

3 Estereoscopia

Estereoscopia é uma técnica usada na visão computacional para obter informações do espaço tridimensional, através da análise de duas imagens capturadas simultaneamente em pontos diferentes (Trucco e Vierri, 2003).

De maneira geral, um sistema de visão estereoscópica precisa tratar de três problemas principais: calibração, correspondência e reconstrução.

Na calibração, procura-se determinar os parâmetros que descrevem o sistema de aquisição utilizado. O problema da correspondência consiste em determinar qual elemento na imagem capturada sob um ponto de vista corresponde a um dado elemento na imagem capturada sob outro ponto de vista. Por sua vez, através da reconstrução, procura-se recuperar a informação de profundidade com base nos parâmetros obtidos na etapa de calibração e nos pares de pontos correspondentes obtidos na etapa de correspondência.

O presente capítulo apresenta modelos, técnicas e soluções relacionadas aos problemas da estereoscopia. Inicialmente, um modelo matemático bem detalhado de uma câmera é descrito. Esse modelo relaciona um ponto no plano da câmera com uma reta no espaço 3D. E no final do capítulo, mostra-se que a estereoscopia é baseada na localização de um ponto 3D no espaço através das projeções desse ponto em duas imagens e na intersecção das retas associadas a essas duas projeções.

3.1. Modelo da Câmera

A maior parte dos algoritmos desenvolvidos para visão computacional parte do princípio que o modelo da câmera segue o modelo *pinhole* ("buraco de agulha"), apresentado na Fig. 3.1.

O modelo *pinhole* é composto por um plano π (plano da imagem), e por um ponto C no espaço (centro de projeção da câmera). A distância entre o plano da imagem e o centro de projeção é medida por um parâmetro da câmera

denominado distância focal, geralmente representado por f . A reta que passa por C perpendicular ao plano π é denominada eixo óptico, e a interseção entre π e o eixo óptico é denominada ponto principal, c . O sistema de coordenadas 3-D, cuja origem é C , onde o plano π é ortogonal ao eixo Z , é conhecido como sistema de coordenadas da câmera.

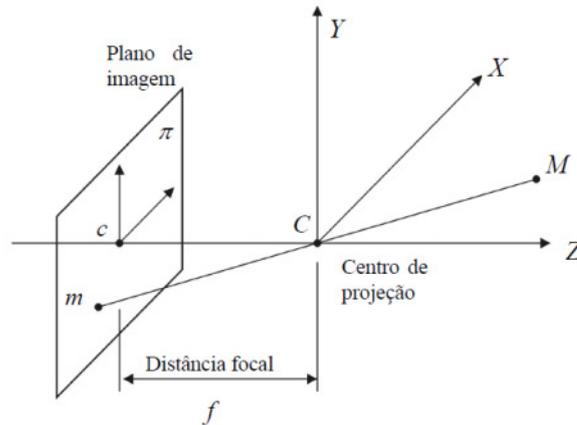


Figura 3.1 Geometria do modelo de câmera *pinhole*.

Na Fig. 3.1, por semelhança de triângulos, podemos calcular a relação entre um ponto no espaço $M = [X, Y, Z]^T$ e a sua projeção $m = [x, y]^T$ no plano da imagem,

$$m = \begin{pmatrix} f \frac{X}{Z} & f \frac{Y}{Z} \end{pmatrix} \quad (3.1)$$

Em diferentes aplicações, algoritmos de visão computacional utilizados na recuperação de informação da cena observada são necessários para relacionar as coordenadas dos pontos 3-D no mundo com as coordenadas de sua projeção no plano da imagem. As expressões definidas na Eq. 3.1 descrevem estas relações com base no sistema de coordenadas da câmera. No entanto, o sistema de coordenadas de interesse pode ser diferente do sistema da câmera. A origem do sistema pode estar definida por um marco do ambiente e apresentar qualquer orientação, conhecidos na literatura de visão computacional como sistema de coordenadas do mundo.

A relação entre as coordenadas de um ponto no mundo e as coordenadas de sua projeção na imagem envolve dois grupos de parâmetros, chamados parâmetros intrínsecos e extrínsecos.

3.1.1. Parâmetros Intrínsecos

Os parâmetros intrínsecos relacionam as coordenadas em *pixels* de um ponto no sistema de coordenadas da imagem com as suas correspondentes no sistema de coordenadas da câmera. Estes parâmetros caracterizam as influências ópticas, geométricas e digitais da câmera utilizada.

Para o modelo de câmera *pinhole*, podemos considerar três conjuntos de parâmetros intrínsecos principais.

O primeiro conjunto é formado pela distância focal, f , que representa a distância em *pixels* entre o centro de projeção e o plano da imagem.

O segundo conjunto de parâmetros é diretamente empregado na transformação de um ponto no sistema de coordenadas da imagem em *pixels* (x_{im} , y_{im}) para o sistema de coordenadas da câmera (x^c , y^c) de acordo com as seguintes relações (Trucco e Verri, 1998):

$$x^c = -(x^{im} - o_x) s_x \quad (3.2)$$

$$y^c = -(y^{im} - o_y) s_y \quad (3.3)$$

onde (o_x , o_y) são as coordenadas em *pixels* do centro da imagem e (s_x , s_y) o tamanho efetivo do *pixel* em milímetros nas direções horizontal e vertical.

O terceiro conjunto de parâmetros intrínsecos está relacionado com as distorções geométricas introduzidas pelo conjunto óptico utilizado. Na maioria dos casos, tais influências ópticas podem ser modeladas como simples distorções radiais utilizando-se as seguintes relações (Hartley e Zisserman, 2003):

$$x^c = -x_d (1 + k_1 r_d^2 + k_2 r_d^4) \quad (3.4)$$

$$y^c = -y_d (1 + k_1 r_d^2 + k_2 r_d^4) \quad (3.5)$$

com

$$r_d^2 = x_d^2 + y_d^2 \quad (3.6)$$

onde (x_d , y_d) são as coordenadas dos pontos distorcidos e k_1 e k_2 correspondem aos coeficientes de distorção.

Pode-se observar nas Eqs. (3.4) e (3.5) que a distorção é sempre nula no centro da imagem, e aumenta conforme o distanciamento do centro. Os coeficientes de distorção são, geralmente, muito pequenos. Sendo assim, a distorção radial pode ser desprezada quando alta precisão não é necessária em todas as regiões da imagem, ou quando os pontos periféricos são descartados. Todavia, como k_2 é bem menor que k_1 , é possível considerar apenas o valor de k_1 para modelar a distorção radial, mantendo ainda uma boa precisão (Trucco e Verri, 2003).

E, por último, o fator *skew* que determina o grau de perpendicularidade entre as direções dos eixos do *pixel* do sensor da câmera, sendo nulo nos casos em que as direções são perpendiculares.

Assim, define-se uma matriz K dos parâmetros intrínsecos da câmera:

$$K = \begin{bmatrix} -f_x & S & c_x \\ 0 & -f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (3.7)$$

onde tem-se que:

- f_x e f_y são proporcionais às distâncias focais definidas em *pixels* por unidade de distância em cada direção;
- c_x e c_y são as coordenadas correspondentes ao ponto principal da câmera e,
- S é o coeficiente *skew* da câmera.

Um modelo simplificado da matriz K considera $S = 0$, o que não é sempre verdade nas câmeras CCD utilizadas comumente.

3.1.2. Parâmetros Extrínsecos

Os parâmetros extrínsecos são definidos como um conjunto de parâmetros geométricos que relacionam, de modo único, o sistema de coordenadas da câmera com o sistema de coordenadas do mundo (Trucco e Verri, 2003). Esta transformação é especificada através de um vetor de translação T e uma matriz de rotação R , como mostrado na Fig. 3.2. O vetor de translação descreve a posição relativa entre as origens dos dois sistemas de coordenadas. A matriz de rotação

descreve a orientação relativa entre os eixos dos dois sistemas de coordenadas. Comumente, a rotação encontra-se representada na forma de vetor, v_{rot} (1×3), que define o eixo de rotação e seu módulo representa o ângulo de rotação, θ :

$$\theta = \|v_{rot}\| \quad (3.8)$$

A fórmula de Rodrigues permite calcular a matriz de rotação, R (3×3), como mostrado na Eq. 3.9:

$$R = \begin{bmatrix} \cos \theta + \omega_x^2 (1 - \cos \theta) & \omega_x \omega_z (1 - \cos \theta) - \omega_z \sin \theta & \omega_y \sin \theta + \omega_x \omega_z (1 - \cos \theta) \\ \omega_z \sin \theta + \omega_x \omega_y (1 - \cos \theta) & \cos \theta + \omega_y^2 (1 - \cos \theta) & -\omega_x \sin \theta + \omega_y \omega_z (1 - \cos \theta) \\ -\omega_y \sin \theta + \omega_x \omega_z (1 - \cos \theta) & \omega_x \sin \theta + \omega_y \omega_z (1 - \cos \theta) & \cos \theta + \omega_z^2 (1 - \cos \theta) \end{bmatrix} \quad (3.9)$$

onde ω_x , ω_y e ω_z são as componentes do vetor de rotação.

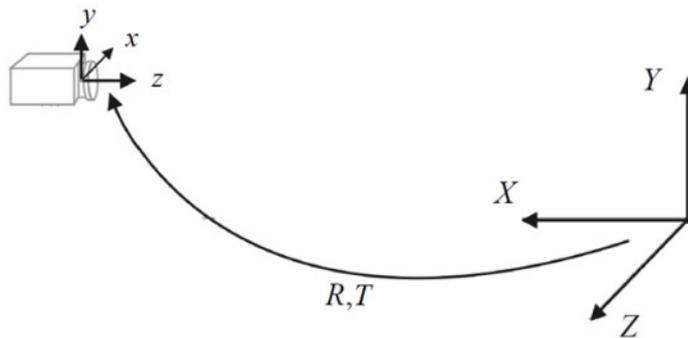


Figura 3.2 Transformação das coordenadas do mundo para as coordenadas da câmera.

Assim, a relação entre as coordenadas de um ponto no sistema de referência da câmera, M_c , e as coordenadas deste mesmo ponto no sistema de referência global, M_w , é dada por:

$$M_c = R(M_w - T) \quad (3.10)$$

onde:

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \quad (3.11)$$

$$T = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix} \quad (3.12)$$

3.2. Sistema de Visão Estereoscópica

Como visto no modelo *pinhole*, descrito na seção anterior, durante a formação da imagem, a câmera projeta um ponto no espaço em um plano bidimensional imaginário chamado de plano de imagem. No entanto, este processo não é reversível, visto que um ponto no plano de imagem pode corresponder à projeção de qualquer ponto sobre uma determinada reta no espaço. Em consequência, todas as informações tridimensionais são perdidas. Por este motivo, os sistemas de visão estereoscópica fazem uso de uma segunda câmera, introduzindo uma restrição adicional necessária para que seja possível realizar o processo de reconstrução, utilizando-se para isto técnicas de triangulação com as quais a informação 3D pode ser completamente recuperada.

O modelo simplificado (Trucco e Verri, 2003) da Fig. 3.3 apresenta a geometria básica presente em um sistema de visão estereoscópica composto por duas câmeras do modelo *pinhole*.

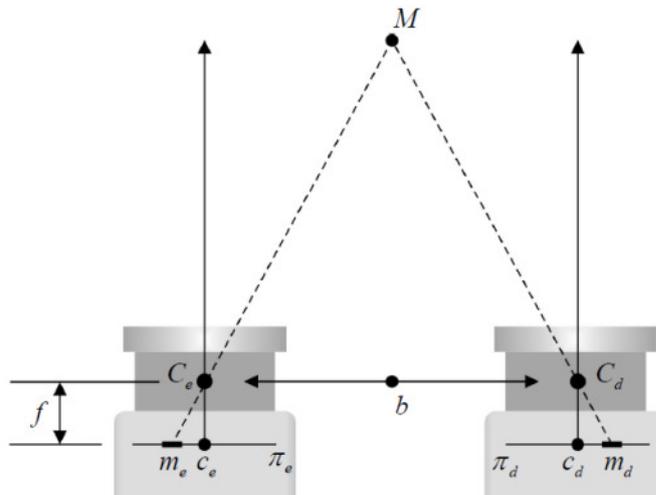


Figura 3.3 Modelo simplificado de um sistema de visão estéreo.

Os planos de imagem das câmeras da esquerda π_e e da direita π_d são coplanares, e os centros de projeção estão representados por C_e e C_d , respectivamente. Os eixos ópticos são paralelos, o que significa que a interseção dos mesmos encontra-se infinitamente distante das câmeras.

A posição de M no espaço é determinada por triangulação, ou seja, pela interseção dos raios definidos pelos centros de projeção e as imagens de M , m_e , e m_d . A distância, b , entre os centros de projeção C_e e C_d , é denominada linha base.

No entanto, a configuração do sistema de visão estereoscópica assumida neste modelo simplificado dificilmente é alcançada na prática, uma vez que é muito difícil construir um sistema binocular onde as câmeras estejam exatamente alinhadas conforme apresentado. Por este motivo, a calibração do sistema estéreo fornece parâmetros necessários, orientação e posicionamento das câmeras, de maneira a compensar as diferenças obtidas.

3.2.1. Calibração

O processo de calibração de câmeras é uma etapa importante na metodologia. Como resultado deste processo, obtém-se os valores dos parâmetros das câmeras, que depois serão utilizados para modelar o seu funcionamento.

Os parâmetros de um sistema de visão estereoscópica obtidos no processo de calibração estéreo são apresentados na Fig. 3.4.

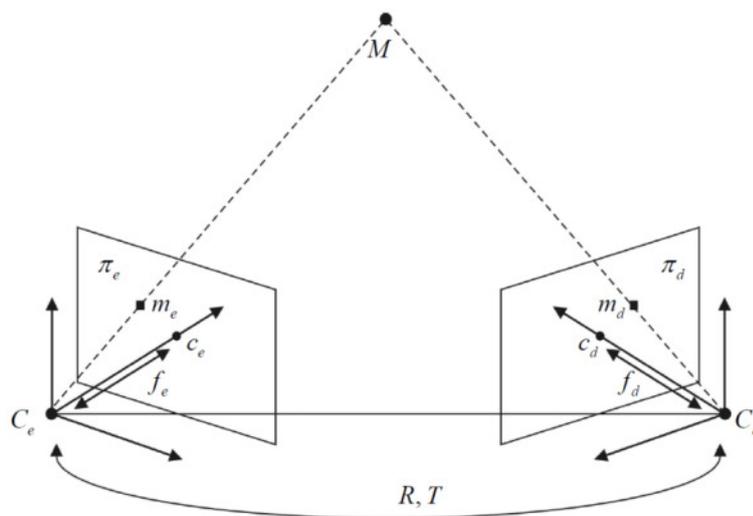


Figura 3.4 Parâmetros extrínsecos e intrínsecos do sistema de visão estéreo.

Neste caso, os parâmetros intrínsecos correspondem aos mesmos descritos na Seção 3.1.1 para cada uma das câmeras que compõe o sistema. Por outro lado, os parâmetros extrínsecos diferem ligeiramente dos parâmetros presentes no modelo de câmera *pinhole*. Em um sistema de visão estéreo, os parâmetros extrínsecos descrevem a transformação que relaciona os sistemas de coordenadas das duas câmeras um ao outro.

3.2.2. Correspondência

Correspondência em estereoscopia refere-se ao procedimento de relacionar elementos presentes nas imagens capturadas pelo sistema de visão estéreo, identificando quais pontos da imagem da esquerda e da direita são projeções do mesmo ponto na cena tridimensional observada. Este é um dos tópicos mais pesquisados em visão computacional, e é uma das principais questões a serem tratadas na estereoscopia.

Neste trabalho, o próprio algoritmo SIFT, descrito no Capítulo 2, é empregado para estabelecer as correspondências entre pontos característicos, como no exemplo mostrado na Fig. 3.5.

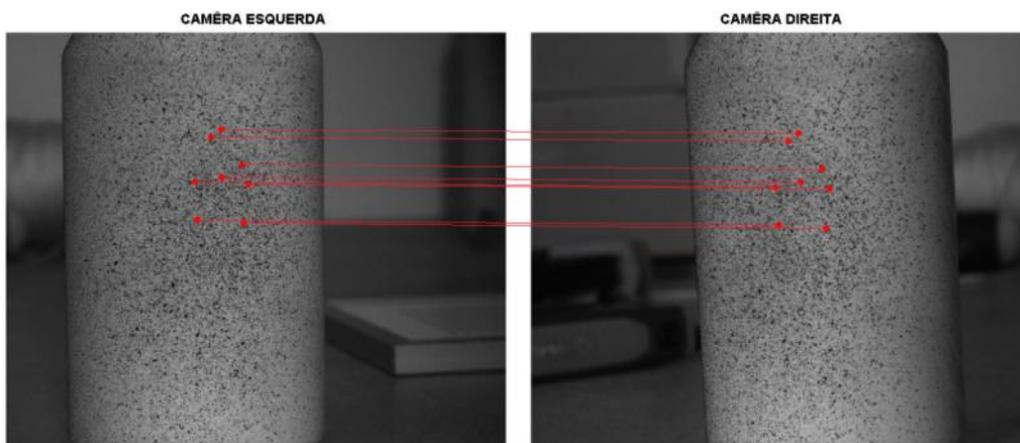


Figura 3.5 Exemplo de correspondência de pontos utilizando o algoritmo SIFT em um par de imagens capturadas simultaneamente de uma lata de refrigerante.

Desse modo, o SIFT pode ser utilizado tanto para correlacionar duas imagens de um objeto obtidas concomitantemente por um sistema binocular e extrair as informações 3-D, quanto para correlacionar imagens obtidas do mesmo objeto em estados distintos de deformação e identificar os deslocamentos.

3.2.3. Reconstrução

A reconstrução está relacionada à recuperação de informação de profundidade a partir de pontos correspondentes identificados nas duas imagens estéreo, e dos parâmetros que descrevem o sistema de visão estéreo. Segundo

Trucco e Verri (2003), para o caso em que os parâmetros intrínsecos e os extrínsecos são conhecidos, a reconstrução métrica pode ser calculada por triangulação.

Conforme exibido na Fig. 3.6, o ponto no espaço, M , projetado no par de pontos correspondentes m_e e m_d , encontra-se na intersecção dos raios P_e e P_d , construídos a partir dos pontos correspondentes projetados $C_e m_e - C_d m_d$. Em teoria, as duas linhas intersectam perfeitamente no espaço, mas isto não acontece na prática devido à presença de erros, como por exemplo: erros de digitalização, ruídos dos sensores, erros no processo de correspondência, etc. Desta forma, a intersecção é estimada como o ponto de mínima distância entre os dois raios (ponto médio).

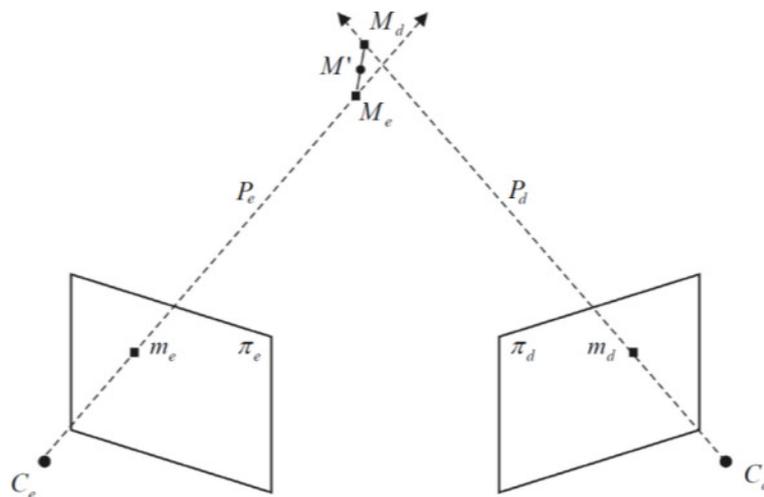


Figura 3.6 Reconstrução do ponto por triangulação com raios que não se intersectam no espaço.

O ponto M no espaço pode ser expresso em dois sistemas de referência $M_e = [X_e Y_e Z_e]$ e $M_d = [X_d Y_d Z_d]$, correspondentes aos sistemas de coordenadas das câmeras da esquerda e da direita, respectivamente. Estes pontos estão relacionados pela transformação homogênea entre seus sistemas de referência:

$$M_e = R M_d + T \quad (3.13)$$

As Eqs. (3.14) e (3.15) constituem os vetores de coordenadas resultantes das projeções em perspectiva usando as matrizes projetivas P_e e P_d .

$$m_e = \frac{M_e}{Z_e} \quad (3.14)$$

$$m_d = \frac{M_d}{Z_d} \quad (3.15)$$

Assim, a Eq. (3.13) é reescrita da forma:

$$m_e \cdot Z_e = R(m_d \cdot Z_d) + T \quad (3.16)$$

e na sua forma matricial, obtém-se:

$$\begin{bmatrix} -Rm_d & m_e \end{bmatrix} \begin{bmatrix} Z_d \\ Z_e \end{bmatrix} = T \quad (3.17)$$

Na Eq. (3.17), as coordenadas desconhecidas Z_d e Z_e podem ser determinadas pela combinação linear calculada por mínimos quadrados a partir da pseudo-inversa conforme apresentado na Eq. (3.18).

$$\begin{bmatrix} Z_d \\ Z_e \end{bmatrix} = \left(\begin{bmatrix} -Rm_d \\ m_e \end{bmatrix} \begin{bmatrix} -Rm_d & m_e \end{bmatrix} \right)^{-1} \begin{bmatrix} -Rm_d \\ m_e \end{bmatrix} T \quad (3.18)$$

E assim,

$$\begin{bmatrix} Z_d \\ Z_e \end{bmatrix} = \left(\begin{bmatrix} \|-Rm_d\|^2 & -Rm_d \cdot m_e \\ m_e \cdot (-Rm_d) & \|m_e\|^2 \end{bmatrix} \right)^{-1} \begin{bmatrix} -Rm_d \cdot T \\ t \cdot m_e \end{bmatrix} \quad (3.19)$$

A inversão da matriz quadrada pode ser feita facilmente,

$$\begin{bmatrix} Z_d \\ Z_e \end{bmatrix} = \frac{\begin{bmatrix} \|m_e\|^2 & -(-Rm_d \cdot m_e) \\ -(m_e \cdot (-Rm_d)) & \|-Rm_d\|^2 \end{bmatrix}}{\| -Rm_d \|^2 \| m_e \|^2 - (-Rm_d \cdot m_e)^2} \begin{bmatrix} -Rm_d \cdot T \\ t \cdot m_e \end{bmatrix} \quad (3.20)$$

E, as coordenadas desejadas Z_d e Z_e são determinadas pelas expressões:

$$\begin{bmatrix} Z_d \\ Z_e \end{bmatrix} = \begin{bmatrix} \frac{\|m_e\|^2 (-Rm_d \cdot T) - (-Rm_d \cdot m_e)(T \cdot m_e)}{\| -Rm_d \|^2 \| m_e \|^2 - (-Rm_d \cdot m_e)^2} \\ \frac{-(m_e \cdot (-Rm_d))(-Rm_d \cdot T) + \|-Rm_d\|^2 (T \cdot m_e)}{\| -Rm_d \|^2 \| m_e \|^2 - (-Rm_d \cdot m_e)^2} \end{bmatrix} \quad (3.21)$$

Assim, as posições M_d e M_e são obtidas a partir das Eqs. (3.14) e (3.15):

$$M_e = m_e \cdot Z_e \quad (3.22)$$

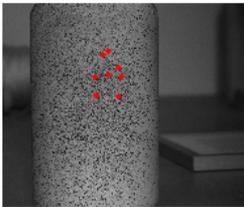
$$M_d = m_d \cdot Z_d \quad (3.23)$$

Finalmente, o ponto triangulado M' é calculado diretamente como o ponto médio da reta que passa por M_d e M_e , dado por:

$$M' = M_d + \frac{1}{2}(M_e - M_d) \quad (3.24)$$

Todos os algoritmos apresentados foram implementados na plataforma Matlab®. Um exemplo da aplicação do algoritmo de triangulação é mostrado na Fig. 3.7. Para isto, utilizam-se como dados de entrada as posições dos pontos SIFT localizados no par estéreo da lata de refrigerante da Fig. 3.5.

CAMÉRA ESQUERDA



CAMÉRA DIREITA

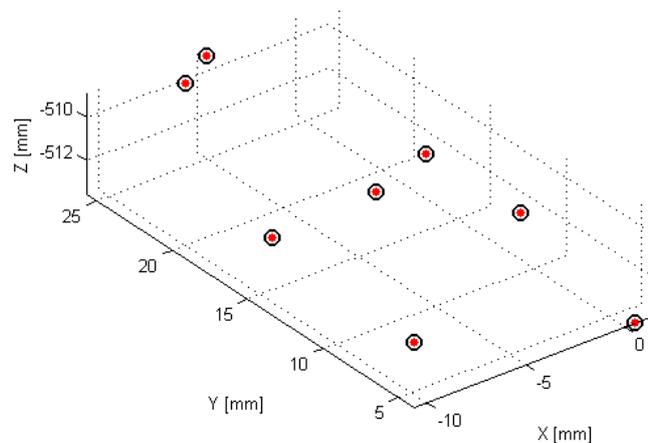
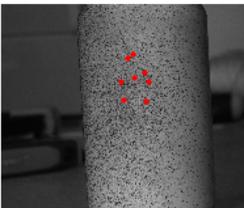


Figura 3.7 Exemplo de recuperação da posição no espaço para um conjunto de pontos localizados no par-estéreo.

No próximo capítulo, métodos sem malha são discutidos, com o objetivo de complementar o algoritmo SIFT e as técnicas de estereoscopia na identificação mais precisa de campos de deslocamento e deformação em espécimes submetidos a carregamentos mecânicos.