

Agradecimentos

A concretização da presente dissertação não teria sido possível sem o apoio de um grupo chegado de pessoas, às quais estou sinceramente agradecido.

Em primeiro lugar agradeço ao Professor Doutor José António Gaspar, o meu orientador científico, pelo empenho na orientação do trabalho aqui apresentado e pela constante transmissão de conhecimentos. Ao Professor Doutor José Santos Victor, o meu Co-orientador, pelo seu vasto trabalho desenvolvido na área de visão.

Agradeço aos meus colegas, também a realizar as suas teses, pelo ambiente favorável e de motivação, nomeadamente ao Paulo Louro e ao Luís Salvado. Obrigado também aos dois pelas partilhas de vida. Ao meu colega João Vitorino, também a escrever a sua Tese sobre processamento de imagem, me permitiu debater inúmeros problemas.

Ao Instituto Superior Técnico e ao Instituto de Sistemas e Robótica pelas instalações e material de apoio ao trabalho.

Aos meus amigos de longa data: Isaac, Rafa, Gémea, Silvestre, por me terem aturado ao longo dos anos. Aos amigos mais recentes: Ana, obrigado pela partilha, Mário, pelos debates construtivos, Sílvia, pela visão realista. Aos vários amigos do IST com quem partilhei momentos que nunca mais esquecerei.

Aos meus familiares: madrinha, Lili, cunhada, mano, Maria Carolina, pelas chamadas de atenção e pura amizade.

Mas a pessoa chave durante todo o tempo que dediquei não só na elaboração da dissertação mas também em todo o curso, foi o meu pai, João. Foi ele que me apoiou na decisão de me deslocar de Santarém para Lisboa e perante situações complicadas da vida, ainda assim, me apoiar e instruir no caminho de ser alguém. Obrigado mãe.

Como fotografo amador gosto muito de fotografar e de tudo o que envolve fotografia. Quis aplicar os conhecimentos de engenharia que o curso me proporcionou de forma a saber ainda mais sobre fotografia e unir estes dois mundos onde vivo. Este projecto trabalha essencialmente com processamento de imagem. Tendo a intenção de aprender mais sobre o método de processamento imagem (técnicas e potencialidades) e saber o que acontece aos pontos que constituem uma fotografia digital, este projecto, entre os vários do Professor José Gaspar, foi o que mais me aliciou.

Resumo

Neste trabalho é abordado o problema de construir uma imagem que contenha toda a superfície interior de um tubo. Este trabalho envolve planificar e combinar imagens adquiridas por uma câmara que se movimenta por dentro do tubo.

Para criar uma imagem mosaico do interior de uma superfície tubular é necessário adquirir as imagens, encontrar pontos correspondentes entre imagens consecutivas e reconstruir o cenário tridimensional, ajustar a nuvem de pontos tridimensional a um modelo mínimo de representação, estimar o caminho da câmara e planificar os pontos reconstruídos para a imagem final.

Este trabalho concentra-se na reconstrução do cenário, na construção do modelo simples de representação de uma superfície tubular e na construção da imagem mosaico.

O modelo mínimo de representação tridimensional é baseado em secções cilíndricas, e é usado tanto para simular estruturas tubulares como para o processo de ajuste da nuvem de pontos reconstruídos. Este modelo também foi construído para tornar a planificação mais fácil.

Palavras chave: *Imagem Mosaico, Algoritmo dos oito-pontos, modelo de ajuste tridimensional, reconstrução tridimensional, pontos correspondentes.*

Abstract

In this work we approach the problem of imaging the complete interior surface of a tubular shape with one single image. This involves dewarping and combining (mosaicing) images acquired by a moving camera.

Our framework for mosaicing the interior of a tubular shape involves acquiring images, finding and reconstructing corresponding feature points, fitting a simple 3D model to the reconstructed points, estimating the camera path and dewarping to one single mosaic image.

This work focus on reconstructing the tubular shape, on the 3D Model fitting and on the mosaic image.

We use a 3D model that is based on cylindrical sections, and is useful for both generating simulated data and implementing the fitting procedure. This 3D model was also constructed to make easier the dewarping to the mosaic image.

Keywords: *Mosaicing tubular shapes, eight-point Algorithm, 3D model fitting, 3D reconstruction, matching.*

Conteúdo

Agradecimentos	i
Resumo	iii
Abstract	v
1 Introdução	1
1.1 Trabalho Relacionado	3
1.2 Organização da Tese	3
2 Introdução Teórica	5
2.1 Geometria Projectiva	5
2.2 Modelo de uma câmara fotográfica perspectiva	5
2.3 <i>Dewarp</i> de polar para cartesiano	7
2.4 Correção do <i>Dewarp</i> devido ao efeito perspectiva	8
2.5 Pontos correspondentes entre imagens	8
2.6 Homografia	9
2.7 Estimação da matriz H	11
3 Reconstrução Tridimensional	13
3.1 Estimação da Matriz Essencial	14
3.2 Factorização da Matriz Essencial e Reconstrução	16
3.3 Normalização do Algoritmo dos 8 pontos	17
3.4 Estimação da Localização da Câmara	18
3.5 Resultados	18
3.5.1 Estudo do efeito do ruído na estimação de movimento de translação	18
3.5.2 Estudo do efeito do ruído na estimação de movimento de rotação e translação	20
3.5.3 Experiência 1 - Reconstrução de uma estrutura composta por dois troços cilíndricos	22
3.5.4 Experiência 2 - Reconstrução de uma estrutura tubular considerando profundidade de visão limitada	24
4 Modelo do Tubo	27
4.1 Introdução	27

4.2	Modelo Tridimensional	27
4.3	Representação Mínima da Estrutura Tubular	27
4.4	Resultados	29
5	Abertura do Tubo	31
5.1	Introdução	31
5.2	<i>Dewarping</i>	31
5.3	Resultados	32
6	Conclusões e Trabalhos Futuros	35
	Bibliografia	37
A	Testes Adicionais em Estimção de Movimento e Reconstrução	39
B	Manual do Programa	43
B.1	Construção do Tubo	43
B.2	Captura de Imagens	44
B.3	Reconstrução Tridimensional	45
B.4	Ajuste ao Modelo Tubular	46

Lista de Figuras

1.1	(a) Câmara catadióptrica com espelho convexo (www.0-360.com), (b) Lente olho de peixe Nikon Fc-E8 (www.nikon.com), imagem adquirida por uma lente olho de peixe Sunex (www.sunex.com) e imagem panorâmica obtida a partir da imagem anterior.	2
2.1	(a) Modelo Câmara Pontual - <i>Pin-hole</i> , (b) Representação alternativa de <i>Pin-hole</i> , (c) Sensor <i>CCD</i> de uma câmara	6
2.2	(a) Imagem perspectiva com a selecção da zona para <i>Dewarp</i> e algumas imagens panorâmicas prontas a serem unidas, (b) Transformação polar \rightarrow cartesiano	7
2.3	Imagem panorâmica sem (a) e com (b) correcção de perspectiva.	8
2.4	Pontos correspondentes entre duas imagens	9
2.5	Mapeamento de um ponto de um plano em duas imagens diferentes.	10
3.1	Duas projecções x_1, x_2 do ponto p . Transformação Euclideana entre duas imagens dada por (R, T) . A intersecção da linha (o_1, o_2) com cada plano de imagem tem o nome de epipolo e_1, e_2 . As linhas l_1, l_2 são as linhas epipolares, e são a intersecção entre o plano (o_1, o_2, p) e os dois planos de imagem.	14
3.2	Quatro possíveis soluções da estimação de E. Entre a solução da esquerda e da direita, a baseline é invertida. Entre as soluções de cima e de baixo, a câmara B é rodada 180° . Notar que só em (a) é que o ponto está a frente das duas câmaras.	17
3.3	Estimação de movimento de translação: (a) localização do cubo de pontos 3D utilizados na estimação do movimento; (b) translações da câmara desde o ponto central (câmara no centro com dimensão maior) para várias latitudes e longitudes. (c) estimação do movimento da câmara considerando ruído de 1 pixel nos pontos correspondentes.	19
3.4	Erro médio na estimação do movimento de translação. Erro médio para (a) ruído de 1pixel e (b) ruído de 5pixels nos pontos de imagem correspondentes.	20
3.5	Estimação de movimento compreendendo simultaneamente translação e rotação: (a) rotações e translações da câmara desde o ponto central (câmara no centro com dimensão maior) para várias latitudes e longitudes, e cubo de pontos 3D utilizados na estimativa do movimento.; (b) Estimação do movimento da câmara com ruído de 1 pixel nos pontos correspondentes.	21
3.6	Erro médio na estimação do movimento translação e rotação. Ruído de 1pixel (a) e de 5pixels (b) nos pontos correspondentes.	22

3.7	Experiência 1, estimação do movimento da câmara num tubo com duas secções: (a) imagens do interior; (b,c) posição estimada para observações sem ruído e com ruído; (d,e) estrutura reconstruída respectivamente para (b) e (c).	23
3.8	Experiência 2, estimação de movimento e de estrutura limitando a profundidade máxima do campo de visão: imagens obtidas no interior da estrutura, 640x480pixels (a) e respectiva reconstrução para ruído de 0.01 pixels; (c) reconstrução utilizando imagens de maior resolução, 3000x2000 pixels, com um ruído de 0.1 pixels.	25
4.1	(a) Modelo do tubo. (b) Exemplo de um tubo.	28
4.2	Ajuste tridimensional para cilindros: (a) Ajustar uma nuvem de pontos a um cilindro (b) remover zonas sobrepostas de superfícies cilíndricas (c) reajustar os cilindros para construir um eixo coerente entre todos os cilindros (d) ajuste final dos cilindros.	28
4.3	Exemplo de secções cónicas	29
4.4	Ajustamento das superfícies cilíndricas: (a) Pontos reconstruídos, (b)(c)(d) Optimização das ligações entre as secções, <i>Azul</i> círculos do modelo tridimensional, <i>Magenta</i> Pontos tridimensionais originais, <i>Azul claro</i> distância entre os pontos originais e a superfície tubular.	29
4.5	Reconstrução da superfície tubular aplicando o modelo.	30
5.1	Dewarp do modelo de representação mínima do tubo	31
5.2	Tubo tridimensional	32
5.3	Imagens capturadas pela câmara (a,c). Imagens dewarped (b,d).	32
5.4	Imagem Mosaico final.	33
5.5	Tubo tridimensional	33
5.6	Imagens capturadas pela câmara (a,c). Imagens dewarped (b,d).	34
5.7	Imagem Mosaico final.	34
A.1	Localização descentrada dos pontos tridimensionais: (a) longitude 0°-90° (b) longitude 90°-180° (c) longitude 180°-270° (d) longitude 270°-360°. Ruído de 1 pixel.	40
A.2	Estudo da influência do ruído: Resoluções de imagem de 1/3Megapixels (a,b), 1Mp (c,d), 6Megapixels (e,f), 10Megapixels (g,h) com níveis de ruído de 0.001% (a,c,e,g) e de 0.01% (b,d,f,h).	41
A.3	Estudo da influência do ruído aleatório, combinado a resolução da imagem, sobre a reconstrução.	42
B.1	Sequência dos blocos de código	43
B.2	Uma secção cilíndrica do tubo final.	44
B.3	Exemplo de um tubo com oito secções.	45
B.4	Exemplo de três imagens do interior do tubo	45
B.5	Exemplo de reconstrução de uma estrutura tubular.	48
B.6	Optimização das secções cónicas.	48
B.7	Aplicação de secções cilíndricas aos pontos reconstruídos.	48
B.8	Secções cilíndricas sem zonas sobrepostas	49
B.9	Modelo representativo da estrutura tubular.	50
B.10	Optimização do modelo e das superfícies cónicas.	50

Capítulo 1

Introdução

Inspeção visual é uma actividade ubíqua numa sociedade moderna. A inspeção visual aparece em actividades tão simples quanto comprar comida ou roupa, até actividades complexas como os exames médicos de endoscopia. É normal encontrar inspeção visual na indústria, por exemplo na verificação de garrafas efectivamente cheias ou de rolhas bem fabricadas, e em trabalho de manutenção de estruturas, como por exemplo a verificação do estado de pontes ou de túneis.

Neste trabalho estamos interessados em particular na inspeção visual de estruturas tubulares tais como condutas de água, gás ou túneis, e consideramos vídeo como o sensor essencial para a obtenção de informação. Apesar da utilização do vídeo, a inspeção visual do interior de estruturas tubulares ainda é naturalmente um processo moroso dada a natureza sequência intrínseca ao vídeo: estruturas longas implicam vídeos longos e a tecnologia actual ainda não ajuda suficientemente na selecção da informação a apresentar a uma pessoa. Neste trabalho procuramos mostrar que a utilização de vídeo pode ser melhorada com a criação automática de sínteses que facilitam a análise das estruturas.

No processo de inspeção visual de uma estrutura tubular existe normalmente uma grande repetição da informação adquirida ao longo do tempo. Uma secção da estrutura tubular é normalmente observada em várias imagens antes de sair do campo de visão da câmara. É assim objectivo deste trabalho contribuir para encontrar formas de seleccionar informação relevante, mais concretamente criar uma representação global em que cada ponto da superfície interior é observado uma única vez. Por outras palavras, pretende-se criar uma única imagem que represente toda a superfície interior do tubo, realizando um mosaico com partes das imagens do vídeo.

No caso de um tubo cilíndrico, sem curvas, e no caso de a câmara ser posicionada no eixo do tubo e perfeitamente alinhada, a criação da imagem mosaico resume-se a uma transformação de coordenadas polares para coordenadas cartesianas seguido da definição de pontos correspondentes em cada par de imagens e a implementação do método da homografia [4]. Neste trabalho desejamos permitir uma forma livre para o tubo e também movimento livre da câmara dentro do tubo, pelo que a transformação de coordenadas e o método de homografia não é suficiente para criar a imagem mosaico.

A construção de imagens mosaico implica portanto conhecer o movimento relativo da câmara dentro do tubo. Para reconstruir o movimento é necessário definir pontos correspondentes entre as imagens, podendo em seguida estimar a matriz essencial [3]. A matriz essencial contém a rotação e a translação da câmara entre um par de imagens de um sistemas estéreo. Conhecendo o caminho percorrido pela câmara é possível reconstruir o cenário tridimensional. Com o cenário tridimensional reconstruído o processo de planificação para imagem mosaico con-

siste em abrir o tubo e torná-lo numa imagem bidimensional.

O facto de a reconstrução da estrutura tubular resultar essencialmente numa nuvem de pontos 3D, não contendo por exemplo a informação explícita de eixo ou de raio da estrutura, implica que não está definida implicitamente uma planificação da estrutura tubular. É então necessário proceder à modelação da estrutura sob a forma de uma superfície 3D que por um lado representa fielmente os pontos reconstruídos e por outro lado permite uma planificação unívoca.

A planificação unívoca de uma estrutura tubular geral, i.e. uma estrutura contendo curvas ao longo do percurso e tendo secção de raio variável, implica uma modelação local da própria estrutura. Em particular consideramos que a estrutura pode ser representada por uma sequência de secções tubulares cilíndricas. Mais concretamente, o modelo tridimensional simplificado organiza os pontos reconstruídos numa lista de círculos que representam as secções tubulares. A lista de círculos compreende três parâmetros: o ponto 3D central do círculo, o raio e o vector normal ao plano do círculo.

Um aspecto que é definitivamente relevante para o problema de inspecção visual de estruturas tubulares é a escolha da geometria (óptica) da câmara de vídeo. Existem várias geometrias possíveis como por exemplo a câmara catadióptrica omnidireccional (Fig.1.1(a)), ou a câmara equipada com uma lente de grande angular (olho de peixe, Fig.1.1(b)) ou simplesmente a câmara perspectiva. A escolha da geometria para além de definir o campo de visão, influencia a resolução disponível para cada azimute ou elevação medidos em relação ao eixo óptico da câmara. Por exemplo uma lente olho de peixe apresenta uma diminuição acentuada da resolução quando o ângulo aumenta desde o centro para a periferia da imagem. Trabalhos recentes de investigação em modelação da geometria apontam contudo para representar em primeira aproximação os vários tipos de câmara como câmaras de centro de projecção único. Este é também o caminho seguido neste trabalho: começamos pelos fundamentos de geometria e de processamento assumindo centro único. Também em primeira aproximação consideramos a resolução da câmara uniforme no plano de imagem, ou seja começamos por utilizar o modelo simples de câmara perspectiva, deixando para análise posterior outras geometrias que se esperam trazer vantagens no campo de visão, na resolução e/ou na minimização do tempo computacional para obter os mosaicos.

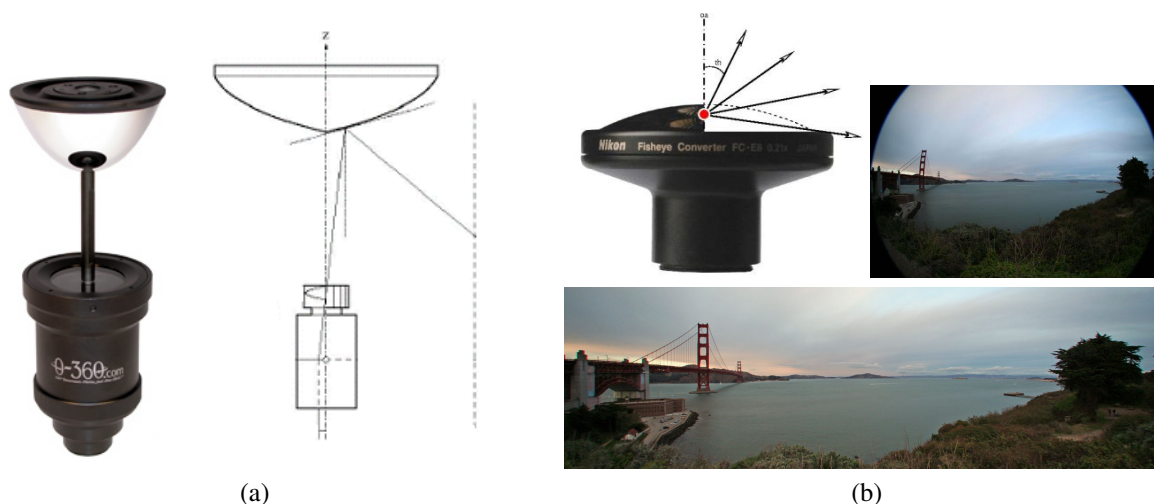


Figura 1.1: (a) Câmara catadióptrica com espelho convexo (www.0-360.com), (b) Lente olho de peixe Nikon Fc-E8 (www.nikon.com), imagem adquirida por uma lente olho de peixe Sunex (www.sunex.com) e imagem panorâmica obtida a partir da imagem anterior.

1.1 Trabalho Relacionado

Criação de imagem mosaico é frequentemente utilizada para aumentar o campo de visão através da união de várias imagens. Por exemplo em [8], um avião equipado com uma câmara adquire imagens na direcção do movimento, e mapeia estas imagens sobre uma superfície cilíndrica virtual de eixo coincidente com a trajectória do avião. A superfície cilíndrica é finalmente aberta, resultando num mapa onde se observam sem distorção por exemplo estradas ou parques de estacionamento. Neste caso, a superfície cilíndrica por estar ligada à trajectória e representar com precisão o próprio cenário dado serem muito elevadas as distâncias entre avião e objectos observados - a discretização da imagem implica que a informação de profundidade deixa de relevante a partir de um certo limiar. No nosso caso desejamos inspeccionar estruturas tubulares cujo raio tem uma dimensão na ordem de grandeza da própria câmara, pelo que pequenos afastamentos da câmara em relação ao eixo da estrutura implicam efeitos de paralaxe significativos.

O facto de considerarmos estruturas tubulares *finas*, e portanto ser necessário considerar efeitos de paralaxe quando a câmara se afasta do eixo, implica um estudo cuidado da geometria do sistema: conhecimento da localização da câmara e conhecimento da geometria da própria estrutura tubular. No sentido de manter a generalidade do trabalho é importante introduzir métodos de estimação do movimento da câmara e de reconstrução do cenário (estrutura tubular).

A reconstrução de cenários é uma área bem conhecida em processamento de imagem. Tradicionalmente a reconstrução de cenários começa por identificar pontos correspondentes numa sequência de imagens. Esta identificação pode ser feita de uma maneira robusta usando por exemplo as funções SIFT [6]. Depois a reconstrução do movimento da câmara entre as imagens consecutivas pode ser determinada por estimação e factorização da matriz essencial [4], assumindo que se sabe a calibração da câmara.

Actualmente, usando sequência de imagens que mostrem movimentos suaves da câmara, a reconstrução de cenário está muito mais desenvolvida com as técnicas de processamento SLAM (*Simultaneous Localization and Mapping*) ou vSLAM (*Visual SLAM*). Estas técnicas usam essencialmente um veículo autónomo que inicia a sua marcha num local indeterminado e ao longo do tempo vai construindo uma mapa do seu caminho e do ambiente que o rodeia [5].

No caso deste trabalho, a reconstrução da estrutura tubular envolve um dos métodos descritos nos parágrafos anteriores. Em particular desenvolvemos nesta dissertação trabalho baseado na estimação e factorização da matriz essencial. Dada a reconstrução de um cenário, isto é, uma nuvem de pontos, é pretendido que esta nuvem seja descrita num modelo simples de representação do cenário. Este modelo tridimensional proporciona um método fácil de criar a planificação da superfície do tubo para uma imagem mosaico.

Parte do trabalho descrito nesta dissertação foi também publicado em workshop nacional. Mais especificamente o modelo 3D para uma estrutura tubular e o seu ajuste um ambiente simulado foi publicado em Novembro de 2007, no workshop RecPad 2007 [9].

1.2 Organização da Tese

Esta tese descreve um problema de processamento de imagem, mais precisamente reconstrução tridimensional e sua planificação para uma imagem panorâmica. Esta tese está organizada em três capítulos principais.

O **capítulo 3** descreve o problema da reconstrução tridimensional a partir de imagens perspectivas. Descreve como, a partir de pontos correspondentes entre imagens e a estimação da matriz essencial, se reconstrói um cenário

e se estima o caminho percorrido pela câmara ao longo do tempo.

O **capítulo 4** discrimina um modelo mínimo de representação tridimensional de uma superfície tubular. Este modelo tem como objecto tornar mais simples o dewarp do tubo para imagem panorâmica. Além de reduzir a informação de todos os pontos tridimensionais a uma lista de secções cilíndricas, possibilita também a própria construção da superfície tubular.

O **capítulo 5** apresenta o processo de abertura do tubo. Dá uso ao modelo mínimo tridimensional e descreve como a imagem final inclui toda a superfície interna do tubo.

Por fim no **capítulo 6** são tecidas as conclusões do trabalho e possíveis trabalhos futuros.

Antes destes capítulos são descritos alguns trabalhos relacionados e é feita uma introdução teórica, **capítulo 2** sobre Geometria Projectiva, modelo de uma câmara fotográfica, construção de imagens panorâmicas e Homografia.

Capítulo 2

Introdução Teórica

Nesta introdução teórica é descrito como pode ser construída uma imagem mosaico usando principalmente numa técnica chamada *Homografia*. É descrito também como uma câmara fotográfica é representada algebricamente. É ainda descrito como cada fotografia tem de ser transformada para uma imagem panorâmica para que, aplicando a Homografia, as várias imagens se possam unir e criar a imagem final, mosaico.

2.1 Geometria Projectiva

Na geometria cartesiana, as entidades geométricas podem ser representadas de forma algébrica, segundo um sistema de coordenadas. A um ponto corresponde um par de coordenadas (x, y) , em \mathbb{R}^2 . Considerando \mathbb{R}^2 como sendo um espaço vectorial, o par de coordenadas (x, y) é um vector. No caso de uma recta de equação $ax + by + c = 0$, esta também pode ser representada pelo vector (a, b, c) . Diferentes escolhas de a , b e c dão origem a diferentes rectas. No entanto, se o vector (a, b, c) for multiplicado por uma constante o resultado é a mesma recta e se o vector (x, y) for multiplicado por uma constante não dá o mesmo ponto. É por este motivo que o vector (a, b, c) é homogéneo e o vector (x, y) não é.

Portanto, as equações da geometria projectiva, expressas em coordenadas cartesianas, não são lineares. Se essas equações forem expressas em coordenadas homogéneas tornam-se lineares. coordenadas cartesianas: $x = (x, y)^T$
coordenadas homogéneas: $x = (x_1, x_2, x_3)^T$

$$x = \frac{x_1}{x_3} \quad y = \frac{x_2}{x_3} \quad (2.1)$$

2.2 Modelo de uma câmara fotográfica perspectiva

O modelo mais usado em visão por computador para modelar uma câmara é o modelo de câmara pontual, *pinhole camera model*, como representado na Fig.2.1(a). É um modelo simples e eficaz para modelar a grande maioria de câmaras existentes no mercado.

Este modelo considera que os raios de luz emitidos ou reflectidos por um dado objecto passam todos por um "furo", pinhole, sendo projectados numa superfície plana, i.e., no plano da imagem π . O modelo de câmara pontual consiste num plano, o plano da imagem, e um ponto no espaço euclideano, o centro da projecção. A distância entre

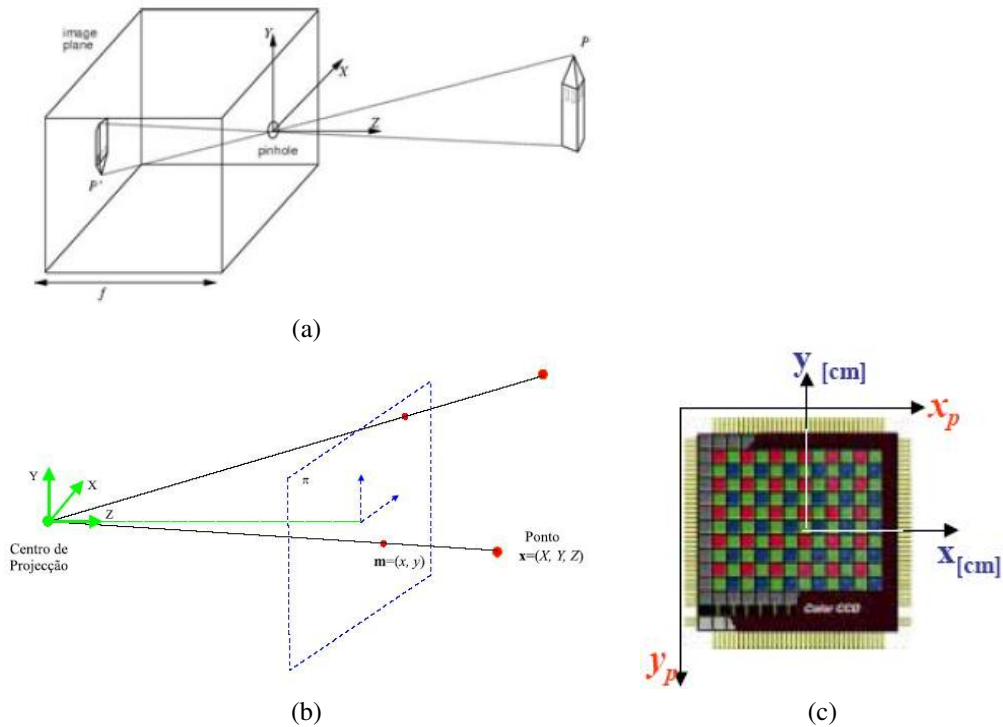


Figura 2.1: (a) Modelo Câmara Pontual - *Pin-hole*, (b) Representação alternativa de *Pin-hole*, (c) Sensor *CCD* de uma câmara

o plano da imagem e o centro de projecção é denominada como distância focal f . A linha que passa pelo centro de projecção e é perpendicular ao plano da imagem é denominada como eixo óptico. A intersecção entre o eixo óptico e o plano da imagem é um ponto denominado *centro da imagem*.

O referencial ortonormado da câmara é ainda definido através do centro de projecção e pelo eixo óptico que coincide com o eixo Z do referido referencial. A utilização da geometria projectiva permite que o modelo de projecção perspectiva seja descrito por uma relação linear. Assim, uma câmara pode ser descrita, mais facilmente, através do mapeamento projectivo linear entre o espaço projectivo P^3 e P^2 , eq.2.2. O modelo *Pin-hole* pode ser representado alternativamente segundo a Fig.2.1(b).

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (2.2)$$

$$m = PM \quad (2.3)$$

Dado um ponto 3D no mundo de coordenadas $(X_w, Y_w, Z_w)^T$, a sua projecção no plano de imagem em coordenadas de imagem (u, v) , é a matriz de projecção P . Esta matriz inclui os parâmetros intrínsecos K e parâmetros extrínsecos $[R, t]$, presentes na Eq.2.4.

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} {}^cR_W & {}^ct_W \end{bmatrix} \cdot \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (2.4)$$

onde cR_W é a matriz (3×3) de rotação do mundo definida no referencial da câmara, ct_W são as coordenadas da origem do referencial do mundo expressas no referencial da câmara, e K , a matriz de dimensão (3×3) é matriz de parâmetros intrínsecos da câmara.

Os parâmetros intrínsecos expressos na matriz K , dependem apenas da própria câmara. Transformam as coordenadas no referencial da câmara, da projecção do raio óptico no plano de imagem, para as coordenadas de imagem, Eq.2.5.

$$K \doteq \begin{bmatrix} f \cdot k_u & f \cdot k_\theta & u_0 \\ 0 & f \cdot k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.5)$$

Na matriz K , o f é a distância focal (em milímetros), o ponto $(u_0, v_0)^T$ (em pixels) são as coordenadas da intercepção do eixo óptico com o plano de imagem, ou ponto central, k_u e k_v são factores de escalamento da dimensão dos pixels na horizontal e na vertical (inverso de milímetros), e k_θ é o factor de skew entre os eixos da imagem (em radianos).

A imagem de um sensor de uma câmara fotográfica pode ser vista na Fig.2.1(c), onde é representado a dimensão dos pixels da imagem (k_u e k_v) e a posição do ponto principal (u_0, v_0) .

2.3 Dewarp de polar para cartesiano

Dewarping é uma transformação que altera a disposição dos pontos de uma imagem. No caso das imagens perspectiva do interior do tubo, um anel de pontos é modificado de forma a criar uma imagem panorâmica, visível na Fig.2.2(a). Este *Dewarp* é uma transformação de coordenadas polares para coordenadas cartesianas.

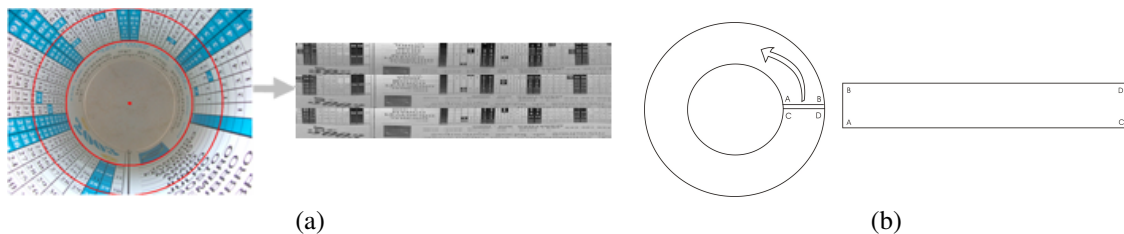


Figura 2.2: (a) Imagem perspectiva com a selecção da zona para *Dewarp* e algumas imagens panorâmicas prontas a serem unidas, (b) Transformação polar \rightarrow cartesiano

O anel de pontos que sofrerá o *Dewarp*, zona de interesse, é definido com 3 pontos: o ponto central do anel, um ponto de raio interior e um ponto de raio exterior. A transformação aplicada pode ser vista na Fig.2.2(b).

2.4 Correção do *Dewarp* devido ao efeito perspectiva

Nestas imagens perspectiva do tubo, há ainda um pormenor a ter em atenção. A superfície do tubo mais perto da câmara, situada na parte exterior da imagem, tem uma representação (em número de pixels) muito maior que a superfície do tubo mais longe da câmara, situada na parte interior da imagem. Este pormenor faz com que as imagens panorâmicas geradas tenham uma espécie de compressor na direcção do caminho, que pode ser vista na Fig.2.3(a).

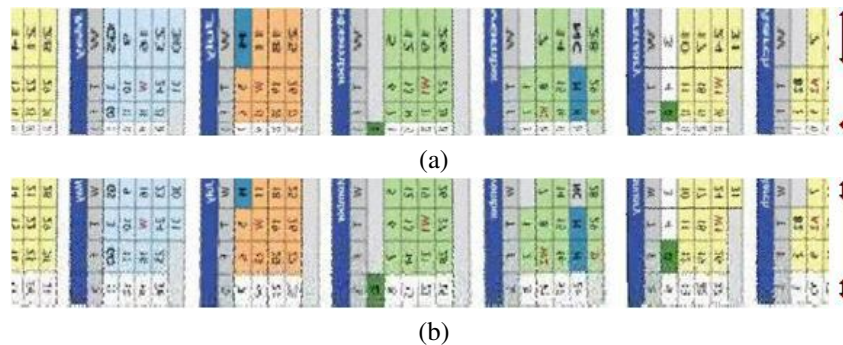


Figura 2.3: Imagem panorâmica sem (a) e com (b) correcção de perspectiva.

Assim é necessário interpolar a imagem (de uma forma crescente) para que a representação da superfície do tubo, ao longo da imagem, seja homogénea. Correção já visível na Fig.2.3(b).

2.5 Pontos correspondentes entre imagens

A detecção de pontos correspondentes entre imagens consecutivas pode ser feita de forma automática. Por exemplo, com SIFT (Scale Invariant Features Transform) de David Lowe [7]. SIFT é uma técnica que detecta e extrai descritores locais, razoavelmente invariáveis a mudanças de iluminação, ruído de imagem, rotação, escala e pequenas mudanças de perspectiva. Os descritores obtidos com a técnica SIFT são altamente distintos. Por exemplo, num banco de dados extensos com descritores para diversas imagens, um determinado ponto correspondente tem alta probabilidade de ser encontrado. Outro aspecto importante da técnica SIFT é a geração de um elevado número de descritores, conseguindo-se cobrir densamente uma imagem quanto a escalas e localizações.

A obtenção de descritores SIFT é feita através das seguintes etapas:

1. Detecção de extremos:

Pontos extremos são pontos onde existe uma elevada variação de contraste por comparação com pontos vizinhos. São encontrados todos estes pontos, utilizando-se filtros gaussianos, de modo a identificar pontos de interesse invariáveis à escala, rotação e localização na imagem.

2. Localização de pontos chave:

Todos os pontos detectados como extremos podem ser pontos chave. Pontos Chave, ou pontos de interesse, são pontos onde as medidas de estabilidade são respeitadas. Por exemplo pontos com baixo nível de contraste são descartados.

3. Definição de orientação:

A orientação de cada ponto chave é definida através do gradiente local da imagem.

4. Descritor de pontos chave:

Construção dos descritores ao medir-se gradientes locais em uma região vizinha a cada ponto de interesse.

Estas medidas são então transformadas para uma representação que permite níveis significativos de distorção e mudança na iluminação. Em tarefas de comparação de imagens e reconhecimento, estes descritores SIFT são extraídos das imagens para então serem comparados.

Um exemplo dessa comparação pode ser visto na Fig.2.4:

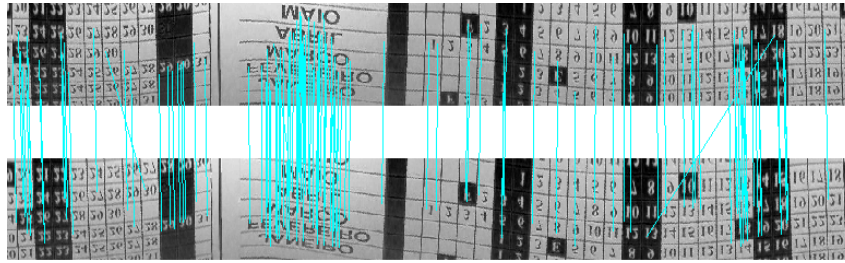


Figura 2.4: Pontos correspondentes entre duas imagens

2.6 Homografia

Uma homografia é uma transformação que relaciona as coordenadas dos pontos no mundo real, com as suas projecções na imagem. Um ponto x , do plano no mundo real, passa a ficar relacionado com a respectiva projecção x' , na imagem. A coordenada Z tem sempre o valor zero.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{14} \\ p_{21} & p_{22} & p_{24} \\ p_{31} & p_{32} & p_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \quad (2.6)$$

A matriz P , (3×4) , passa a ser (3×3) e toma a designação de homografia ou matriz H . Um ponto x , do plano no mundo real, passa a ficar relacionado com a respectiva projecção x' , na imagem. Essa relação é definida pela Eq.2.7:

$$\begin{bmatrix} x'_1 \\ x'_2 \\ x'_3 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad (2.7)$$

Esta igualdade pode resumir-se por $x' = Hx$. A multiplicação da matriz H por qualquer valor escalar não nulo não provoca alteração na homografia, pelo que esta se designa por matriz homogénea. Tal como na representação homogénea de um ponto, é a razão entre os elementos da matriz que é importante manter, Eq.2.1.

Existem oito razões independentes entre os nove elementos da matriz pelo que se diz que uma homografia tem oito graus de liberdade. Recorrendo à Eq.2.7 e à Eq.2.1 conclui-se que cada correspondência entre um ponto no mundo real e a sua imagem dá origem a às equações 2.8 e 2.9.

$$x' = \frac{x'_1}{x'_3} = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}} \quad (2.8)$$

$$y' = \frac{x'_2}{x'_3} = \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}} \quad (2.9)$$

A Eq.2.8 e a Eq.2.9 podem ser reescritas da forma:

$$x' (h_{31}x + h_{32}y + h_{33}) = h_{11}x + h_{12}y + h_{13} \quad (2.10)$$

$$y' (h_{31}x + h_{32}y + h_{33}) = h_{21}x + h_{22}y + h_{23} \quad (2.11)$$

Verifica-se, assim, que cada correspondência origina duas equações lineares relativamente aos elementos de H . Quatro pontos dão origem a oito equações lineares de onde é possível extrair os valores de H , a menos de um factor escalar. Existe, no entanto, uma restrição importante - para que as equações sejam linearmente independentes os pontos escolhidos não podem ser colineares.

Para relacionar duas imagens o processo de homografia também é possível. Essa relação está esquematizada na Fig.2.5.

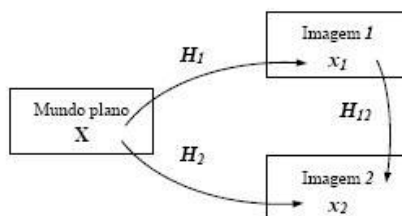


Figura 2.5: Mapeamento de um ponto de um plano em duas imagens diferentes.

A figura 2.5 mostra que para um mesmo ponto X visto nas duas imagens é possível obter o sistema da Eq.2.12.

$$\begin{cases} x_1 = H_1 X \\ x_2 = H_2 X \end{cases} \quad \begin{cases} X = H_1^{-1} x_1 \\ x_2 = H_2 H_1^{-1} x_1 \end{cases} \quad x_2 = H_{12} x_1 \quad H_{12} = H_2 H_1^{-1} \quad (2.12)$$

Sendo H_1 e H_2 matrizes de homografia, as suas inversas são também matrizes de homografia, tal como o produto de uma delas pela inversa da outra. Assim, a matriz H_{12} é também uma matriz de homografia e relaciona as duas imagens.

É este tipo de relação entre imagens de um mesmo objecto que está na origem da construção de mosaicos de imagens, o objectivo deste trabalho. Portanto é necessário calcular as várias matrizes H_{12} que relacionam cada par de imagens.

2.7 Estimação da matriz H

Existem vários métodos para estimação da matriz H . Como H tem oito graus de liberdade uma forma de a estimar é fixando uma das variáveis com um valor arbitrado. Assim, fixando o último elemento $h_{33} = 1$ e tomando um par de pontos correspondentes de coordenadas na primeira imagem (x_1, y_1) e na segunda imagem (x'_1, y'_1) , as equações 2.10 e 2.11 tomam a forma das equações 2.13 e 2.14.

$$x_1 h_{11} + y_1 h_{12} + h_{13} - x'_1 x_1 h_{31} - x'_1 y_1 h_{32} = x'_1 \quad (2.13)$$

$$x_1 h_{21} + y_1 h_{22} + h_{23} - y'_1 x_1 h_{31} - y'_1 y_1 h_{32} = y'_1 \quad (2.14)$$

Sendo apenas necessário, no mínimo, quatro pontos correspondentes entre duas imagens para estimar a homografia, a Eq.2.15 mostra o caso de se utilizar mais do que quatro pontos:

$$\begin{bmatrix} x_i & y_i & 1 & 0 & 0 & 0 & -x'_i x_i & -x'_i y_i \\ 0 & 0 & 0 & x_i & y_i & 1 & -y'_i x_i & -y'_i y_i \end{bmatrix}_A \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{bmatrix}_h = \begin{bmatrix} x'_i \\ y'_i \\ b \end{bmatrix} \quad (2.15)$$

Como a matriz A é uma matriz quadrada, sendo possível calcular a sua inversa, os valores do vector h podem ser obtidos através da Eq.2.16:

$$h = A^{-1}b \quad (2.16)$$

Com quatro correspondências, desde que não sejam colineares, é possível determinar a solução exacta para a matriz H .

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix} \quad (2.17)$$

Capítulo 3

Reconstrução Tridimensional

No caso de estruturas tubulares perfeitamente cilíndricas, sem curvas, percorridas por uma câmara posicionada e alinhada exactamente com o eixo do cilindro, o mosaico do interior da estrutura tubular consistiria somente na realização de um remapeamento de imagens polares para rectangulares seguido pelo registo das várias imagens remapeadas. O registo seria então realizado por uma simples homografia, tal como descrito no capítulo de introdução teórica.

No nosso caso pretendemos considerar estruturas tubulares que incluem não só secções rectas mas também curvas, para além de desejarmos permitir à câmara um movimento livre dentro do tubo. A generalidade da estrutura e a liberdade de movimento da câmara implicam que o remapeamento das imagens para o mosaico deixa de ser constante. É portanto importante estudar o problema de estimar o remapeamento a aplicar a cada imagem adquirida dentro da estrutura tubular.

Existem vários caminhos que podem ser explorados para resolver o problema de estimar um remapeamento variável imagem a imagem. Uma forma possível seria a aproximação de que o remapeamento de cada imagem consiste numa modificação *pequena* em relação a um remapeamento standard para a câmara posicionada perfeitamente sobre o eixo do tubo. Este caminho envolve estudar e escolher os parâmetros de deformação mais importantes. Este estudo não é fácil se não se dispuser de uma caracterização completa da geometria. Por esta razão seguimos uma ideia mais convencional de caracterizar completamente a geometria do problema: reconstruir a localização da câmara a cada instante e simultâneamente reconstruir a estrutura tubular. Dada a estrutura tubular e a respectiva textura fornecida pelas imagens, temos então o conjunto completo de dados que permite obter um mosaico.

Para realizar uma reconstrução tridimensional da estrutura tubular é essencialmente necessária uma sequência de imagens obtidas ao longo da estrutura. Considerando que cada nova imagem tem uma sobreposição com a imagem anterior suficiente para permitir um bom registo e estimação de movimento entre imagens, torna-se então de seguida possível reconstruir o cenário tridimensional.

A estimação da localização da câmara ao longo da estrutura tubular é obtida neste trabalho aplicando o *Algoritmo dos 8 pontos* de Longuet-Higgins [6]. Os principais passos do algoritmo são resumidamente os seguintes:

- Correspondência de pontos entre imagens consecutivas.
- Estimação da matriz Essencial.
- Estimação do movimento de rotação e translação (à parte um factor de escala) entre imagens.

- Uniformização do factor de escala na translação ao longo de toda a sequência de imagens.

De seguida detalham-se os passos mais importantes do *Algoritmo dos 8 pontos*, que permitirá primeiro estimar o movimento da câmara e consequentemente reconstruir a estrutura tubular.

3.1 Estimação da Matriz Essencial

A matriz Essencial contém a informação da rotação e da translação da câmara entre duas imagens. Na Fig.3.1 mostra-se um sistema estéreo e a geometria epipolar subjacente à aplicação do Algoritmo dos 8 pontos.

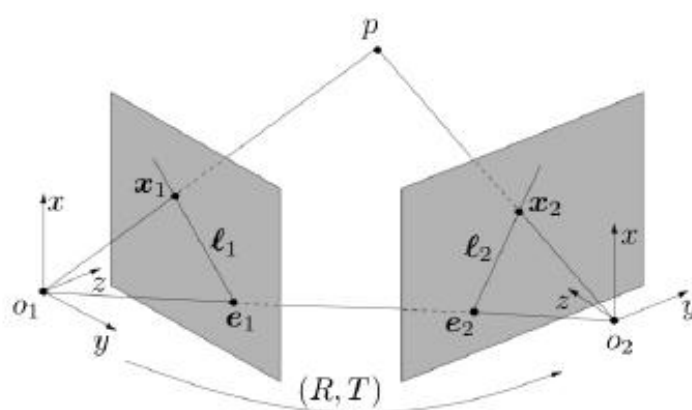


Figura 3.1: Duas projecções x_1, x_2 do ponto p . Transformação Euclidiana entre duas imagens dada por (R, T) . A intersecção da linha (o_1, o_2) com cada plano de imagem tem o nome de epípolo e_1, e_2 . As linhas l_1, l_2 são as linhas epipolares, e são a intersecção entre o plano (o_1, o_2, p) e os dois planos de imagem.

No sistema estéreo, define-se o plano 'epipolar' a partir de 3 pontos: os dois centros ópticos o_1 e o_2 e um ponto p visível pelas duas câmaras. Esse plano intersecta o plano de imagem das câmaras e origina uma recta epipolar l . A recta l_1 está limitada pelo epípolo e_1 (ponto na imagem da câmara 1 onde se vê o centro óptico da câmara 2) e pelo ponto na imagem x_1 (ponto na imagem 1 onde se vê o ponto p). O mesmo acontece se falarmos num ponto da imagem da câmara 2. Este confinamento epipolar é determinante na representação de imagens stéreo e na análise de movimento.

A geometria epipolar permite retirar uma dimensão ao espaço de procura do ponto correspondente, visto que o ponto correspondente na segunda imagem só pode estar numa linha da imagem e não em qualquer ponto do plano da imagem.

Dado ser nulo o volume de um plano pode-se escrever: equação 3.1.

$$o_1 \vec{p} \cdot [o_1 \vec{o}_2 \times o_2 \vec{p}] = 0 \quad (3.1)$$

Assim dado um ponto da imagem 1 \hat{x}_1 , o seu correspondente na outra imagem \hat{x}_2 tem de estar sobre recta epipolar l_2 e pode ser calculado por:

$$\hat{x}_1 \cdot [t]_{\times} R \cdot \hat{x}_2^T = 0 \quad (3.2)$$

onde:

- R é uma matriz rotação entre câmaras
- t é o vector translação entre câmaras, $[t]_{\times}$ é uma forma matricial baseada em t que permite realizar um produto externo

Reescrevendo a Eq.3.2 obtém-se a matriz Essencial, Eq.3.3. Nesta equação são retirados os parâmetros intrínsecos aos pontos correspondentes, Eq.3.4.

$$\hat{x}_2^T E \hat{x}_1 = 0 \quad (3.3)$$

$$\hat{x} = K'_i \cdot x \quad (3.4)$$

A matriz Essencial E é uma matriz (3×3) de característica 2. A matriz E contém a matriz rotação e o vector de translação entre os dois centros ópticos $[R, t]$, ou dito de outra forma, E define a mudança de coordenadas de uma câmara para a outra.

A matriz Fundamental F é também uma matriz (3×3) de característica (rank) 2 e é uma representação algébrica que por construção contém parâmetros intrínsecos do sistema. Se os parâmetros intrínsecos não forem retirados dos pontos correspondentes, a Eq. 3.3 é reescrita desta forma:

$$x_2^T F x_1 = 0 \quad (3.5)$$

A relação da matriz Fundamental com a matriz Essencial é dada pela equação Eq. 3.6:

$$E = K_1 F K_2' \quad (3.6)$$

A estimação de E é então realizada retirando os parâmetros intrínsecos ao pontos correspondentes e resolvendo o sistema 3.7:

$$\begin{cases} \min_E \sum_{j=1}^n (\hat{x}_{2j}^T E \hat{x}_{1j})^2 \\ \|E\|_E = 1 \end{cases} \quad (3.7)$$

Este sistema é resolvido usando um conjunto de equações lineares da forma $Ae = 0$ onde e é um vector coluna de nove entradas que contem a solução da matriz E . A matriz de medição A é descrita por, Eq. 3.8:

$$\begin{bmatrix} x_1 x_2 & x_1 y_2 & x_1 & y_1 x_2 & y_1 y_2 & y_1 & x_2 & y_2 & 1 \end{bmatrix}_n \underset{A}{=} \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{21} \\ e_{22} \\ e_{23} \\ e_{31} \\ e_{32} \\ 1 \end{bmatrix} = 0 \quad (3.8)$$

e

e tem como entradas as coordenadas dos pontos correspondentes \hat{x}_1 e \hat{x}_2 . A matriz Essencial E é uma reformulação do vector solução e , cujo valor é determinado à parte do factor escala, Eq. 3.9. Isto quer dizer que o sistema tem ainda a restrição de $\|e\| = 1$.

$$E = \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & 1 \end{bmatrix} \quad (3.9)$$

3.2 Factorização da Matriz Essencial e Reconstrução

A factorização da matriz Essencial permite obter os parâmetros rotação R e translação t da matriz da câmara entre cada par de imagens. O algoritmo de 8 pontos tem a grande vantagem de ser linear e portanto rápido e de fácil implementação. Se forem usados apenas oito pontos correspondentes a solução é resolver um sistema linear de equações, se forem usados mais do que oito pontos então é necessário resolver o método dos mínimos quadrados para encontrar a matriz mínima. É aplicada a factorização SVD que subdivide a matriz Essencial em três matrizes Eq.3.10, onde U e V são matrizes ortogonais, D é matriz diagonal. A matriz diagonal mostra que E é de rank 2 pois $D = \text{diag}(1, 1, 0)$.

$$[U, D, V] = \text{SVD}(E) \quad (3.10)$$

$$E = UDV^T \quad (3.11)$$

$$E = RS \quad (3.12)$$

A matriz Essencial pode ser também decomposta na multiplicação da matriz rotação R com a matriz anti-simétrica S . A matriz rotação R é determinada segundo a Eq.3.14 e a matriz anti-simétrica S é determinada segundo a Eq.3.15, onde Z e Y são duas matrizes descritas a seguir Eq.3.13. A matriz S contém o vector t também, Eq.3.16. O vector t pode ser encontrado de duas maneiras. Por S , Eq.3.16, ou através da terceira coluna de U , Eq.3.18.

$$Z = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad Y = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (3.13)$$

$$R = UYV^T \quad \text{ou} \quad R = UY^TV^T \quad (3.14)$$

$$S = VZV^T \quad (3.15)$$

$$S = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \quad (3.16)$$

Vector translação tirando da matriz anti-simétrica S:

$$t = (t_x, t_y, t_z) \quad (3.17)$$

Vector translação tirando da terceira coluna de U:

$$t = U(:, 3) \quad (3.18)$$

A factorização da Essencial origina duas hipóteses para a matriz rotação, e origina um vector translação com norma igual a 1. Desta forma, existem quatro combinações possíveis $[R, t]$. Para encontrar a estimativa correcta é necessário reconstruir um ponto tridimensional e analisar se é positiva a profundidade desse ponto em relação às duas câmaras, isto é, se o ponto está à frente das duas câmaras, Fig.3.2.

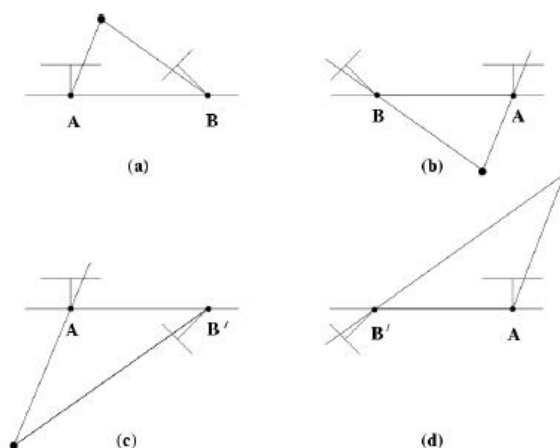


Figura 3.2: Quatro possíveis soluções da estimação de E. Entre a solução da esquerda e da direita, a baseline é invertida. Entre as soluções de cima e de baixo, a câmara B é rodada 180°. Notar que só em (a) é que o ponto está a frente das duas câmaras.

3.3 Normalização do Algoritmo dos 8 pontos

Um melhoramento possível ao algoritmo de 8 pontos é o algoritmo normalizado de 8 pontos, que consiste em equalizar os pontos correspondentes. Ou seja, é diminuir os valores absolutos das coordenadas de forma a que os seus erros (originados por uma correspondência imperfeita) também sejam menores. É redefinir os pontos de forma a que a localização do ponto de coordenadas $(1, 1)$ esteja no centro da imagem, em vez de estar no canto superior esquerdo, e ainda é aplicado um escalamento de forma a que a média das normas dos pontos passa a ser raiz de dois.

Esta normalização provoca uma matriz de medição Eq.3.8 mais equilibrada, pois as colunas onde existe uma multiplicação deixam de ter valores absolutos muito maiores que as restantes. Também desta forma e nestas colunas, a importância do erro nas correspondências é diminuído.

3.4 Estimação da Localização da Câmara

Como já foi referido a matriz E estima a alteração da localização da câmara entre duas imagens, portanto considera-se que a câmara da primeira imagem tem rotação e translação nulas $[R, t] = [I, 0]$ para que cada matriz E estimada seja a alteração da câmara entre imagens. Para encontrar cada $[R, t]$ é primeiro construída uma matriz de medição constituída pelas coordenadas dos pontos correspondentes, Eq.3.8. Depois aplica-se o método dos mínimos quadrados para determinar a matriz mínima segundo o sistema 3.7. A seguir é feita uma decomposição SVD da matriz essencial e um teste de profundidade para então encontrar matriz rotação e o vector translação. O caminho percorrido pela câmara é determinado pela multiplicação das rotações e pela soma das translações. Cada novo R estimado é multiplicado pela rotação entretanto acumulada e cada novo t estimado é somado com a translação entretanto acumulada. É de notar que para se visionar o caminho percorrido pela câmara e as várias rotações da mesma é necessário mudar de referencial. A estimação da matriz essencial determina os parâmetros da matriz da câmara, ou seja, determina a alteração que a câmara sofreu analisando uma alteração dos pontos do mundo no referencial da própria câmara. Com já foi referido a factorização da Essencial origina um vector translação com norma igual a um. Esta restrição impõe que a estimação seja desprovida de escala. Para encontrar o escalamento do caminho é necessário comparar distâncias entre pontos tridimensionais. Em cada sistema estéreo é feito uma reconstrução de dois pontos tridimensionais para que no próximo sistema estéreo se possa calcular o escalamento. Este escalamento divide a distância anterior pela actual e multiplica pelo escalamento acumulado. Desta forma além do sentido do movimento o vector translação representa o caminho percorrido pela câmara.

3.5 Resultados

Nesta secção realiza-se primeiro um estudo do efeito do ruído existente na imagem sobre a estimação do movimento de uma câmara. Seguidamente realizam-se duas experiências de navegação e reconstrução de uma estrutura tubular.

3.5.1 Estudo do efeito do ruído na estimação de movimento de translação

Para estudar a influência do ruído a nível da imagem sobre a estimação de movimento de translação consideramos um cenário virtual em que uma câmara observa uma estrutura 3D à sua frente e realiza vários movimentos de translação. A estrutura 3D observada é um cubo representado por pontos aleatórios distribuídos uniformemente. O cubo tem meio metro de aresta e está localizado a dois metros da câmara, centrado com o eixo óptico. Os movimentos de translação são realizados de forma a amostrar todas as direcções possíveis. Considerando que o eixo óptico da câmara intersecta os pólos norte e sul de uma esfera de raio um metro, que envolve a câmara, os movimentos de translação são realizados para longitudes entre -180° e 180° (saltos de 15°) e latitudes entre -80° e 80° (saltos de 30°). Ver figura 3.3(b).

A figura 3.3(c) mostra estimativas de movimento de translação considerando um ruído de 1 pixel¹ nos pontos correspondentes, sendo a resolução das câmaras de 640*480 pixels. A figura mostra que para este nível de ruído as estimativas de localização estão próximas das latitudes e longitudes de referência, bem assim como as estimativas da orientação indicam poses verticais estipuladas na experiência.

¹Ruído de 1 pixel designa de forma abreviada ruído aleatório uniforme entre 0 e 1 pixel. Esta forma abreviada será utilizada seguidamente com o objectivo de simplificar a descrição das várias experiências realizadas.

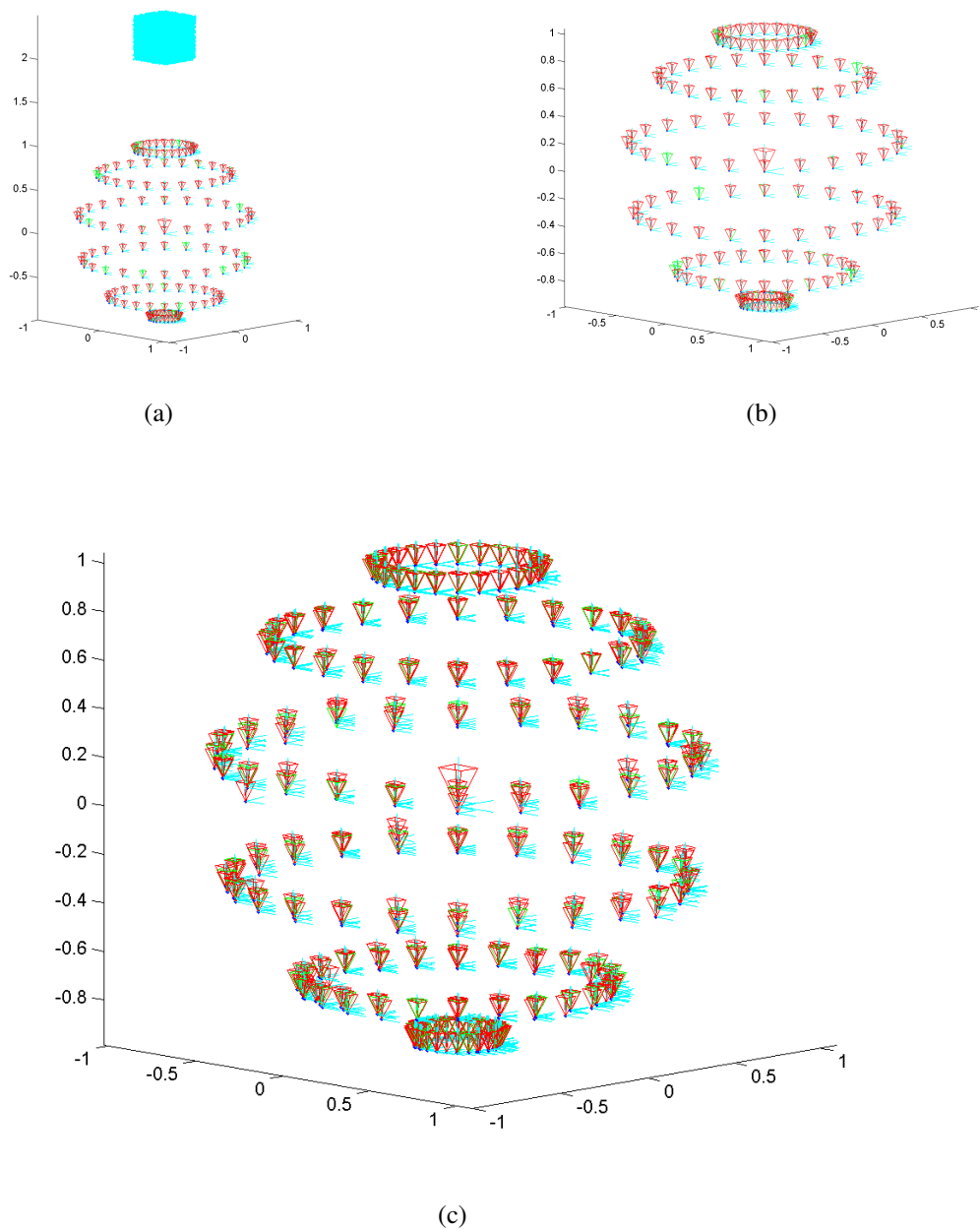


Figura 3.3: Estimação de movimento de translação: (a) localização do cubo de pontos 3D utilizados na estimação do movimento; (b) translações da câmara desde o ponto central (câmara no centro com dimensão maior) para várias latitudes e longitudes. (c) estimação do movimento da câmara considerando ruído de 1 pixel nos pontos correspondentes.

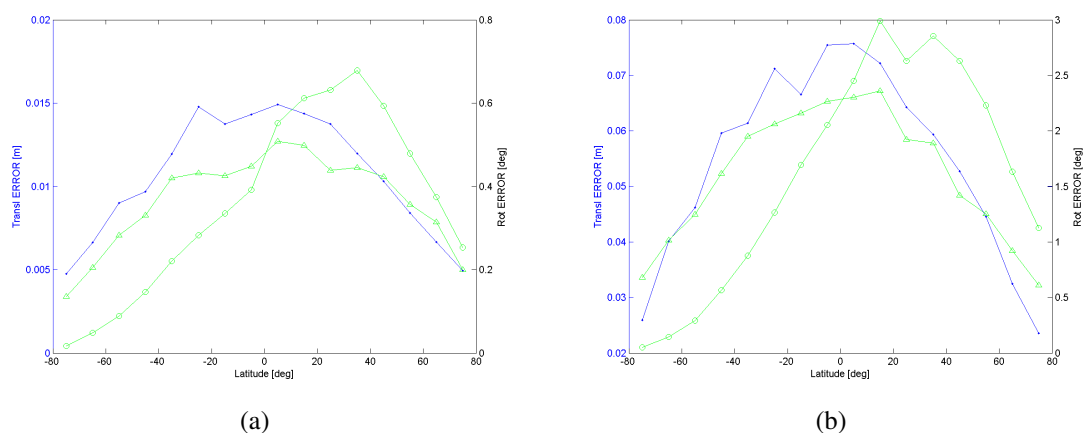


Figura 3.4: Erro médio na estimação do movimento de translação. Erro médio para (a) ruído de 1pixel e (b) ruído de 5pixels nos pontos de imagem correspondentes.

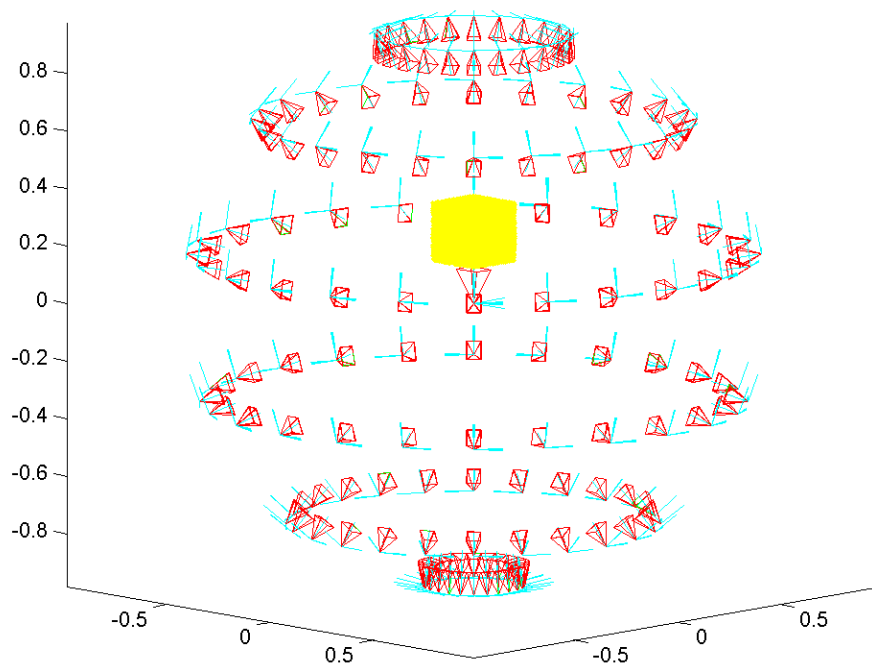
No sentido de avaliar mais precisamente os erros de estimação de localização e de pose, apresentam-se dois gráficos de análise de erro, Figs. 3.4(a) e (b). O primeiro gráfico de análise do erro corresponde à experiência ilustrada na Fig. 3.3(c) e o outro a uma experiência semelhante na qual o nível de ruído foi aumentado para 5 pixels nos pontos de imagem correspondentes. O erro da localização indicado nos gráficos é a distância euclidiana entre a localização exacta e a localização estimada da câmara. No caso da pose, o erro é visualizado por dois ângulos, nomeadamente o ângulo de rotação do eixo óptico e o ângulo de rotação do eixo X da câmara. Para 1 pixel de ruído, os erros médios de localização e de pose da câmara são de aproximadamente 1cm e de 0.5° respectivamente. Para 5 pixel de ruído, os erros médios de localização e de pose aumentam, passando a ser de aproximadamente 5cm e 2.5° . Verifica-se que para latitudes próximas de zero (equador da esfera) o erro na estimativa é maior do que nas outras latitudes. Este efeito está relacionada com a pose relativa do movimento e da estrutura 3D observada para realizar a estimação do movimento. Notar que os gráficos de erro são função somente das várias latitudes, uma vez que a simetria do problema torna praticamente irrelevante a variação da longitude.

3.5.2 Estudo do efeito do ruído na estimação de movimento de rotação e translação

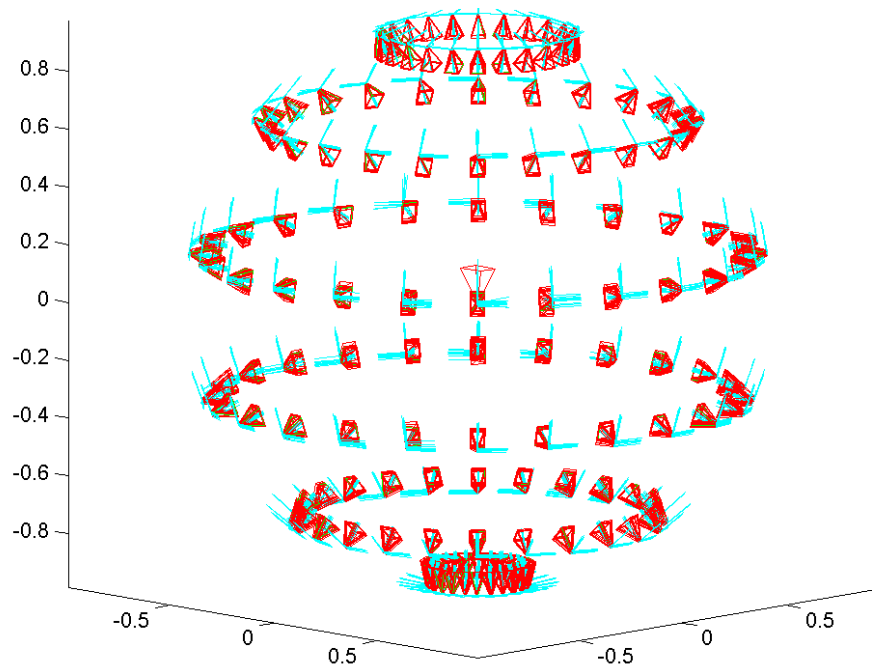
Neste estudo a câmara move-se de uma posição central para localizações a um metro de distância e, para cada translação, realiza uma rotação de forma a observar a posição de partida. O conjunto das localizações da câmara após translação distribuiu-se assim de novo sobre uma superfície esférica (ver Fig 3.5(a)). Os pontos 3D do cenário virtual utilizado para estimar o movimento estão de novo limitados a um volume cúbico, no entanto neste caso o cubo situa-se dentro da esfera de raio unitário. O cubo está centrado com o eixo óptico da câmara na sua posição original no centro da esfera, a uma distância de quinze centímetros, e tem vinte centímetros de lado.

A figura 3.5(a) ilustra os movimentos de translação e rotação, bem assim como o cubo contendo pontos 3D usados para estimar o movimento. Na figura 3.5(b) podem então ser vistas estimativas de movimento com ruído de 1 pixel nos pontos correspondentes nas imagens.

Os gráficos da Fig.3.6(a) mostram o erro do movimento estimado para câmaras com rotação e translação, e com ruído de 1pixel nos pontos correspondentes. Os gráficos da Fig.3.6(b) mostram o erro do movimento estimado para câmaras também com rotação e translação mas com ruído de 5pixels nos pontos correspondentes. Comparando



(a)



(b)

Figura 3.5: Estimação de movimento compreendendo simultaneamente translação e rotação: (a) rotações e translações da câmera desde o ponto central (câmera no centro com dimensão maior) para várias latitudes e longitudes, e cubo de pontos 3D utilizados na estimativa do movimento.; (b) Estimação do movimento da câmera com ruído de 1 pixel nos pontos correspondentes.

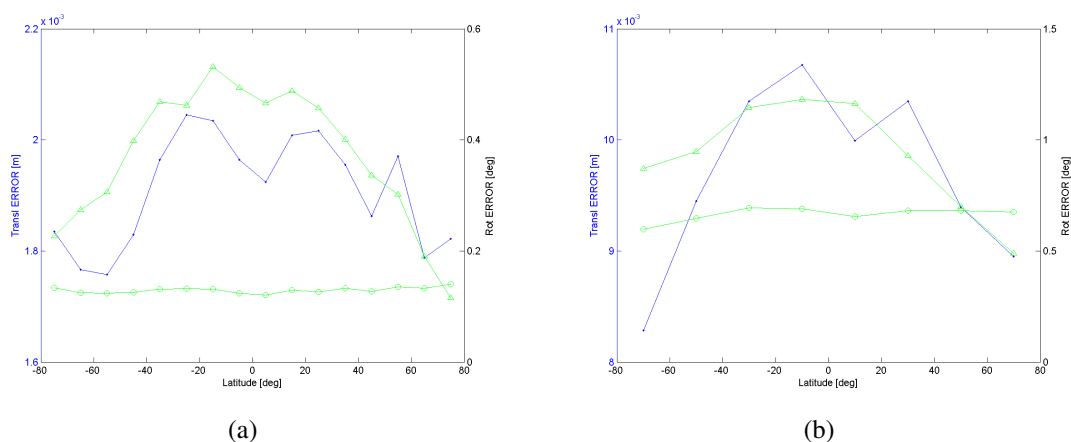


Figura 3.6: Erro médio na estimação do movimento translação e rotação. Ruído de 1pixel (a) e de 5pixels (b) nos pontos correspondentes.

com a Fig.3.4, nota-se também um erro maior para latitudes próximas de zero. No caso de um 1 pixel de ruído e por cada metro de movimento, o erro médio da localização da câmara é de dois milímetros. No caso de 5 pixels de ruído é de dez milímetros. Para 1 pixel de ruído, o erro da pose da câmara em relação ao eixo óptico é de quinze centésimas de grau e em relação ao eixo X é de quatro décimas de grau. Para 5 pixels de ruído o erro da pose da câmara em relação ao eixo óptico é de sessenta e cinco centésimas de grau e em relação ao eixo X é de um grau.

3.5.3 Experiência 1 - Reconstrução de uma estrutura composta por dois troços cilíndricos

Na primeira experiência é reconstruído uma estrutura tubular composta por dois troços cilíndricos, representando um percurso plano seguido de uma pequena subida. Na figura 3.7(a) são mostradas três exemplos de imagens do interior do tubo. A partir de pares de imagens consecutivas da estrutura tubular, são realizadas estimativas das várias localizações da câmara ao longo do percurso (ver Fig.3.7(b)). Na figura 3.7(c) mostram-se estimativas de localização das câmaras existindo neste caso ruído de 5 pixel sobre os pontos correspondentes. Nota-se que o ruído ao longo das estimativas degrada cumulativamente a estimativa da localização da câmara.

Dados pontos correspondidos ao longo do tempo e as localizações da câmara, é então realizada a reconstrução. O facto de estar a experiência decorrer dentro de um ambiente simulado constituído por pontos, tendo a estrutura global pouca curvatura, permite visualizar todo o cenário em cada imagem. Assim a reconstrução pode ser realizada praticamente de forma completa somente com imagens do início do percurso. Na figura 3.7(d) são usadas apenas as duas primeiras imagens de todo o conjunto de imagens. Nota-se que sem ruído a reconstrução é perfeita. Quando é introduzido ruído, como indicado para a figura 3.7(c), nota-se uma crescente má reconstrução ao longo do tubo, figura 3.7(e). Este resultado é esperado pois a representação do final do tubo é de menor qualidade, em número de pixels na imagem, do que o início do tubo. Pontos mais longe da câmara são mais sensíveis ao ruído.

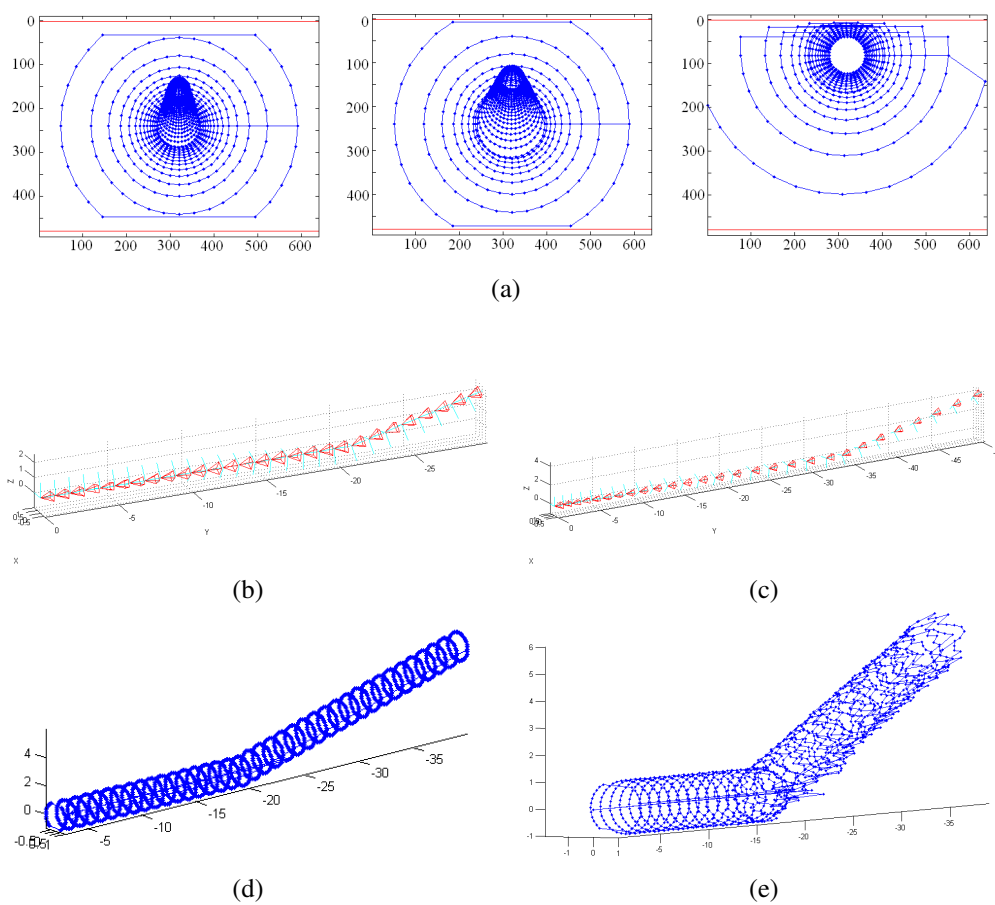


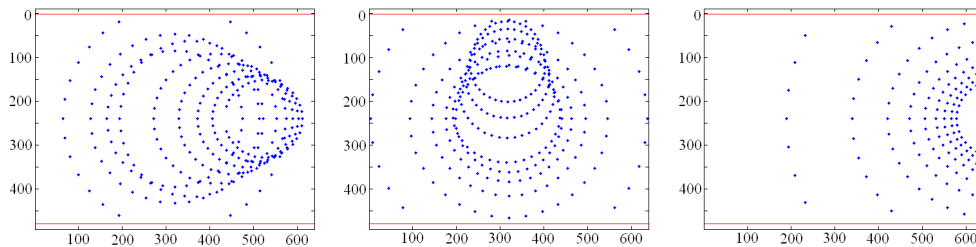
Figura 3.7: Experiência 1, estimação do movimento da câmara num tubo com duas secções: (a) imagens do interior; (b,c) posição estimada para observações sem ruído e com ruído; (d,e) estrutura reconstruída respectivamente para (b) e (c).

3.5.4 Experiência 2 - Reconstrução de uma estrutura tubular considerando profundidade de visão limitada

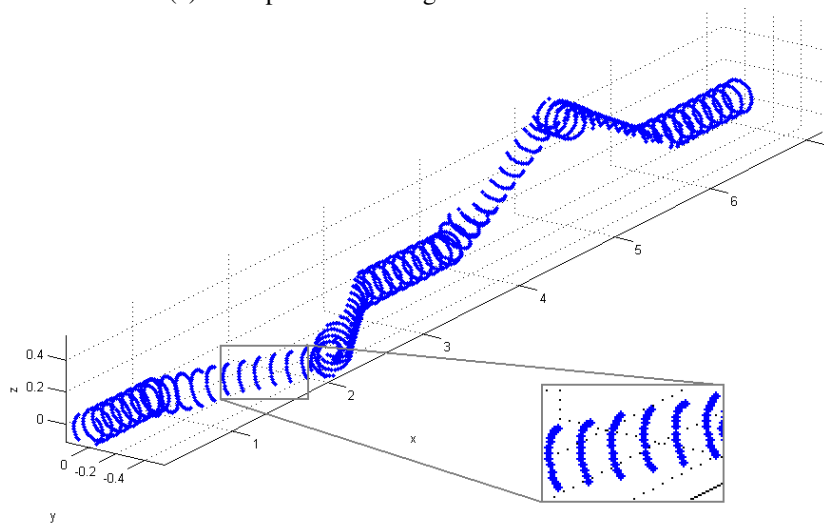
Na segunda experiência é também reconstruída uma estrutura tubular com vários troços. É novamente simulada uma câmara a navegar dentro do tubo representado por pontos. Considera-se em particular que a câmara viaja sobre o eixo do tubo estando sempre orientada no sentido positivo do eixo Oz. Esta restrição determina que os parâmetros extrínsecos da câmara se resumem a uma translação. Nesta experiência é também colocada a hipótese de que a câmara observa somente até uma profundidade máxima. Aproxima-se assim a situação real de uma câmara que por limitações de visibilidade ou de resolução vê a cada momento só uma parte da estrutura tubular.

A Fig.4.1(a) mostra imagens adquiridas no cenário, e ilustra em particular a visualização parcial do tubo (somente uma face) devido à orientação da câmara e à limitação da profundidade máxima visível.

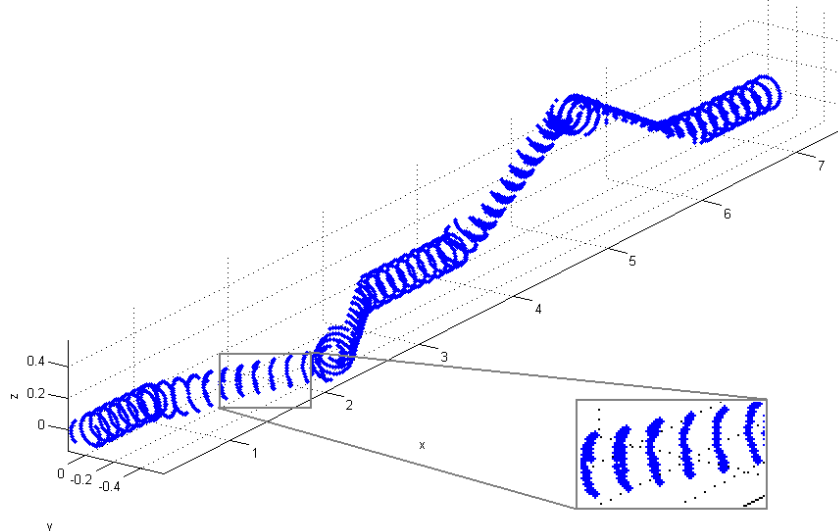
Nas Figs.3.8(b) e (c) mostra-se a estrutura tubular reconstruída a partir de todas as imagens (640x480 pixels) afectadas por um ruído 0.01 pixel, considerando-se no segundo caso imagens de maior resolução (3000x2000 pixels) mas também com mais ruído (uniforme, 0.1 pixels). As restrições de visibilidade implicam regiões da estrutura tubular não reconstruídas. Os detalhes indicados nas Figs.3.8(b) e (c) mostram precisamente secções da estrutura tubular reconstruídas somente na face direita devido ao facto de a câmara não ter obtido imagem da face esquerda. Nota-se ainda nestes detalhes uma reconstrução mais ruidosa associada às imagens mais ruidosas.



(a) Exemplo de três imagens do interior do tubo.



(b) 640x480pixels ; Erro máximo na correspondência: 0.01pixels



(c) 3000x2000pixels ; Erro máximo na correspondência: 0.1pixels

Figura 3.8: Experiência 2, estimação de movimento e de estrutura limitando a profundidade máxima do campo de visão: imagens obtidas no interior da estrutura, 640x480pixels (a) e respectiva reconstrução para ruído de 0.01 pixels; (c) reconstrução utilizando imagens de maior resolução, 3000x2000 pixels, com um ruído de 0.1 pixels.

Capítulo 4

Modelo do Tubo

4.1 Introdução

Com a reconstrução tridimensional do tubo e a estimação do caminho da câmara é possível planificar para uma imagem 2D a superfície interna do tubo. Neste capítulo é descrito um modelo que organiza os dados tridimensionais e que os prepara para a planificação. Este modelo reduz o tubo reconstruído a um conjunto de secções tubulares que são definidas por círculos. Este modelo também permite criar a própria superfície tubular a ser usada na planificação.

4.2 Modelo Tridimensional

O modelo tridimensional do tubo é constituído por uma lista de círculos. Cada secção tubular é definida por um círculo que tem os seguintes três parâmetros: (a) ponto central m_i (b) raio r_i e (c) vector normal ao plano do círculo v_i . Estes círculos estão representados graficamente na Fig.4.1(a). e na Fig.4.1(b) é representado uma estrutura tubular simples contendo vários círculos definidos ao longo de sete troços cilíndricos.

Este modelo pode ser usado tanto para criar estruturas, como a mostrada na Fig.4.1(b), como para obter imagens perspectiva simuladas do interior do tubo, que servirão para a representação mínima da estrutura tubular. Simular a navegação da câmara no interior do tubo envolve especificar os parâmetros intrínsecos da câmara e o seu movimento. A representação mínima da estrutura tubular é descrita na secção 4.3.

4.3 Representação Mínima da Estrutura Tubular

O modelo mínimo de representação tridimensional de uma superfície tubular pretende fazer uma pré análise da estrutura do tubo para simplificar a planificação da superfície para uma imagem. Este modelo resume a reconstrução da estrutura tubular a uma sequência de secções circulares definidas por círculos.

Com os pontos tridimensionais reconstruídos o modelo anteriormente definido é aplicado. Este modelo resume-se em três passos: primeiro é definida a melhor superfície cilíndrica que representar cada nuvem de pontos (reconstruída a partir de cada par de imagens); depois são redefinidas as superfícies para tirar as zonas cilíndricas

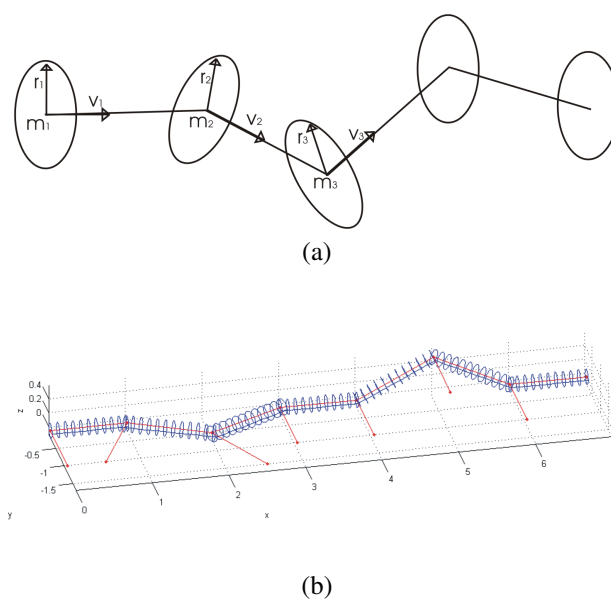


Figura 4.1: (a) Modelo do tubo. (b) Exemplo de um tubo.

sobrepostas; por fim as superfícies são reajustadas de modo a construir um caminho contínuo do eixo da estrutura tubular.

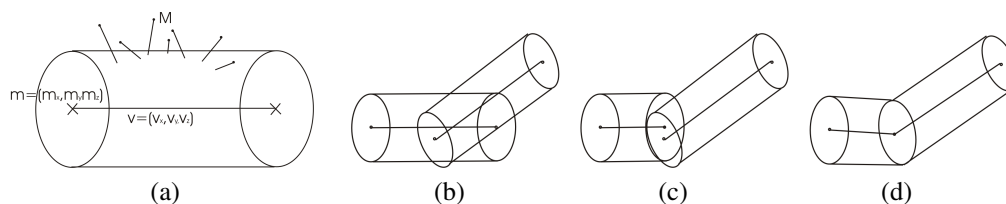


Figura 4.2: Ajuste tridimensional para cilindros: (a) Ajustar uma nuvem de pontos a um cilindro (b) remover zonas sobrepostas de superfícies cilíndricas (c) reajustar os cilindros para construir um eixo coerente entre todos os cilindros (d) ajuste final dos cilindros.

A Fig.4.2a representa como é feito o ajustamento de uma superfície cilíndrica a uma nuvem de pontos. Este ajustamento é feito encontrando a superfície cilíndrica que detenha a menor distância média a todos os pontos da nuvem de pontos. Sendo $\{M_i\}$ os pontos tridimensionais reconstruídos, sendo (m, v, r) o ponto central, a direcção do eixo e o raio, então a optimização da superfície é conseguida através da Eq.4.1.

$$\hat{\theta} = \arg_{\theta} \min \sum_i (\| (M_i - m) - \text{proj}_v (M_i - m) \| - r)^2 \quad (4.1)$$

onde θ contém apenas os graus de liberdade mínimos de (m, v, r) , ou seja $\theta = [m_x \ m_y \ v_x \ v_y \ r]$ e onde $\text{proj}_v(x) = x^T v v / \|v\|^2$. A coordenada z não é estimada porque para o círculo que pretende representar a superfície, é indiferente a sua localização segundo o eixo dos zz . Dado um tubo que se forma ao longo deste eixo a terceira coordenada é encontrada usando o centro de massa de cada nuvem de pontos.

A remoção de zonas sobrepostas entre superfícies cilíndricas é apenas uma redução do comprimento de cada superfície de forma a não se sobrepor com a próxima, Fig.4.2c.

O conjunto das superfícies cilíndricas limpas de sobreposição não têm uma intercepção correcta dos seus eixos. Assim é necessário definir um caminho contínuo entre todos os eixos. Redefinindo os círculos final e inicial de superfícies cilíndricas consecutivas para só um círculo de compromisso entre os dois, Fig.4.2d, é possível encontrar um eixo contínuo de toda a superfície tubular. Estes novos círculos têm rotação e raio independentes o que provocam superfícies não só cilíndricas, mas também cónicas. Na Fig.4.3 pode ser visto um exemplo gráfico de tal situação.

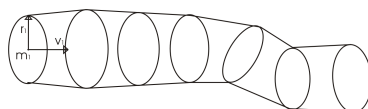


Figura 4.3: Exemplo de secções cónicas

4.4 Resultados

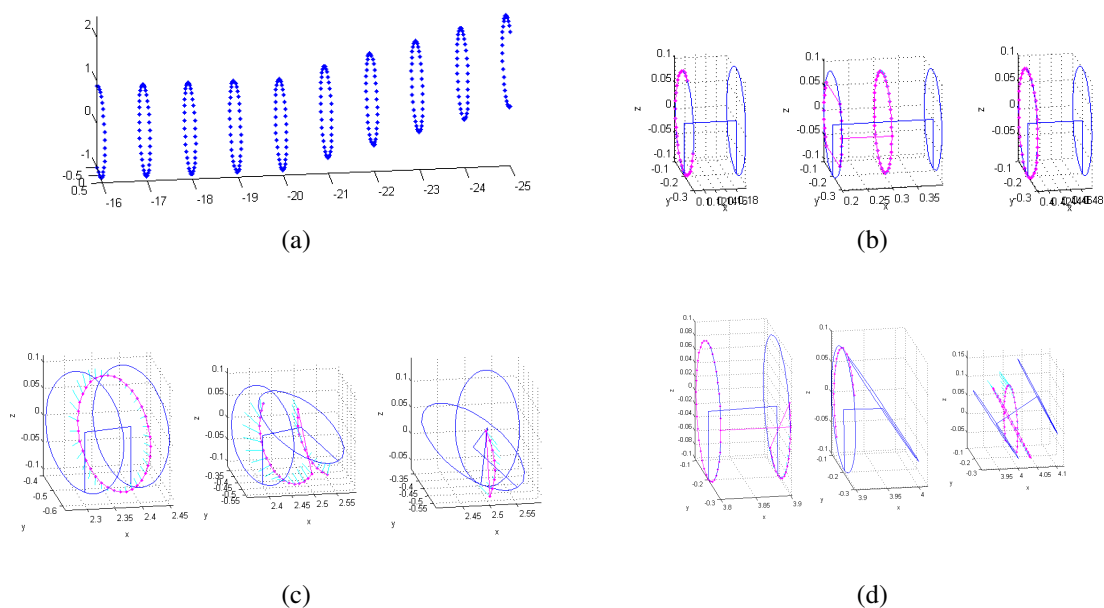


Figura 4.4: Ajustamento das superfícies cilíndricas: (a) Pontos reconstruídos, (b)(c)(d) Otimização das ligações entre as secções, *Azul* círculos do modelo tridimensional, *Magenta* Pontos tridimensionais originais, *Azul claro* distância entre os pontos originais e a superfície tubular.

A Fig.4.4 representa o re-ajustamento das superfícies cilíndricas depois da remoção das zonas sobrepostas. Este processo de otimização re-ajusta os círculos definidos pelo modelo. É de notar as dificuldades de otimização nos casos da Fig. 4.4c e Fig. 4.4d. Mantendo fixo o primeiro e o quarto círculo, os dois do meio são reajustados de forma a suprimir estimações cilíndricas erradas, ou seja, os dois círculos são reajustados de modo a criar um superfície tubular continua e sem variações abruptas. Na iteração seguinte são os círculos dois e cinco que são fixos, assim o terceiro é novamente reajustado.

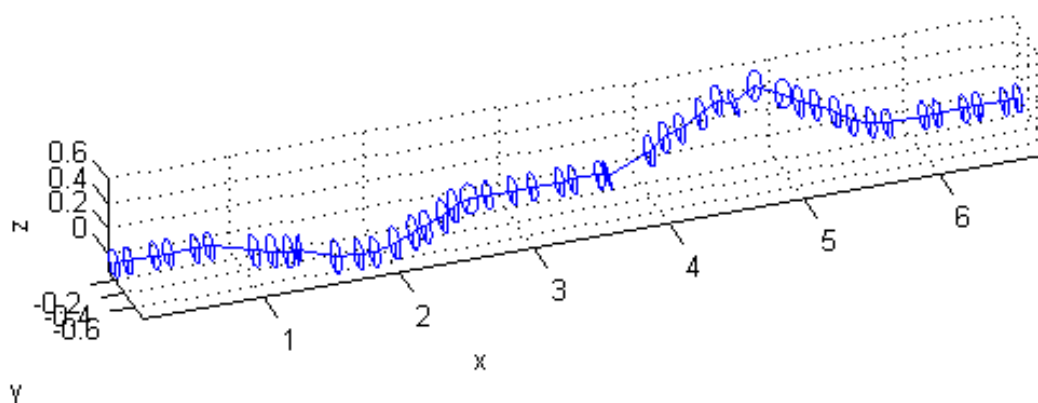


Figura 4.5: Reconstrução da superfície tubular aplicando o modelo.

A razão de se utilizarem quatro círculos deve-se ao facto de ser um meio termo entre suavização de más estimações e demasiada suavização do tubo. No caso de se reajustar apenas um círculo nota-se uma intensificação de superfícies cónicas (pouca robustez aos erros). No caso se reajustar 3 círculos, se se estiver perante uma boa estimacão de uma curvatura esta vai ser erradamente suavizada.

A Fig.4.5 mostra a reconstrução da superfície tubular. Nota-se uma grande semelhança com a estrutura original apesar da remoção de sobreposições das superfícies cilíndricas e dos processos de reajustamento.

Capítulo 5

Abertura do Tubo

5.1 Introdução

Com a simplificação imposta pela representação mínima referida no capítulo anterior, a abertura do tubo para uma imagem bidimensional está facilitada. Esta abertura é um processo de *dewarp* que transforma a informação tridimensional da superfície tubular numa imagem panorâmica que mostra o interior de todo o tubo. O processo de abertura do tubo está facilitado porque o modelo de representação mínima origina uma lista de círculos representativos da estrutura tubular que torna este dewarp numa mudança de coordenadas.

5.2 Dewarping

O processo de abertura do tubo é uma transformação de coordenadas polares para coordenadas cartesianas. Os círculos do modelo de representação mínima são abertos para formar as colunas que irão preencher a imagem final. Para ajudar a este processo é definido uma linha de chão que liga cada círculo. Esta linha de chão representa exactamente o chão do tubo, ou seja, o ponto mais baixo na direcção do eixo vertical. Esta linha de chão é a linha média da imagem final, e é perpendicular às colunas que são obtidas dos círculos tridimensionais. Mas com este processo de redução tridimensional para bidimensional alguma informação terá de ser perdida. Esta imagem final dá uma ideia rápida da superfície interna do tubo mas altera as suas proporções. Nesta imagem as curvas e as mudanças de raio são difíceis de avaliar.

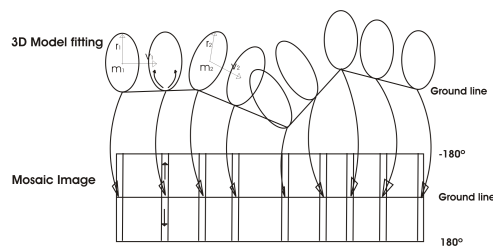


Figura 5.1: Dewarp do modelo de representação mínima do tubo

5.3 Resultados

Um caso simples desta abertura é quando o tubo é perfeitamente cilíndrico e a câmara está perfeitamente posicionada e alinhada com o eixo. Este caso foi simulado e os resultados pode ser vistos de na Fig. 5.2. As imagens perspectiva são usadas no modelo do tubo e como era de esperar existe sobreposições de informação. Para remover esta sobreposição é usado a Homografia. A estimação da homografia é referida no capítulo 2.6.

Nas Fig.5.3a e Fig.5.3c e Fig.5.6a e Fig.5.6c são mostradas imagens perspectivas do interior do tubo. Nas Fig.5.3a e Fig.5.6-direita são mostradas as imagens das secções tubulares reconstruídas. É de notar a sobreposição das imagens dewarped. O resultado final do processo origina a Fig.5.4 e a Fig.5.7 que mostra que o interior do tubo tinha representado no primeiro conjunto de imagens um calendário, e no segundo conjunto, um mapa mundo. Os rectângulos verdes mostram, respectivamente, a localização das imagens dewarped da Fig.5.6b,d.

Estas imagens perspectivas têm uma resolução de 640x480 pixels.

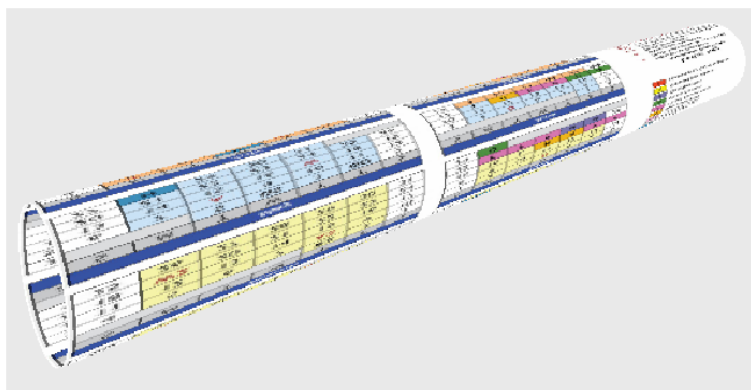


Figura 5.2: Tubo tridimensional

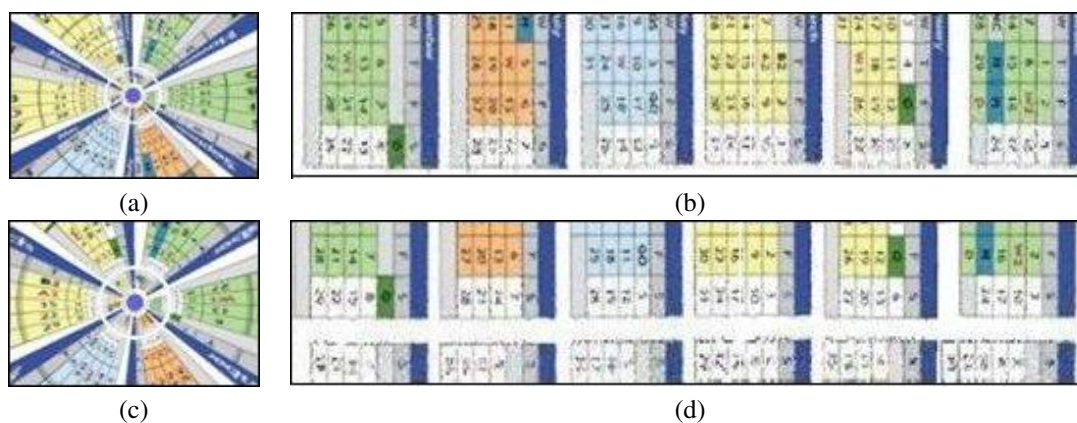


Figura 5.3: Imagens capturadas pela câmara (a,c). Imagens dewarped (b,d).

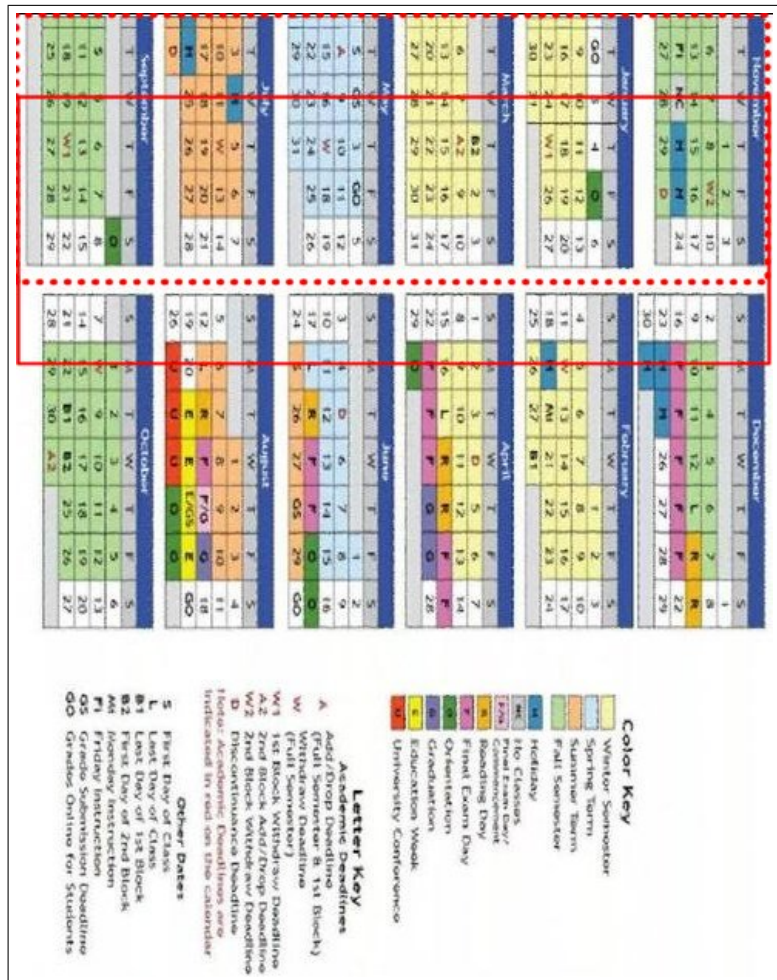


Figura 5.4: Imagem Mosaico final.

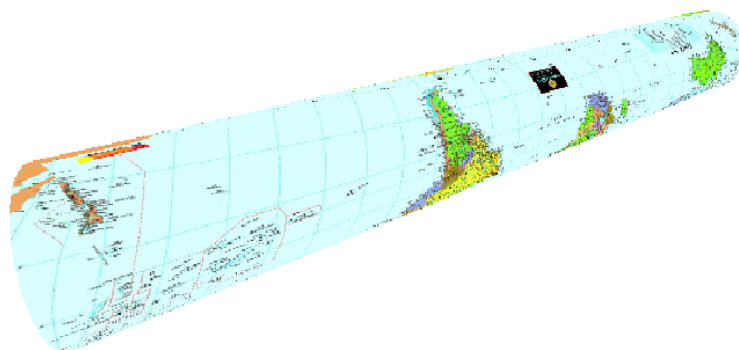


Figura 5.5: Tubo tridimensional

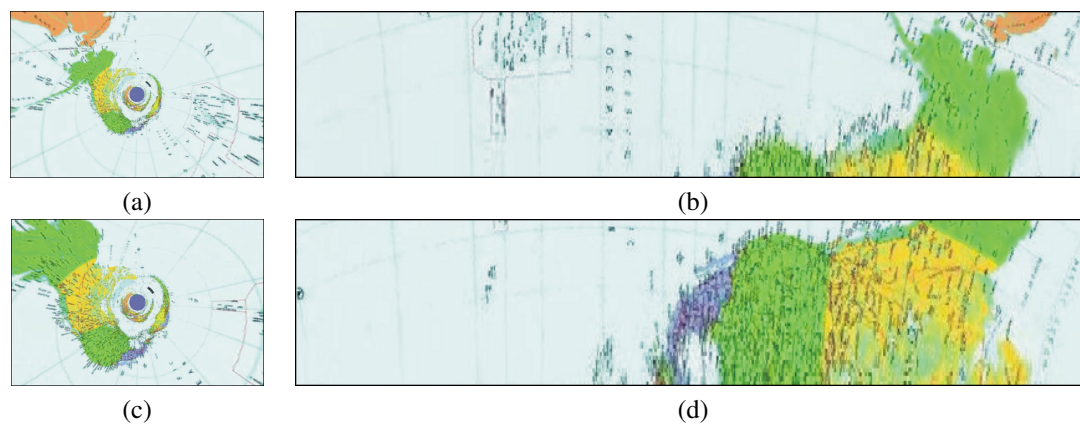


Figura 5.6: Imagens capturadas pela câmara (a,c). Imagens dewarped (b,d).

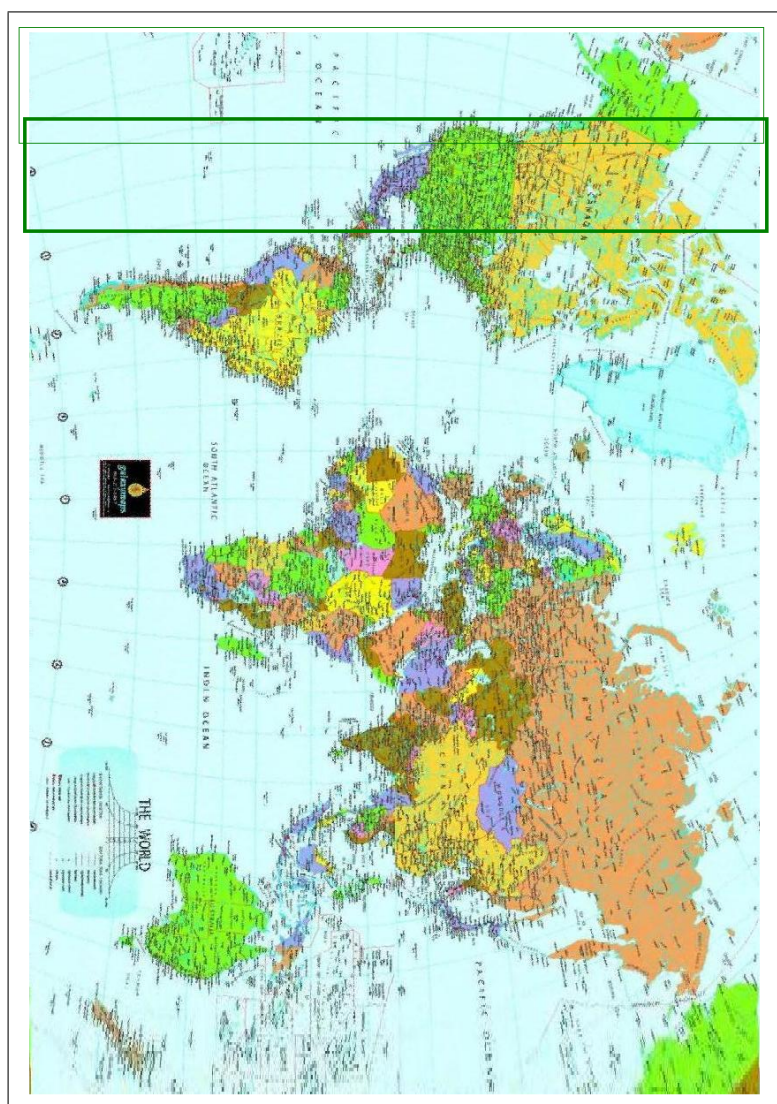


Figura 5.7: Imagem Mosaico final.

Capítulo 6

Conclusões e Trabalhos Futuros

Nesta dissertação estudou-se o problema de inspecção visual do interior de uma estrutura tubular, tendo sido estabelecido como objectivo principal a síntese do vídeo do interior da estrutura em uma única imagem (mosaico). Foram desenvolvidos dois métodos de obtenção do mosaico: um método baseado no mapeamento de cada imagem de coordenadas cartesianas para polares, método este que obtém e combina sequencialmente faixas da estrutura tubular, e um método baseado na reconstrução 3D da estrutura tubular.

A criação de imagens mosaico baseada no mapeamento de coordenadas cartesianas para polares, começa precisamente pela transformação de coordenadas da qual resulta uma faixa representativa de uma planificação de uma secção da estrutura tubular por cada imagem. As várias faixas são então justapostas sequencialmente por intermédio do registo de pontos correspondentes seguido do cálculo da transformação (homografia) que liga as imagens. Este processo de obtenção de imagens mosaico é eficaz se a estrutura tubular tiver um diâmetro aproximadamente constante ao longo de um percurso, que deve ser essencialmente rectilíneo, e a câmara estiver bem alinhada com o eixo da estrutura.

A obtenção do mosaico por via da reconstrução da estrutura 3D permite o estudo de casos mais gerais, por exemplo estruturas tubulares contendo curvas ou movimento da câmara contendo afastamentos do eixo. Neste método a obtenção do mosaico é decomposta em três etapas essenciais: (i) reconstrução 3D de pontos característicos observados em imagens da estrutura, (ii) ajuste de modelos cilíndricos localmente às várias secções da estrutura tubular, e (iii) planificação das várias secções cilíndricas e justaposição para obtenção do mosaico global.

A reconstrução 3D da estrutura tubular envolve o registo de pontos correspondentes entre imagens consecutivas (por exemplo utilizando pontos característicos do tipo SIFT), e posterior estimação de movimento da câmara por estimação e factorização da matriz essencial. O processo de estimação e factorização da matriz essencial revela sensibilidades diferentes para várias direcções de movimento: translações ortogonais em relação ao eixo óptico da câmara revelam-se experimentalmente mais sensíveis provavelmente devido à estrutura estar concentrada à frente da câmara e portanto motivar efeitos de paralaxe menos expressivos nestes movimentos.

Obtida a reconstrução 3D da estrutura tubular, a nuvem de pontos resultante é então substituída por uma forma sintética utilizando um modelo 3D paramétrico. Este modelo descreve a geometria da estrutura através de uma sequência de secções tubulares que por sua vez são definidas por círculos. O ajuste de modelos 3D cilíndricos a secções locais da estrutura tubular, por via de minimização de erro quadrático médio, revelou-se também eficaz: os pontos 3D reconstruídos permitiram em todas as experiências realizadas estimar facilmente raios e eixos das secções cilíndricas locais.

A representação paramétrica da estrutura tubular define intrinsecamente uma planificação. Mais precisamente, a modelação da estrutura baseada numa sequência de círculos permite definir a planificação como a abertura dos círculos pelos seus topos formando assim as colunas da imagem mosaico final. A escolha de uma representação baseada em círculos para a estrutura tubular revelou-se prática e eficiente para resolver o problema por exemplo quando comparada com algumas opções de triangulação 3D que foram consideradas.

A aplicação de câmaras específicas para a inspecção visual do interior de estruturas tubulares continua a ser um aspecto a suscitar grandes expectativas. Um tipo interessante de câmara para o nosso problema é a câmara catadióptica omnidireccional cujo espelho convexo pode ser desenhado de acordo com as estruturas tubulares a inspecionar. Com esta câmara é possível colocar o campo de visão direccionado ortogonalmente ao eixo da câmara e assim observar ortogonalmente a superfície interior da estrutura tubular. Também a resolução pode ser desenhada de forma a ser uniforme sobre a superfície interior da estrutura.

Alternativamente às câmaras catadiópticas omnidireccionais, podem ser usadas também câmaras equipadas com lentes olho de peixe. Neste caso obtêm-se campos de visão que contêm automaticamente a região tubular mais próxima da câmara (superfície tubular lateral em relação à câmara posicionada e alinhada com o eixo da estrutura). Permanece no entanto a questão da resolução ser menor na periferia da imagem. Este pode contudo ser um problema resolvido tecnologicamente, atendendo a que recentemente se tem observado um crescimento extraordinário da resolução das câmaras de vídeo: se há alguns anos a resolução era normalmente inferior a 0.5 megapixel, hoje já é fácil encontrar câmaras de 10 megapixel. O aumento de resolução actualmente disponibilizado comercialmente permite considerar, também como alternativa, a utilização de uma câmara perspectiva simples. As câmaras perspectiva observam na região central da imagem a superfície omnidireccional à frente da câmara, e portanto observam *antecipadamente* os dados, no entanto desperdiçam de alguma forma a resolução disponível na câmara.

No trabalho da dissertação considerámos essencialmente câmaras perspectiva. Apesar de existir alguma perda de resolução nestas câmaras, os métodos propostos são gerais tanto em termos conceptuais como de implementação. As técnicas fundamentais de estimação dos mosaicos, por exemplo nos aspectos de reconstrução da estrutura, são iguais para todas as câmaras de centro de projecção único que sejam calibradas. Propuseram-se portanto metodologias de obtenção de mosaicos do interior de estruturas tubulares, úteis para inspecção visual, mas manteve-se ainda em aberto a possibilidade de melhorar os resultados pelo estudo de geometrias alternativas de câmaras. Com este estudo adicional esperam-se certamente melhorias na qualidade dos mosaicos, minimizando os requisitos dos sensores das câmaras e os requisitos computacionais.

Ainda em relação a trabalho futuro, uma metodologia alternativa, promissora, será uma metodologia de compromisso entre as duas apresentadas na tese. Mantendo a ideia de um mapeamento (transformação) a ser aplicado a cada imagem, e trazendo alguns dos parâmetros mais importantes do método baseado na reconstrução da estrutura tubular, pode ser criada uma metodologia que aplica um mapeamento variável às imagens dependente de um conjunto de parâmetros pequeno mas relevante. Nesta metodologia combinar-se-iam a generalidade da estrutura e do movimento da câmara com a minimização do cálculo computacional envolvido.

Acreditamos que no futuro a inspecção visual será cada vez mais auxiliada por métodos de síntese dos dados observados (vídeo). Representações baseadas em mosaicos poderão então ser utilizadas para a criação de relatórios com a forma de gráficos de uma ou algumas (poucas) páginas que permitam a inspecção visual rápida de túneis, condutas ou mesmo de outras estruturas tubulares, eventualmente orgânicas tais como sistemas digestivo ou respiratório, que tenham formas rígidas ou lentamente modificadas por comparação com o processo de observação.

Bibliografia

- [1] Z. Y. Hu F. C. Wu and F. Q. Duan. 8-point algorithm revisited: Factorized 8-point algorithm. *National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences*, 2005.
- [2] Nuno Gracias and José Santos-Victor. Underwater mosaicing and trajectory reconstruction using global alignment. *VisLab-TR 12/2001, - IEEE OCEANS 2001, Honolulu*, November 2001.
- [3] R. Hartley. In defense of the eight-point algorithm. *IEEE-T. PAMI*, 1997.
- [4] Richard Hartley and Andrew Zisserman. Multiple view geometry in computer vision. *Cambridge University Press*, 2000.
- [5] N. Karlsson, E. Di Bernardo, J. Ostrowski, L. Goncalves, P. Pirjanian, and M. Munich. The vslam algorithm for robust localization and mapping. *Proc. IEEE Int. Conf. on Robotics and Automation*, pages 24 – 29, 2005.
- [6] Longuet-Higgins. A computer algorithm for reconstructing from two projections. *Nature*, page 293:133–135, 1981.
- [7] David Lowe. Distinctive images features from scale-invariant keypoints. *International Journal of computer Vision*, 2004.
- [8] B. Rousso S. Peleg I. Finci A. Rav-Acha. Universal mosaicing using pipe projection. *Institute of Computer Science*, November 2001.
- [9] Luís Ruivo, José Gaspar, and José Santos Victor. Mosaicing the interior of tubular shapes: 3d-model fitting. *13ª Conferência Portuguesa de Reconhecimento de Padrões, 2007* <http://users.isr.ist.utl.pt/~jag/publications/07-recpad-tube-mosaicing.pdf>.

Apêndice A

Testes Adicionais em Estimação de Movimento e Reconstrução

Nesta secção mostram-se resultados de testes adicionais do efeito do ruído e da resolução da imagem sobre a estimação do movimento e da reconstrução.

A Fig.A.1 mostra a estimação da localização da câmara quando é aplicado um ruído aleatório uniforme entre 0 e 1 pixel nos pontos correspondentes e quando os pontos tridimensionais não estão centrados com o eixo vertical da esfera. Nota-se que o erro da estimação é predominante na direcção dos pontos.

A Fig.A.2 mostra um estudo do efeito do ruído nos pontos correspondentes sobre a reconstrução da estrutura tubular. São testadas várias resoluções de imagem para níveis de ruído aleatório uniforme, $\eta = u(0, b)$: $b = 0.001 \times W$ e $b = 0.01 \times W$, onde W é a largura máxima da imagem. A figura mostra que para $b = 0.01 \times W$ a reconstrução já não é considerada aceitável pois existe uma duplicação do comprimento da estrutura e consequente duplicações de informação. Para o nível $b = 0.001 \times W$ todas as reconstruções são adequadas. É de notar que o aumento da resolução da imagem confere uma maior tolerância ao erro de correspondência.

A Fig.A.3 mostra um estudo do efeito do erro em imagens com resolução de 640x480 pixels e com 3000x2000 pixels. No caso de maior resolução a correspondência entre pontos pode ter um erro de uma décima de pixel, obtendo-se ainda nesse caso uma reconstrução visualmente comparável à reconstrução com menor resolução (mas muito menor ruído).

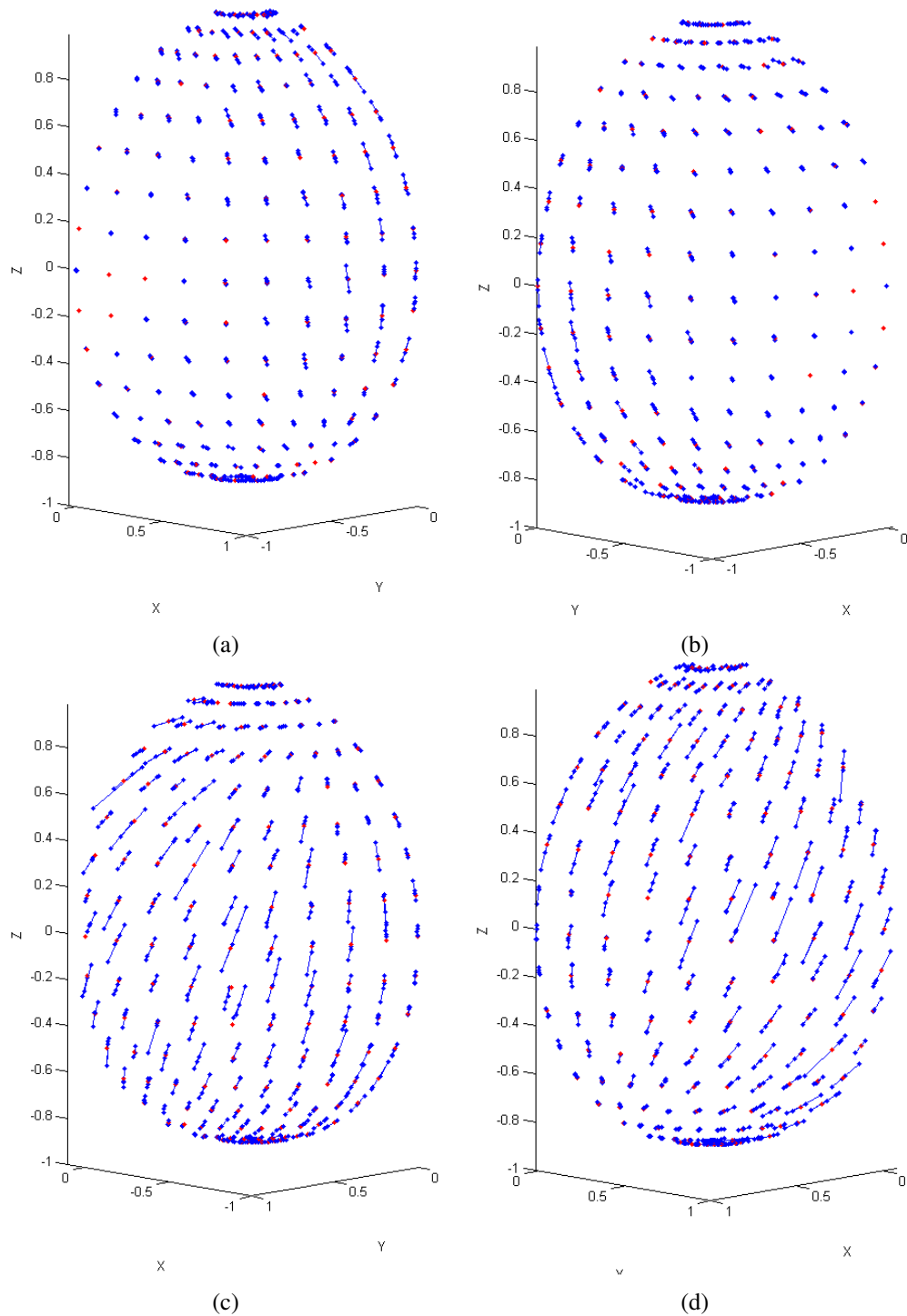


Figura A.1: Localização descentrada dos pontos tridimensionais: (a) longitude 0° - 90° (b) longitude 90° - 180° (c) longitude 180° - 270° (d) longitude 270° - 360° . Ruído de 1 pixel.

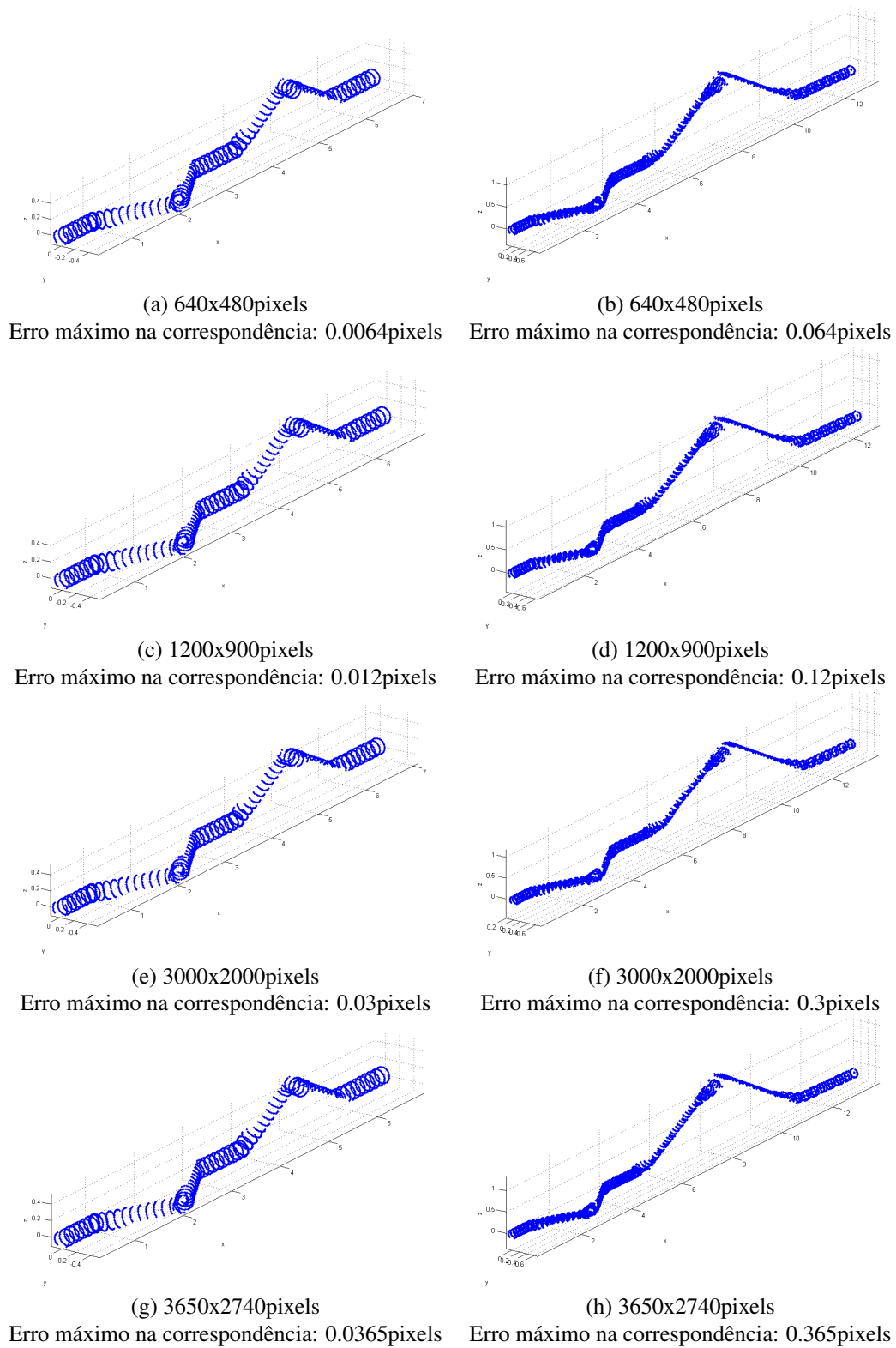
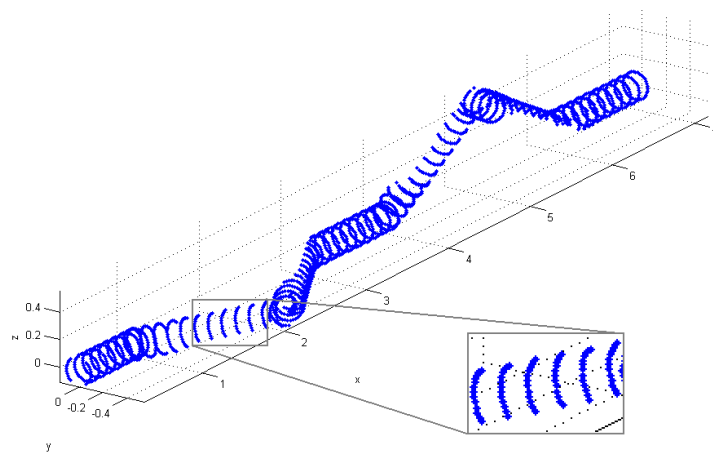
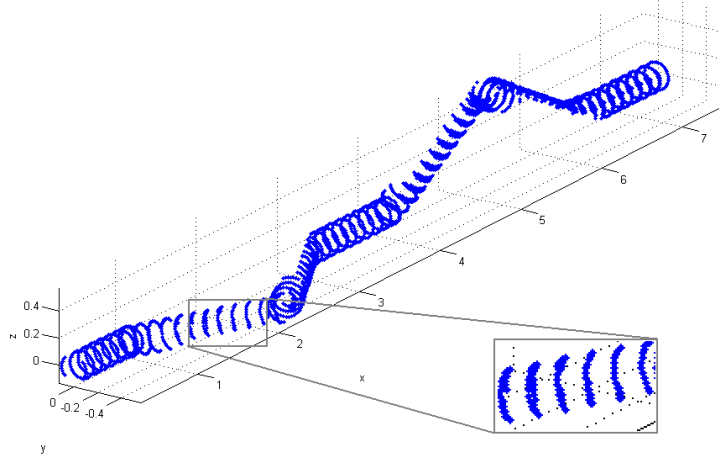


Figura A.2: Estudo da influência do ruído: Resoluções de imagem de 1/3Megapixels (a,b), 1Mp (c,d), 6Megapixels (e,f), 10Megapixels (g,h) com níveis de ruído de 0.001% (a,c,e,g) e de 0.01% (b,d,f,h).



(a) 640x480pixels ; Erro máximo na correspondência: 0.01pixels



(b) 3000x2000pixels ; Erro máximo na correspondência: 0.1pixels

Figura A.3: Estudo da influência do ruído aleatório, combinado a resolução da imagem, sobre a reconstrução.

Apêndice B

Manual do Programa

O código está implementado em Matlab e divide-se em quatro blocos principais: construção de tubo, captura de imagens, reconstrução 3D e ajuste de modelo tubular simples. A Fig.B.1 mostra a relação e as estruturas de dados que são transferidas entre os vários blocos.

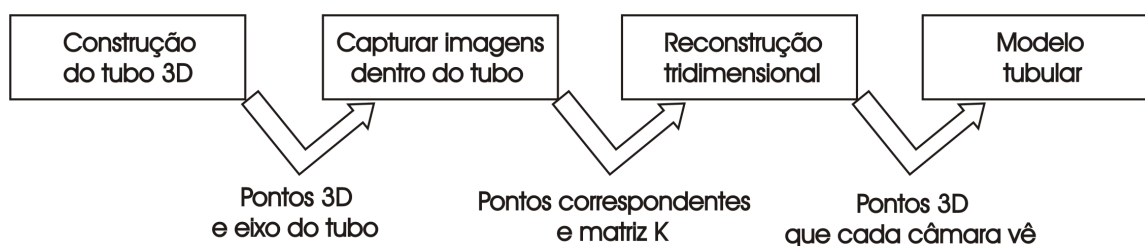


Figura B.1: Sequência dos blocos de código

De seguida, detalham-se para cada bloco as variáveis de entrada, as variáveis de saída, o algoritmo implementado e um exemplo de aplicação da função.

B.1 Construção do Tubo

O processo começa com a construção do tubo. Implementa-se um tubo com várias secções cilíndricas que são unidas entre si para originar um tubo final com curvas. Na Tab. B.1 são descritos as variáveis de entrada e de saída da função *tube_mk*.

O algoritmo consiste em criar secções cilíndricas como as da Fig. B.2 que são dispostas sequencialmente de maneira a construir um tubo com curvas. A forma cilíndrica é estabelecida criando onze circunferências de pontos igualmente espaçadas entre si. Cada secção tem um metro de comprimento e dez centímetros de raio.

Para unir as várias secções, o último ponto do eixo da secção anterior é sobreposto ao primeiro ponto do eixo da secção seguinte. A localização de cada secção cilíndrica é feita através de uma rotação e translação de forma a respeitar o eixo do tubo. Para estes cálculos a função $[xyz2, abc, abc2] = cylinder_at(P(i, :), P(i + 1, :), Ox2)$ constrói e orienta cada secção do tubo, onde P é *xyzCenter*, ou seja, o primeiro ponto do eixo de cada secção, onde $Ox2$ é o vector que irá provocar que cada círculo seja ortogonal ao eixo. A matriz *abc* contém direcção da secção antes da rotação e a matriz *abc2* contém a nova direcção da secção. *xyz2* são os pontos da secção na

		$function [xyz, xyzCenter] = tube_mk(opStr, options)$
Input		
String : opStr 'shCylinder' 'shTube1' 'tubeImgs' *		- Mostra apenas um secção cilíndrica 3D - Mostra o tubo inteiro 3D - Retorna as coordenadas dos pontos e o eixo do tubo
Struct : options * 'shFlagOriginal', 1		- Activa a hipótese de mostrar o tubo
Ouput		
Matrix ($n \times 3$): xyz		- Pontos tridimensionais do tubo
Matrix ($k \times 3$): xyzCenter		- Primeiro ponto do eixo de cada secção do tubo

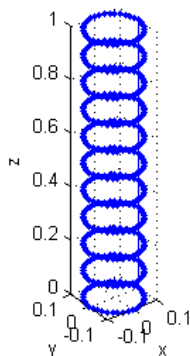
Tabela B.1: Função *tube_mk*: Variáveis de entrada e de saída.

Figura B.2: Uma secção cilíndrica do tubo final.

localização pretendida. A função $[R, t, scale] = procrustes_rt(abc2, abc)$ encontra a rotação R , a translação t e a escala $scale$ pretendidas para alterar a posição de cada secção.

Exemplo de utilização:

```
>> tube_mk (mesmo que tube_mk('shTube1'))
```

Resultado: Visualização do tubo, Fig. B.3.

B.2 Captura de Imagens

Com o tubo construído é então possível adquirir as imagens necessárias ao processo de reconstrução. A função *tube_imgs* usa a função *tube_mk* para construir o tubo e coloca sobre o seu eixo uma câmara que vai adquirir as imagens perspectiva do seu interior. Na Tab. B.2 são descritos as variáveis de entrada e de saída da função *tube_imgs*.

A câmara navega alinhada com o eixo. Foi imposta uma rotação zero para que assim se simule, com as várias curvas do tubo, perspectivas onde apenas é visível parte da superfície local do tubo.

A câmara tem um campo de visão horizontal de 44 graus e uma resolução de 640*480 pixels (sensor 4/3).

Como já foi referido, é usado a função *tube_mk* para se construir uma estrutura tubular onde se possa obter imagens do interior. Em cada 'lStep' é criada uma câmara usando a função *camera_mk*($W, H, HFOV, cRw, ctw$) onde apenas *ctw* é alterado. Depois os pontos visíveis pela câmara são encontrados com a função

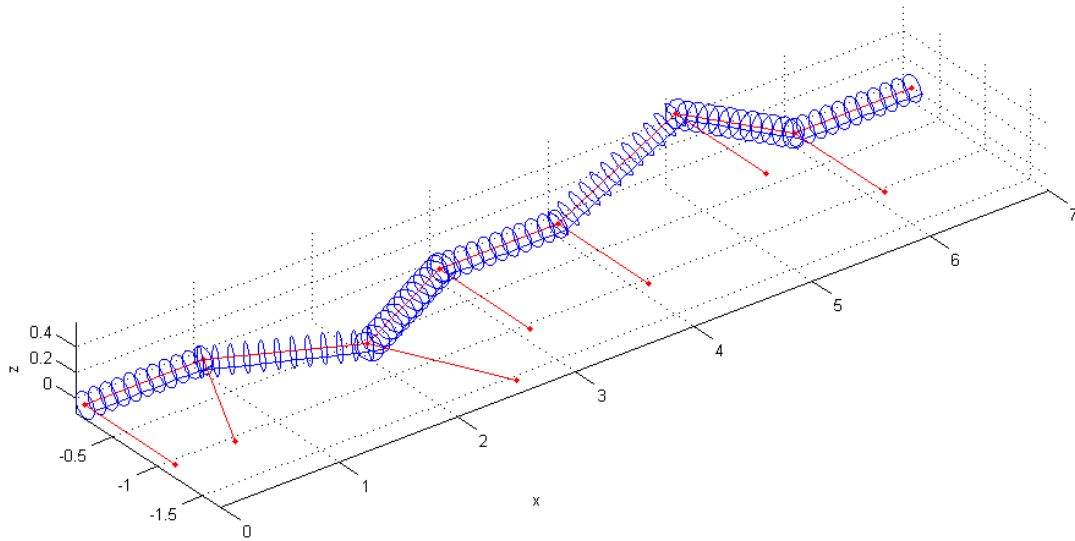


Figura B.3: Exemplo de um tubo com oito secções.

$visible_points(xyz', cRw, ctw, HFOV, VFOV, maxZFOV)$ e finalmente projectados para constituir cada imagem do interior do tubo $uv = P \times xyz$.

No caso de 'tube1imgs' a função $world2match(world, indList)$ realiza a correspondência entre os pontos das imagens e retorna as suas coordenadas.

Exemplo de utilização:

```
» tube_imgs (mesmo que tube_imgs('shTube1imgs'))
```

Resultado: Mostra as imagens dentro do tubo, Fig.B.4.

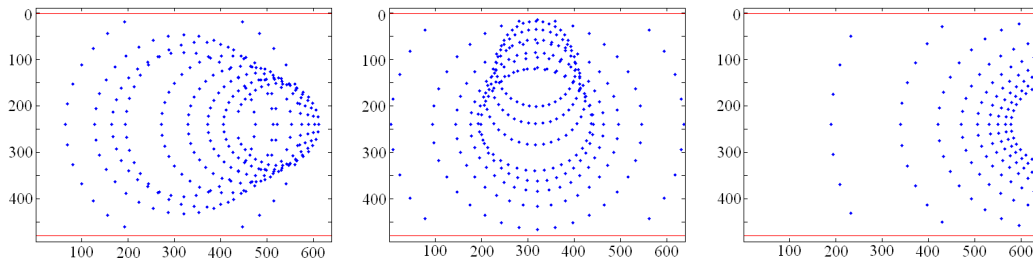


Figura B.4: Exemplo de três imagens do interior do tubo

B.3 Reconstrução Tridimensional

Com as imagens do interior da estrutura tubular e os respectivos pontos correspondentes é possível obter a reconstrução do cenário de forma a ser aplicado o modelo tubular. A função $tube_reconstr$ usa a função $tube_imgs$ para adquirir as imagens e os pontos correspondentes necessários à reconstrução.

Na Tab. B.3 são descritas as variáveis de entrada e de saída da função $tube_reconstr$.

<i>function [imatch, camK] = tube_imgs(opStr, options)</i>	
Input	
String : opStr 'shTubeImgs' 'tubeImgs' *	- Mostra as imagens da câmara adquiridas dentro do tubo - Retorna os pontos correspondentes entre imagens e a matriz intrínseca da câmara
Struct : options 'shFlagOriginal' 'shFlagImgs' 'tubeSec' 'IStep' 'maxZFOV' 'scale' 'HFOV' * 'ret_world'	- Mostra a estrutura tubular original. - Mostra as imagens adquiridas no interior da estrutura. - Define os pontos de curva do tubo e a direcção de cada troço. (ex: [.2 0 0; .2 0 1; .7 0 2; .2 0 3; .2 0 4; .2 -.5 5; .2 0 6; .2 0 7]). - Distância entre câmaras (ex: 0.015) - Distância máxima que a câmara vê (ex: 1) - Escala todas as dimensões das imagens, mas não altera a perspectiva porque normaliza todas as dimensões. (ex: 2) - Campo de visão horizontal da câmara, altera a amplitude da lente da câmara (ex: 44°) (VFOV=3/4*HFOV) - Retorna o tubo tridimensional
Output	
Struct: imatch Matrix(3 × 3): camK	- Retorna as coordenadas dos pontos correspondentes (<i>xy1, xy2, m12</i>) - Matriz intrínseca da câmara

Tabela B.2: Função tube_imgs: Variáveis de entrada e de saída.

A reconstrução tridimensional consiste em encontrar a localização da câmara em cada imagem para assim colocar no referencial do mundo os pontos reconstruídos. A função $xyz = Reconst(cam1, cam2, xy1, xy2)$ permite reconstruir um ponto 3D sabendo as matrizes das duas câmaras e um par de pontos correspondentes. A função $[info, R2, t2] = essential_matrix(m1', m2b', options)$ e usando, no mínimo, oito pares de pontos correspondentes, determina a transformação entre duas câmaras. A função $[xyzList, RtIncr, scale] = tube_and_motion(imatch, camK, options)$ calcula todas as transformações entre câmaras sucessivas e reconstrói localmente os pontos tridimensionais. A função $[world1, cams] = do_scale(xyzList, RtIncr, scale, options)$ aplica o escalonamento aos pontos reconstruídos obtendo a estrutura tridimensional. A função $[world2, indList] = clean_world2(world1, options)$ retira a sobreposição de pontos tridimensionais e retorna a estrutura tridimensional sem sobreposições mais uma lista de índices dos pontos que cada câmara vê.

Exemplo de utilização:

» *tube_reconstr* (mesmo que *tube_reconstruct('shStructure')*)

Resultado: Mostra o cenário reconstruído a partir das imagens, Fig.B.5. É de notar que pontos não visíveis nas imagens não são reconstruídos.

B.4 Ajuste ao Modelo Tubular

O modelo tubular ajusta a estrutura reconstruída a uma lista de secções tubulares. Analisa um conjunto de pontos reconstruídos e adapta esse conjunto a uma secção tubular. Depois são removidas eventuais zonas sobrepostas entre secções e finalmente é feito um reajustamento das secções para cobrir possíveis pontos reconstruídos não cobertos pelas secções tubulares.

Na Tab. B.4 são descritos as variáveis de entrada e de saída da função *tube_fit*.

<i>function [world, indList] = tube_reconstr(opStr, options)</i>	
Input	
String : opStr 'shStructure' 'tubeImgs' *	- Mostra o cenário reconstruído a partir das imagens. - Retorna o cenário reconstruído.
Struct : options 'lStep' 'maxZFOV' 'scale' 'HFOV' 'shFlagOriginal' 'shFlagImgs' 'shFlagReconts' 'shFlagAllReconts' 'Noise' 'tubeSec'	- Distância entre câmaras (ex: 0.015). - Distância máxima que a câmara vê (ex: 0.8). - Escala todas as dimensões das imagens, mas não altera a perspectiva porque normaliza todas as dimensões. - Campo de visão horizontal da câmara, altera a amplitude da lente da câmara (ex: 44) (VFOV=3/4*HFOV). - Mostra a estrutura tubular original. - Mostra as imagens adquiridas no interior da estrutura. - Mostra a estrutura tubular reconstruída a partir das imagens. - Se 1, mostra, localmente, a estrutura tubular reconstruída. - Se 2, mostra a estrutura tubular reconstruída incluindo as sobreposições. - Introduce o nível de ruído na correspondência entre pontos (ex: 0.0001). - Define os pontos de curva do tubo e a direcção de cada troço. (ex: [.2 0 0; .2 0 1; .7 0 2; .2 0 3; .2 0 4; .2 -.5 5; .2 0 6; .2 0 7]). [y -z x] - o x tem de ter espaçamento sempre igual a 1.
Output	
Se 'tubeImgs' world IndList	- Retorna a estrutura reconstruída e as câmaras (<i>xyz, camK, cams</i>) - Lista de índices dos pontos que cada câmara vê
Se 'shStructure'	- Mostra uma imagem da reconstrução

Tabela B.3: Função tube_reconstr: Variáveis de entrada e de saída.

Esta função usa a função *tube_reconstr* para obter a estrutura reconstruída do tubo. Com a função *cylinder_fit(mi)* é adaptado um troço dessa reconstrução a uma secção cilíndrica (usa a *fminsearch* - para encontrar a superfície que melhor adapta a nuvem de pontos). Depois remove as sobreposições entre superfícies com a função *cut_overlap(tube, xyz, indList)* e com a função *reassign_xyz2sections(tube2, xyz, indList)* actualiza os pontos reconstruídos incluídos pelas secções tubulares. A função *tube_fit_main(tube, xyz, indList)* aplica uma função de custo às novas secções tubulares de forma a encontrar uma superfície contínua. As secções tubulares são redefinidas localmente e como os eixos de cada secção podem não coincidir, então é encontrado um círculo intermédio que faça a ponte entre as secções criando assim mais uma secção. Desta forma as secções tubulares passaram a ser secções cónicas. O reajustamento final a cada secção é realizando usando *fminsearch('tube_fit_aux', x0, options)* de modo ao modelo tubular representar o melhor possível a nuvem de pontos do tubo reconstruído.

Exemplo de utilização:

» *tube_fit*

Resultado: Mostra o tubo com as secções cilíndricas do modelo, mostra a optimização do eixo e das secções cónicas finais.

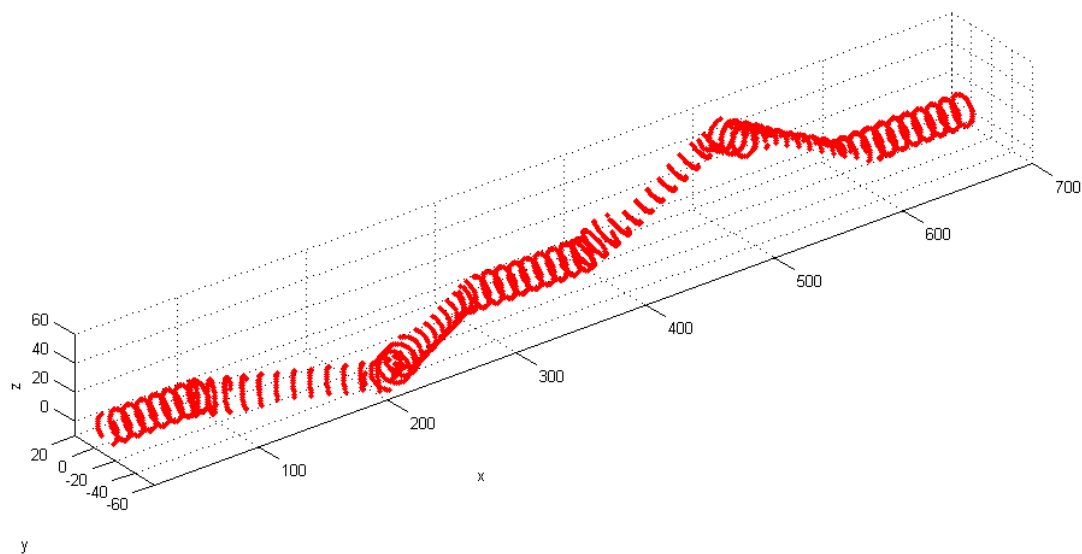


Figura B.5: Exemplo de reconstrução de uma estrutura tubular.

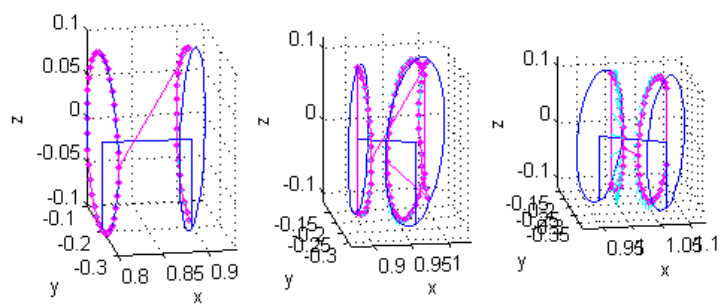


Figura B.6: Optimização das secções cónicas.

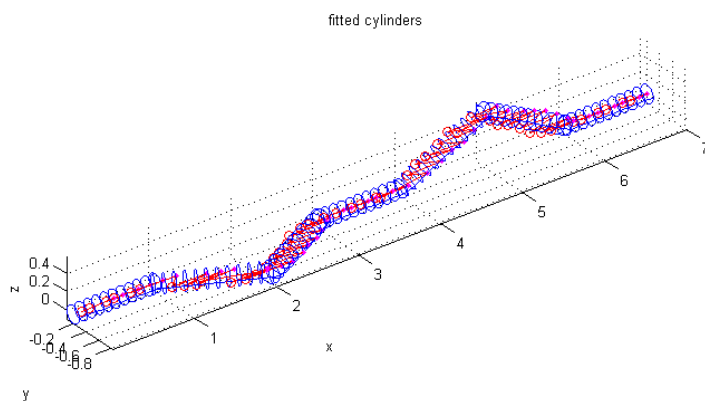


Figura B.7: Aplicação de secções cilíndricas aos pontos reconstruídos.

<i>function tube_fit(opStr, options)</i>	
Input	
String : opStr 'tubeImgs' *	- Para usar as imagens do tubo.
Struct : options * 'shFlag' 'lStep' 'maxZFOV' 'scale' 'HFOV' 'shFlagOriginal' 'shFlagImgs' 'shFlagReconts' 'shFlagAllReconts' 'Noise' 'tubeSec'	<ul style="list-style-type: none"> - Activa (ou desactiva se 0) a hipótese de mostrar o tubo. - Distância entre câmaras (ex: 0.015). - Distância máxima que a câmara vê (ex: 0.8). - Escala todas as dimensões das imagens, mas não altera a perspectiva porque normaliza todas as dimensões. - Campo de visão horizontal da câmara, altera a amplitude da lente da câmara (ex: 44) (VFOV=3/4*HFOV). - Mostra a estrutura tubular original. - Mostra as imagens adquiridas no interior da estrutura. - Mostra a estrutura tubular reconstruída a partir das imagens. - Se 1, Mostra, localmente, a estrutura tubular reconstruída. - Se 2, Mostra a estrutura tubular reconstruída incluindo as sobreposições. - Introduce o nível de ruído na correspondência entre pontos (ex: 0.0001). - Define os pontos de curva do tubo e a direcção de cada troço. (ex: [.2 0 0; .2 0 1; .7 0 2; .2 0 3; .2 0 4; .2 -.5 5; .2 0 6; .2 0 7]). [y -z x] - o x tem de ter espaçamento sempre igual a 1.
Output	
	- Mostra imagens do tubo reconstruído, de aplicação do modelo e do resultado final

Tabela B.4: Função tube_fit: Variáveis de entrada e de saída.

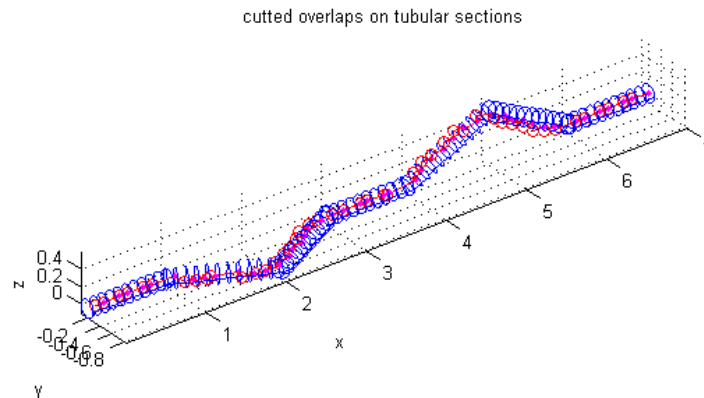


Figura B.8: Secções cilíndricas sem zonas sobrepostas

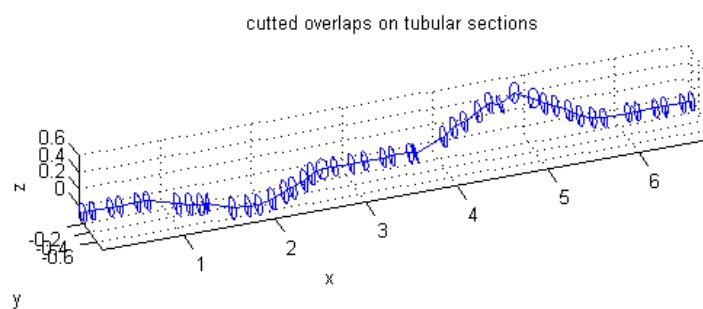


Figura B.9: Modelo representativo da estrutura tubular.

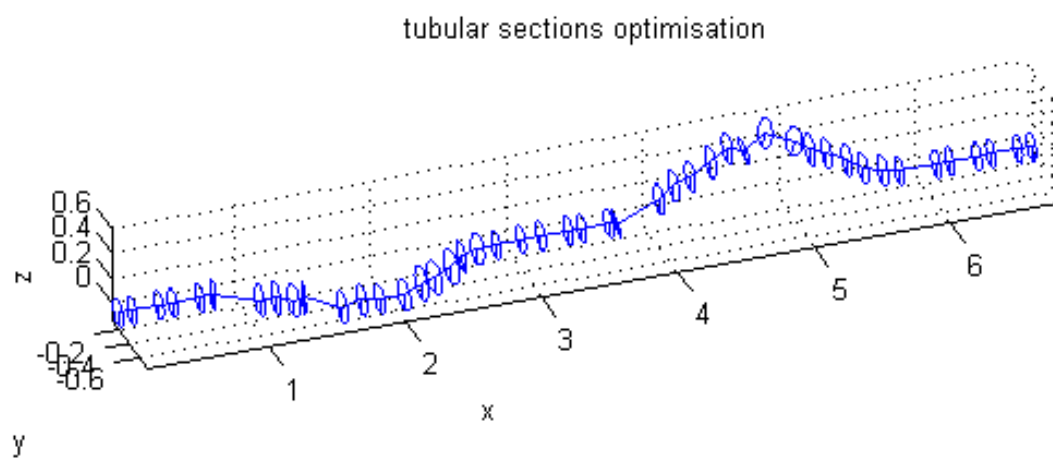


Figura B.10: Optimizaç o do modelo e das superf cies c nicas.