Pontos

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

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

IMAGEM DE EXEMPLO DAS MASCARAS EM 2D

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

IMAGEM 3D DA CENA

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

IMAGEM 3D DA CENA COM INFORMAÇÃO DOS ESPELHOS JÁ INVERTIDA

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

IMAGEM 3D DA CENA SEM CHAO

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

Geração da Mesh

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

Mais algoritmos?

Reconhecimento / matching de objetos?

Data Visualization / Recording´

Surface Reconstruction of Point Clouds Captured with Kinect [pdf]

Visualizador usado é o do PCL

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

Modelos / Nuvens de pontos pontuais para ply ou obj?

Problemas e Soluções

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

“Multiple Kinect Studies - Technical Report”

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

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

Time Critical Isosurface Reﬁnement And Smoothing

Ruído

Pontos a mais que não conseguiram ser filtrados.

Tempo real

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

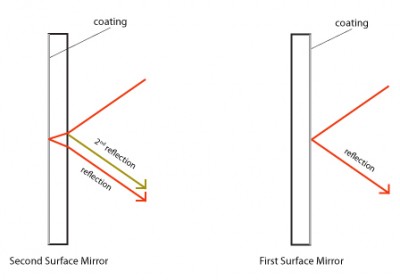
*MeshLab* pode ser usado para modelos fixos e o “lixo” pode ser retirado manualmente.

Buracos de Informação

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

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

Informação Imprecisa



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

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

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

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

Tecnologia

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

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

<http://opencv.org/>

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

<http://pointclouds.org/>

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

Results

Object Acquisition

Show examples of Point Clouds and Meshes.

Continuous acquisition for missing data

Hole filtering technics/3D processing

Real-time Acquisition

Show examples of Point Clouds videos and Fast Mesh Generation

Quality analysis

Performance Analysis.

Conclusion and Future Work

Conclusions

Future Work

Kinect 2 – mais resolução (profundidade e imagem)

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

Appendix

User Guide

Related Work (OLD)

3D Scanning / Images Aquisition

Things

Technology (Instruments)

Describe each technology and point the pros and cons.

Structured Light

Stratified light?

Depth Cameras (emphasis on Kinect)

Ranging Cameras

Triangulation Scanners

**Mesa Imaging SwissRanger 4000 (SR4000)**

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

**PMD Technologies CamCube 2.0**

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

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

**PMD Nano**

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

**Bumblebee 2 e XB3 specs**

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

**Bumblebee 2**

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

**Bumblebee XB3**

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

**Kinect**

**Primesence**

**Other Cameras**

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

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

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

Systems

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

From Capture to Models

Mesh construction from point clouds.

Smoothing

Color issue.

Normal, reflection, etc.

­­­­­­­Summary

Things

1. <http://www.matterform.net/scanner> [↑](#footnote-ref-1)
2. <http://store.makerbot.com/digitizer.html> [↑](#footnote-ref-2)
3. <http://www.primesense.com/solutions/3d-sensor/> [↑](#footnote-ref-3)
4. [http://structure.io](http://structure.io/) [↑](#footnote-ref-4)
5. <http://cad-scan.co.uk/> [↑](#footnote-ref-5)
6. <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.vitronic.de/en [↑](#footnote-ref-25)
26. http://www.optimet.com/ [↑](#footnote-ref-26)
27. <http://msdn.microsoft.com/en-us/library/jj131033.aspx>

    <http://www.engadget.com/2010/06/13/microsoft-kinect-gets-official/> [↑](#footnote-ref-27)
28. Time-of-Flight Cameras and Microsoft Kinect [Book][pdf] [↑](#footnote-ref-28)
29. <http://www.asus.com/Multimedia/Xtion_PRO_LIVE/#specifications> [↑](#footnote-ref-29)
30. <http://www.primesense.com/get-your-sensor2/> [↑](#footnote-ref-30)
31. http://www.guinnessworldrecords.com/records-9000/fastest-selling-gaming-peripheral/ [↑](#footnote-ref-31)
32. 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-32)
33. <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-33)
34. <http://www.space-vision.jp/EP-Body_Scanner.html> [↑](#footnote-ref-34)
35. <http://www.4ddynamics.com/3d-scanners/bodyscanner/> [↑](#footnote-ref-35)
36. <http://ir-ltd.net/> [↑](#footnote-ref-36)
37. <http://tc2.com/index_3dbodyscan.html>

    <http://www.tc2.com/pdf/kx16.pdf> [↑](#footnote-ref-37)
38. <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-38)
39. <http://www.iti.gr/iti/files/document/publications/manuscript6camready_.pdf> [↑](#footnote-ref-39)
40. 3D Face Reconstruction from a Single Camera Using a Multi-Mirror Set-up [pdf] [↑](#footnote-ref-40)
41. Surround Structured Lighting for Full Object Scanning [pdf] [↑](#footnote-ref-41)
42. http://store.makerbot.com/digitizer.html [↑](#footnote-ref-42)
43. http://www.matterform.net/scanner [↑](#footnote-ref-43)
44. 3D Body Scanning in a Mirror Cabinet [pdf] [↑](#footnote-ref-44)
45. Centro de Computação Gráfica (www.ccg.pt) [↑](#footnote-ref-45)