

# OBTENÇÃO DE INFORMAÇÃO 3D A PARTIR DE MOVIMENTO DE CÂMARA: CALIBRAÇÃO, DETECÇÃO E SIMPLIFICAÇÃO DE ENTIDADES, SEGUIMENTO TEMPORAL, TRIANGULAÇÃO

João Manuel R. S. Tavares

Laboratório de Óptica e Mecânica Experimental, INEGI, Departamento de Engenharia Mecânica, FEUP  
[tavares@fe.up.pt](mailto:tavares@fe.up.pt), [www.fe.up.pt/~tavares](http://www.fe.up.pt/~tavares)



## RESUMO

*A obtenção de informação tridimensional (3D) de objectos apresenta importância extrema em muitas aplicações de Visão Computacional. Como exemplos, podem ser citados: inspecção industrial, guiamento de veículos, reconstrução, seguimento e identificação de objectos. Neste artigo, é apresentada uma metodologia para obter informação 3D a partir do movimento de uma câmara, constituída pelas fases: calibração, detecção e simplificação de entidades, seguimento temporal das mesmas e obtenção de coordenadas 3D.*

## 1 – INTRODUÇÃO

A obtenção da estrutura 3D de objectos tem um papel importante em Visão Computacional. Por exemplo, em veículos e *robots* autónomos, reconstrução, seguimento e identificação de objectos, inspecção automática, etc. Actualmente existem várias técnicas para a obtenção de informação tridimensional (Silva, 94; Tavares, 95b; Coelho, 03; Azevedo 2008) que podem ser divididas em:

a) **Activas** – Envolvem a projecção controlada de energia, sendo a energia reflectida detectada por sensores que directa ou indirectamente fornecem informação 3D, ou movimento relativo.

b) **Passivas** – Não envolvem a projecção de energia nem o movimento relativo e usam a iluminação ambiente.

Como exemplos de técnicas activas têm-se:

- **Estereoscopia activa com luz estruturada:** Consiste em projectar um padrão de luz

e obter uma imagem de intensidades com a câmara afastada da fonte de luz. De seguida, estabelece-se correspondências entre os pontos do padrão projectado e os obtidos na imagem. Finalmente, calculam-se as coordenadas 3D usando as equações de projecção para a fonte de luz e para a câmara – procedimento designado por triangulação.

- **Câmara(s) activa(s):** Uma ou mais câmaras obtêm imagens dos objectos com movimento relativo. Assim, após detecção de pontos característicos nas várias imagens e emparelhamento dos mesmos, é possível obter a reconstrução 3D, assim como o movimento relativo e as características da(s) câmara(s).

- **Interferometria:** Consiste em projectar sobre os objectos riscas de sombra igualmente espaçadas. De seguida, deve-se adquirir uma imagem com a câmara deslocada lateralmente em relação ao projector e em frente da qual está colocada uma grelha idêntica à usada para projectar as riscas. A imagem obtida apresenta os contornos (franjas) dos pontos que estão à

mesma distância, mas não dá indicação da variação de distância entre linhas de contorno adjacentes. Informação sobre a referida variação pode ser obtida deslocando os objectos num sentido conhecido ou usando uma segunda grelha.

- **Deteção de ecos:** Consiste na projecção de energia e na medida do tempo que decorre entre a sua emissão e recepção.

Geralmente, as técnicas activas com projecção de energia são aptas para objectos sem pontos característicos. Contudo, apresentam algumas dificuldades: quando se usam padrões não codificados ou em superfícies com descontinuidades torna-se difícil estabelecer as correspondências; quando é utilizado um feixe ou um plano de luz existe a necessidade de “varrer” os objectos, provocando a lentidão do processo e a possível ocorrência de imprecisões; quando são utilizados vários planos de luz, o emparelhamento torna-se difícil; normalmente, surgem problemas nas partes ocultas de objectos.

Por outro lado, como técnicas passivas têm-se:

- **Duas (ou mais) câmaras estáticas (estereoscopia passiva):** Consiste em obter por cada câmara uma imagem. Estabelecer as correspondências entre pontos característicos das imagens obtidas. Determinar as coordenadas 3D por triangulação e com a geometria do *setup*.

- **Forma com base em X:** Procura-se extrair a informação 3D a partir de uma ou mais imagens monoculares obtidas com uma câmara fixa. Podem ser baseadas (X), por exemplo, no brilho, na perspectiva, na textura ou na oclusão.

As técnicas passivas apresentam como vantagem o facto de serem de aplicabilidade geral. Contudo, apresentam como desvantagens: dificuldade de aplicação em caso de objectos sem pontos característicos, complicações no caso de existirem pontos que apenas visíveis numa imagem, lentidão.

Recentemente, técnicas activas com movimento relativo têm merecido relevância, pois permitem a obtenção da informação 3D densa sem impor restrições complexas. Este

trabalho enquadra-se nesta classe, constituindo uma solução global para a obtenção de informação 3D a partir do movimento de uma câmara. A abordagem usada é constituída pelas fases: calibração da câmara, determinação e simplificação das entidades em cada imagem, seguimento das mesmas ao longo do tempo e obtenção das coordenadas 3D dos seus pontos iniciais e finais, Fig. 1.

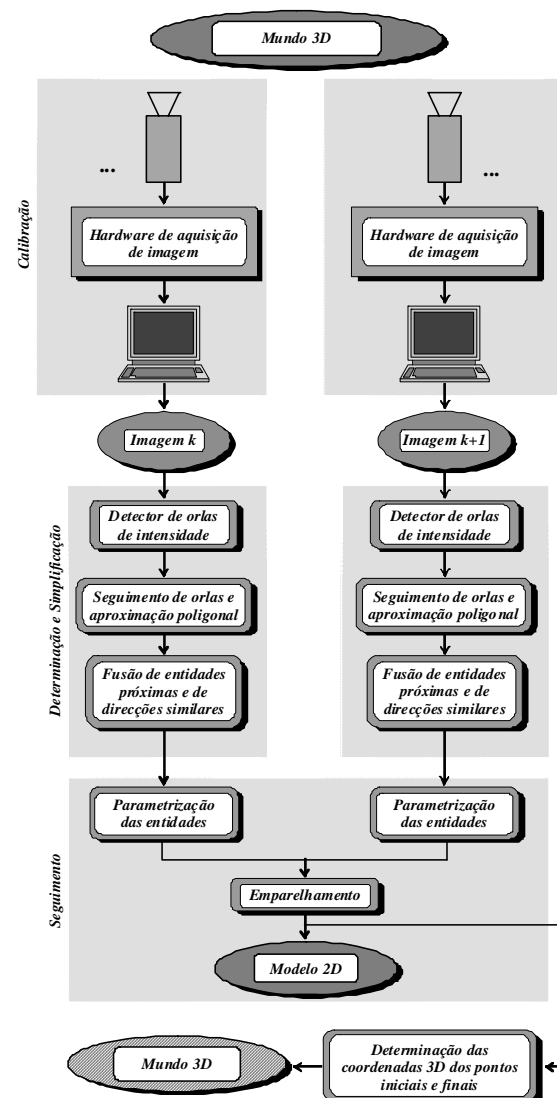


Fig. 1 – Diagrama da abordagem seguida.

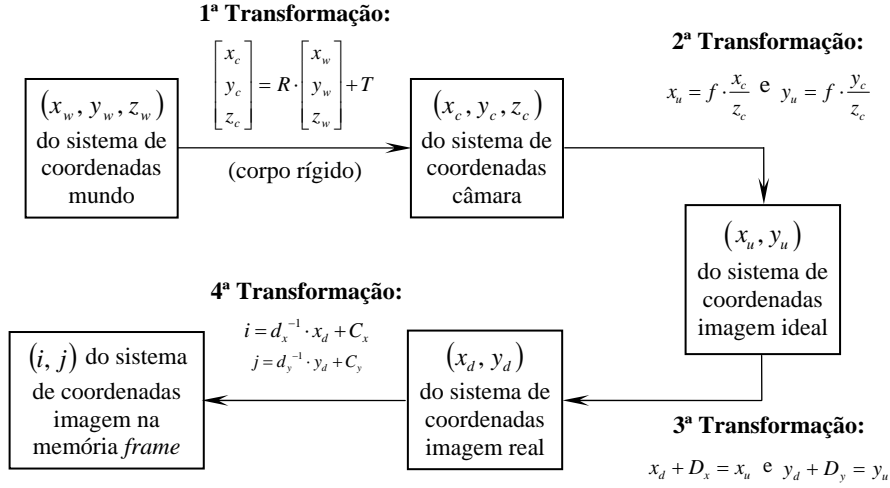
## 2 – METODOLOGIA

### 2.1 – Calibração

Quando se pretende obter informação 3D a partir de imagens, torna-se útil conhecer o modelo da(s) câmara(s) usada(s). Assim, é necessário o conhecimento de como pontos no espaço 3D são transformados em pontos no

plano imagem 2D da câmara para cada posição e orientação da mesma; isto é, é preciso realizar a calibração da câmara. Neste trabalho utilizou-se um método de calibração baseado no proposto em (Tsai, 87; Lenz, 88), que considera pontos coplanares (Tavares, 95a, b). A transformação considerada, do sistema de coordenadas mundo 3D para o sistema de coordenadas imagem distorcidas 2D (na memória *frame*),

está representada na Fig. 2. Nesta figura,  $R$  é a matriz de rotação,  $T$  o vector de translação,  $f$  a distância focal,  $(C_x, C_y)$  as coordenadas discretas do centro óptico,  $D_x = x_d \cdot k_1 \cdot r^2$  e  $D_y = y_d \cdot k_1 \cdot r^2$  (com  $r^2 = x_u^2 + y_u^2$ ) reflectem a influência da distorção radial da lente  $k_1$ , e  $d_x$  e  $d_y$  são as distâncias entre centros dos elementos sensores segundo  $x$  e  $y$ .



**Fig. 2** – Transformação do sistema de coordenadas mundo para o sistema de coordenadas imagem na memória *frame* utilizada na calibração.

O algoritmo de calibração considerado pode ser descrito da seguinte forma:

1) Para cada ponto de calibração  $k$ :

- Detectar a localização  $(i_k, j_k)$  na memória *frame*;
- Fazer  $(C_x, C_y)$  igual às coordenadas do centro da memória *frame*;
- Calcular  $(x_{d_k}, y_{d_k})$  a partir de:

$$i_k = d_x^{-1} \cdot x_{d_k} + C_x \text{ e } j_k = d_y^{-1} \cdot y_{d_k} + C_y.$$

2) Calcular  $t_y^{-1} \cdot r_{11}$ ,  $t_y^{-1} \cdot r_{12}$ ,  $t_y^{-1} \cdot t_x$ ,  $t_y^{-1} \cdot r_{21}$ ,  $t_y^{-1} \cdot r_{22}$ , resolvendo para cada ponto  $k$ :

$$\begin{bmatrix} y_{d_k} x_{w_k} \\ y_{d_k} y_{w_k} \\ y_{d_k} \\ -x_{d_k} x_{w_k} \\ -x_{d_k} y_{w_k} \end{bmatrix}^T \cdot \begin{bmatrix} t_y^{-1} \cdot r_{11} \\ t_y^{-1} \cdot r_{12} \\ t_y^{-1} \cdot t_x \\ t_y^{-1} \cdot r_{21} \\ t_y^{-1} \cdot r_{22} \end{bmatrix} = x_{d_k}.$$

3) Calcular a matriz  $R$  e as duas primeiras componentes do vector  $T$ :

- Calcular  $|t_y|$ :

Considerar  $R'$  uma submatriz de  $R$ :

$$R' = \begin{bmatrix} r'_{11} & r'_{12} \\ r'_{21} & r'_{22} \end{bmatrix} = \begin{bmatrix} r_{11}/t_y & r_{12}/t_y \\ r_{21}/t_y & r_{22}/t_y \end{bmatrix}.$$

Se o determinante de  $R'$  for diferente de zero, então  $t_y^2$  é igual a:

$$S_r - \frac{\sqrt{[S_r^2 - 4 \cdot (r'_{11} \cdot r'_{22} - r'_{21} \cdot r'_{12})^2]}}{2 \cdot (r'_{11} \cdot r'_{22} - r'_{21} \cdot r'_{12})},$$

onde  $S_r = r'_{11}^2 + r'_{12}^2 + r'_{21}^2 + r'_{22}^2$ . Se não:

$$t_y^2 = (r'_i{}^2 + r'_j{}^2)^{-1},$$

onde  $r'_i$  e  $r'_j$  são os elementos da linha ou da coluna da matriz  $R'$  sem nenhum elemento igual a zero.

- Determinar o sinal de  $t_y$  :

Escolher um ponto  $k$  cujas coordenadas imagem  $(i_k, j_k)$  sejam afastadas do centro óptico, considerar o sinal de  $t_y$  positivo e calcular:

$$\begin{aligned} r_{11} &= (t_y^{-1} \cdot r_{11}) \cdot t_y, & r_{12} &= (t_y^{-1} \cdot r_{12}) \cdot t_y, \\ r_{21} &= (t_y^{-1} \cdot r_{21}) \cdot t_y, & r_{22} &= (t_y^{-1} \cdot r_{22}) \cdot t_y, \\ t_x &= (t_y^{-1} \cdot t_x) \cdot t_y, \\ x_{c_k} &= r_{11} \cdot x_{w_k} + r_{12} \cdot y_{w_k} + t_x, \\ y_{c_k} &= r_{21} \cdot x_{w_k} + r_{22} \cdot y_{w_k} + t_y, \end{aligned}$$

se  $i_k$  e  $x_{c_k}$  têm o mesmo sinal, assim como  $j_k$  e  $y_{c_k}$ , então  $t_y$  é positivo, caso contrário,  $t_y$  é negativo.

- Calcular a matriz  $R$  e a componente  $t_x$  do vector  $T$  :

$$\begin{aligned} r_{11} &= (t_y^{-1} \cdot r_{11}) \cdot t_y, & r_{12} &= (t_y^{-1} \cdot r_{12}) \cdot t_y, \\ r_{21} &= (t_y^{-1} \cdot r_{21}) \cdot t_y, & r_{22} &= (t_y^{-1} \cdot r_{22}) \cdot t_y, \\ t_x &= (t_y^{-1} \cdot t_x) \cdot t_y, \end{aligned}$$

$$R = \begin{bmatrix} r_{11} & r_{12} & \sqrt{(1-r_{11}^2-r_{12}^2)} \\ r_{21} & r_{22} & s \cdot \sqrt{(1-r_{21}^2-r_{22}^2)} \\ r_{31} & r_{32} & r_{33} \end{bmatrix},$$

com  $s = -\text{sinal}(r_{11} \cdot r_{21} + r_{12} \cdot r_{22})$ . As componentes  $r_{31}$ ,  $r_{32}$  e  $r_{33}$  são obtidas a partir do produto vectorial das duas primeiras linhas.

- 4) Calcular uma aproximação para  $f$  e para  $t_z$  ignorando a distorção:

Para cada ponto  $k$ , estabelecer:

$$\begin{bmatrix} y_{c_k} & -y_{d_k} \end{bmatrix} \cdot \begin{bmatrix} f \\ t_z \end{bmatrix} = w_k \cdot y_{d_k},$$

onde:

$$y_{c_k} = r_{21} \cdot x_{w_k} + r_{22} \cdot y_{w_k} + t_y$$

e

$$w_k = r_{31} \cdot x_{w_k} + r_{32} \cdot y_{w_k}.$$

Resolver este sistema com um número elevado de pontos de calibração.

Se a aproximação para  $f$  é menor do que zero, então actualizar a matriz  $R$  :

$$R = \begin{bmatrix} r_{11} & r_{12} & -\sqrt{(1-r_{11}^2-r_{12}^2)} \\ r_{21} & r_{22} & -s \cdot \sqrt{(1-r_{21}^2-r_{22}^2)} \\ -r_{31} & -r_{32} & r_{33} \end{bmatrix}.$$

- 5) Calcular a solução exacta para  $f$ ,  $t_z$  e  $k_1$ , resolvendo:

$$\begin{aligned} y_d (1 + k_1 \cdot r^2) &= \\ f \cdot \frac{r_{21} \cdot x_w + r_{22} \cdot y_w + r_{23} \cdot z_w + t_y}{r_{31} \cdot x_w + r_{32} \cdot y_w + r_{33} \cdot z_w + t_z}, \end{aligned}$$

usando como solução inicial as aproximações de  $f$  e de  $t_z$  encontradas anteriormente e zero para  $k_1$ .

Como pontos de calibração são considerados vértices de quadrados definidos no plano de calibração. As suas coordenadas 2D no plano imagem são obtidas por um algoritmo que considera: 1) detecção de orlas de intensidade (ver, por exemplo, (Tavares, 95b)), 2) seguimento das orlas usando informação de intensidade e de direcção, 3) regressão linear das orlas seguidas, 4) determinação dos vértices por intercepção das rectas de regressão (Tavares, 95a, b). Na Fig. 3, está representada uma imagem de calibração e identificados por intermédio de cruces os vértices determinados.

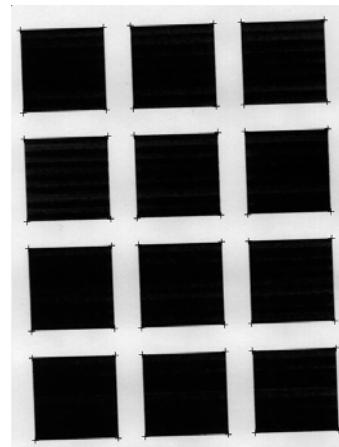


Fig. 3 – Exemplo de uma imagem de calibração e dos pontos de calibração obtidos.

## 2.2 - Detecção e simplificação de entidades

Em cada imagem da sequência desejada é necessário determinar as entidades a processar. Optou-se por considerar as entidades como sendo segmentos de recta, determinados inicialmente por um detector de orlas de intensidade, após o qual é executado um seguimento com histerese usando informação de intensidade e de direcção. As várias orlas seguidas são posteriormente submetidas a uma aproximação poligonal, utilizando-se um algoritmo de faixas dinâmicas (Leung, 90; Tavares, 95b), Fig. 4.

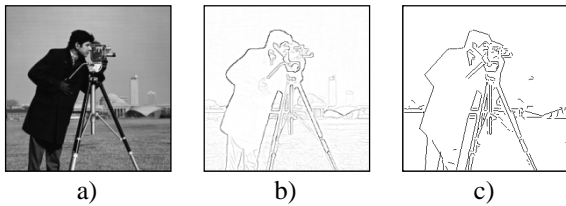


Fig. 4 – a) Imagem Original, b) Orlas determinadas, c) Aproximação poligonal obtida.

Como representado na Fig. 5, uma faixa é definida por uma linha crítica e por duas linhas de fronteira. A linha crítica é definida por dois pontos de referência, o primeiro e o segundo ponto da linha que se pretende aproximar, e as duas linhas fronteira são paralelas à linha crítica e a uma dada distância desta (tolerância de aproximação). As duas linhas de fronteira formam uma faixa para restrição do processo de aproximação. A linha que se pretende aproximar é então percorrida ponto a ponto. O processo pára no primeiro ponto que sai da faixa. A ideia básica do método de aproximação consiste em rodar a faixa de aproximação, utilizando o primeiro ponto como *pivot*, de forma a incluir o maior número possível de pontos da linha que se pretende aproximar. Na impossibilidade de rodar a faixa, sem deixar de incluir um ponto previamente considerado, é definido um novo segmento de recta e o processo de aproximação continua considerando uma nova faixa.

Como segmentos de recta que sejam próximos e de direcções similares provavelmente representam uma mesma entidade,

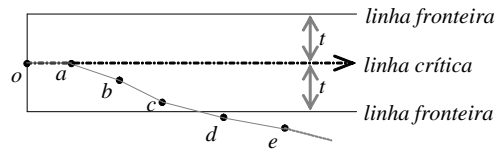


Fig. 5 – Definição de uma faixa de aproximação.

optou-se por fundir os segmentos nessas condições, Fig. 6.

O algoritmo considerado para simplificar segmentos de recta pode ser assim descrito (Tavares, 95a, b, d):

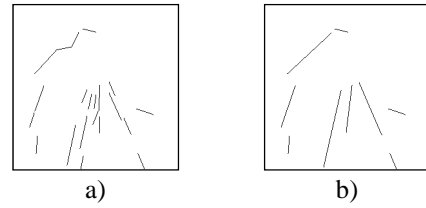


Fig. 6 – Simplificação de segmentos de rectas: a) Segmentos originais, b) Segmentos após fusão.

- 1) Definir as coordenadas do centróide  $(x_G, y_G)$  dos pontos extremos dos segmentos a fundir,  $(a, b)$  e  $(c, d)$ , considerando os comprimentos  $l_i$  e  $l_j$  como “massas” pontuais:

$$\begin{cases} x_G = \frac{l_i \cdot (a_x + b_x) + l_j \cdot (c_x + d_x)}{2 \cdot (l_i + l_j)} \\ y_G = \frac{l_i \cdot (a_y + b_y) + l_j \cdot (c_y + d_y)}{2 \cdot (l_i + l_j)} \end{cases}$$

- 2) Definir a direcção do segmento de fusão  $\theta_r$  como a média pesada das direcções dos segmentos a fundir,  $\theta_i$  e  $\theta_j$ , considerando  $l_i$  e  $l_j$  como pesos:

$$\text{Se } |\theta_i - \theta_j| \leq \frac{\pi}{2} \text{ então } \theta_r = \frac{l_i \cdot \theta_i + l_j \cdot \theta_j}{l_i + l_j},$$

$$\text{Se não } \theta_r = \frac{l_i \cdot \theta_i + l_j \cdot \left( \theta_j - \pi \cdot \frac{\theta_j}{|\theta_j|} \right)}{l_i + l_j}.$$

- 3) Definir um referencial  $(X_G, Y_G)$  centrado no centróide e com o eixo  $X_G$  paralelo à direcção  $\theta_r$ .

4) Os pontos extremos dos segmentos a fundir no referencial  $(X_G, Y_G)$ :

$$\begin{cases} \delta_{xG} = (\delta_y - y_G) \cdot \sin \theta_r + (\delta_x - x_G) \cdot \cos \theta_r \\ \delta_{yG} = (\delta_y - y_G) \cdot \cos \theta_r - (\delta_x - x_G) \cdot \sin \theta_r \end{cases}$$

5) As duas projecções mais afastadas, sobre o eixo  $X_G$  dos pontos  $a$ ,  $b$ ,  $c$  e  $d$  definem o segmento de fusão.

Deste modo, o conjunto de segmentos de recta original reduz-se a um conjunto menor de elementos. No entanto, as entidades resultantes serão mais representativas. Como vantagens desta diminuição podem ser citadas: um menor tempo de computação, entidades de reduzida importância são absorvidas por entidades de maior importância, ou simplesmente desprezadas, o seguimento posterior fica facilitado.

### 2.3 – Seguimento de entidades

Após a determinação das entidades estas devem ser parametrizadas (Deriche, 87, 90). A parametrização escolhida é baseada na posição do ponto médio, na direcção e no comprimento, pois permite a utilização eficiente [Deriche, 90; Gai-Checa, 93] de três filtros de Kalman independentes: um, para as coordenadas do ponto médio; outro, para a direcção; e um último, para o comprimento.

O modelo cinemático associado a cada filtro de Kalman é um modelo de aceleração constante localmente, sendo a aceleração considerada como um processo Gauss-Markov de primeira ordem e igual a uma fracção  $\alpha$  da anterior. Assim, a aceleração estimada tende para zero; isto significa, que se uma entidade desaparece, o filtro diminuíra a ultima aceleração conhecida até zero, através de potências de  $\alpha$ . A utilização de um modelo cinemático deste tipo torna a abordagem mais flexível e adaptável.

#### 2.3.1 – Filtro de kalman

O filtro de Kalman (Maybeck, 79; Tavares, 95b, c; Pinho, 2009) é neste trabalho utilizado para a estimativa temporal do comportamento das entidades. Este filtro é uma abordagem

estatística para estimar um vector de características que variam ao longo do tempo  $\hat{x}_t$  a partir de medidas que incluem ruído  $\hat{z}_t$ . Considerando a estimativa  $\hat{x}_{t+k}$  a partir da medição até ao instante  $k$ , a utilização deste filtro é um esquema recursivo para descrever o modelo dinâmico do sistema, a estatística do erro entre o modelo e a realidade, e a incerteza associada à medição. O filtro de Kalman pode ser interpretado como constituído pelas seguintes fases, Fig. 7:

- **Previsão**

A estimativa para o vector de características  $\hat{x}$  no instante  $t$  é:

$$\hat{x}_t^- = \Phi \cdot \hat{x}_{t-1}^+,$$

onde  $\Phi$  é a matriz cinemática e os índices  $\pm$  indicam estimativa posterior e anterior à fase de medição, respectivamente.

No instante inicial  $t=0$ , considera-se que as características desconhecidas são nulas.

A estimativa para a incerteza associada ao vector de características  $\hat{x}_t$  no instante  $t$  é:

$$P_t^- = \Phi \cdot P_{t-1}^+ \cdot \Phi^T + Q,$$

onde  $Q$  é uma matriz diagonal, cujos elementos são determinados em função da aplicação, traduzindo a inclusão da variância do ruído Gaussiano de média nula esperado, de modo a evitar que o filtro convirja em demasia.

No instante  $t=0$ , considera-se que a matriz de incerteza é diagonal com os respectivos elementos assumindo valores determinados pela experiência desenvolvida.

- **Medição e emparelhamento**

Após a estimação do vector de características e da sua incerteza, deve-se proceder à medição. De notar que não é necessário proceder à medição das derivadas das variáveis de parametrização, pois o filtro permite estimar recursivamente essas derivadas a partir da medição das variáveis associadas (Crowley, 92).

De seguida deve-se seleccionar, de entre as várias entidades candidatas, aquela que apresenta características que melhor se

aproximam às estimadas. Após este emparelhamento, o vector de características medidas  $\hat{z}_t$  assume os valores das características da entidade emparelhada e a matriz de incerteza associada a este vector  $R_t$  as variâncias medidas.

• **Actualização**

Após a medição e o emparelhamento, deve ser calculado o ganho do filtro  $K_t$ , que traduz o peso que se deve atribuir à medição e à estimativa:

$$K_t = P_t^- \cdot H^T \cdot [H \cdot P_t^- \cdot H^T + R_t]^{-1},$$

onde  $H$  é a matriz que transforma o vector de características estimado  $\hat{x}_t$  no vector de características medido  $\hat{z}_t$  e  $R_t$  é a matriz de variância medida.

Após o cálculo do ganho, pode-se realizar a actualização da estimativa utilizando a medição efectuada:

$$\hat{x}_t^+ = [I - K_t \cdot H] \cdot \hat{x}_t^- + K_t \cdot \hat{z}_t,$$

com  $I$  sendo a matriz identidade.

A correcção da matriz de incerteza associada é:

$$P_t^+ = [I - K_t \cdot H] \cdot P_t^- [I - K_t \cdot H]^T + K_t \cdot R_t \cdot K_t^T$$

Assim, na actualização quanto maior o ganho, maior o peso atribuído à medição e menor peso à estimativa. Adicionalmente, se a incerteza associada à medição for menor do que a incerteza associada à estimativa, o filtro atribui maior peso à medição e vice-versa.

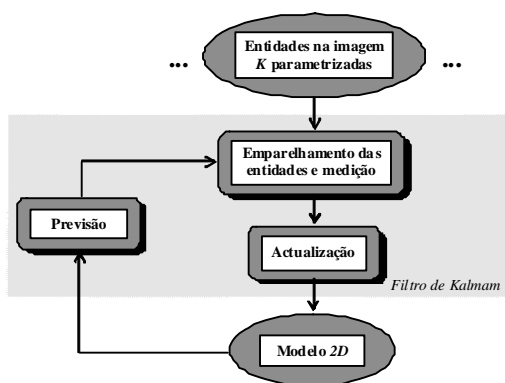


Fig. 7 – Diagrama de funcionamento do filtro de Kalman.

2.3.2 – Distância normalizada de Mahalanobis

No interior da área de medição resultante do filtro para a posição do ponto médio, é necessário associar uma medida do emparelhamento de cada possível entidade candidata com a entidade cujas características foram estimadas. Assim, na primeira tentativa de emparelhamento são utilizadas distâncias normalizadas de Mahalanobis como medida do emparelhamento. Esta distância é definida como a diferença de características normalizada pelas suas variâncias:

$$d_{\chi^2} = \frac{(X_m - X_e)^T \cdot (V_m - V_e)^{-1} \cdot (X_m - X_e)}{2},$$

onde  $X_m$  é o vector de características e  $V_m$  a matriz de variância da candidata ao emparelhamento,  $X_e$  é o vector de características e  $V_e$  a matriz de variância estimadas.

Para o caso de uma característica escalar, a equação anterior simplifica-se:

$$d_{\chi^2} = \frac{(x_m - x_e)^2}{2 \cdot (v_m - v_e)},$$

onde  $x_m$  é a característica e  $v_m$  a variância da candidata,  $x_e$  é a característica e  $v_e$  a variância estimadas.

Esta distância tem uma distribuição  $\chi^2$ , sendo o número de graus de liberdade igual ao de características consideradas.

2.3.3 – Restrições geométricas

Quando o emparelhamento não é conseguido com êxito pela utilização das distâncias de Mahalanobis, este é novamente tentado utilizando três restrições geométricas:

- i. A diferença entre o comprimento da candidata e o comprimento estimado deve ser  $\leq$  a um valor admissível.
- ii. A diferença entre a direcção da candidata e a direcção estimada deve ser  $\leq$  a um valor admissível.
- iii. A diferença entre a posição do ponto médio da candidata e a posição do ponto

médio estimada deve ser  $\leq$  a um valor admissível.

### 2.3.4 – Fase de medição e emparelhamento

A estimação da matriz de incerteza associada ao vector de características estimado para a posição do ponto médio das entidades determina uma área elíptica no interior da qual a entidade cujas características estimadas pelos três filtros deverá encontrar-se. Assim, considera-se a matriz  $M$  :

$$M = H \cdot P_i^- ,$$

a partir da qual é possível determinar a elipse de emparelhamento, centrada na posição estimada pelo filtro de posição do ponto médio, para a entidade em causa:

- Os valores próprios de  $M$  definem os raios maior e menor da elipse;
- Os seus vectores próprios de  $M$  definem os eixos maior e menor da elipse.

Definida a elipse, é necessário realizar o emparelhamento da entidade estimada com a melhor entidade medida. Inicialmente, tal emparelhamento é tentado como:

#### **Início**

{

**Considerar** como candidatas todas aquelas cujo ponto médio é interior à elipse de emparelhamento e ainda não foram emparelhadas.

**Se** não existe nenhuma candidata, **então fim.**

**Fazer** para todas as candidatas:

{

**Calcular** a distância de Mahalanobis entre a direcção da candidata e a prevista. **Se** esta distância é maior que um dado limiar, **então continuar.**

**Calcular** a distância de Mahalanobis entre o comprimento da candidata e o previsto. **Se** esta distância é maior que um dado limiar, **então continuar.**

**Calcular** a distância de Mahalanobis entre a posição do ponto médio da candidata e a prevista. **Se** esta distância é maior que um dado limiar, **então continuar.**

**Calcular** o produto das três distâncias calculadas anteriormente. **Se** este

*produto é menor que o mínimo encontrado até esta altura, então a melhor entidade é esta.*

}  
}

Se o emparelhamento é conseguido com êxito, então o estado medido é o seguinte:

#### • **Filtro para a direcção**

O elemento do vector de características medido assume o valor correspondente à direcção da entidade determinada pelo emparelhamento. Enquanto, o elemento relativo à variância da matriz de incerteza associada à medição, assume o valor da variância calculada na área de emparelhamento.

#### • **Filtro para o comprimento**

O primeiro elemento do vector de características medido assume o comprimento da entidade determinada pelo emparelhamento. Enquanto, o elemento relativo à variância da matriz de incerteza associada à medição assume o valor da variância calculada na área de emparelhamento.

#### • **Filtro para a posição do ponto médio**

Os elementos do vector de características medido relativos assumem as coordenadas do ponto médio da entidade determinada pelo emparelhamento. Enquanto, para a matriz de incerteza associada à medição, os elementos relativos às variâncias assumem as variâncias calculadas na área de emparelhamento.

Quando o emparelhamento não é conseguido, este é novamente tentado utilizando-se restrições geométricas:

#### **Início**

{

**Fazer** para todas as entidades não emparelhadas:

{

**Calcular** a diferença entre a direcção da candidata e a prevista.

**Se** a diferença for maior que o respectivo limiar, **então continuar.**

**Calcular** a diferença entre o comprimento da candidata e o previsto.



*Se a diferença for maior que o respectivo limiar, então continuar.*

*Calcular a diferença entre a posição do ponto médio da candidata e a prevista.*

*Se a diferença for maior que o respectivo limiar, então continuar.*

*Se o produto das três diferenças for menor que o mínimo encontrado até esta altura, então a melhor entidade é esta.*

}  
}

Quando o emparelhamento é desta feita conseguido com êxito, o estado medido é calculado como descrito anteriormente, sendo a área de emparelhamento centrada no ponto médio da entidade emparelhada pela utilização das restrições geométricas.

Se após esta tentativa o emparelhamento não é conseguido com êxito, o estado medido é assumido como:

- **Filtro para a direcção**

O elemento do vector de características medido assume a direcção estimada. Enquanto, o elemento da matriz de incerteza associada à medição assume o valor atribuído na fase de inicialização.

- **Filtro para o comprimento**

O elemento do vector de características medido assume o comprimento estimado. Enquanto, o elemento da matriz de incerteza associada à medição assume o valor atribuído na fase de inicialização.

- **Filtro para a posição do ponto médio**

Os elementos do vector de características medido assumem os valores estimados. Enquanto, para a matriz de incerteza associada à medição os seus elementos assumem os valores atribuídos na fase de inicialização.

Assim, quando o emparelhamento de uma dada entidade não é conseguido com êxito, aumenta-se a incerteza associada à medição de tal maneira que os filtros aumentam o peso atribuído à previsão. Desta forma, na posterior previsão a incerteza virá aumentada, provocando um consequente aumento da área de emparelhamento.

### 2.3.5 – Modelo de entidades

No seguimento das entidades é utilizado um modelo 2D de entidades (Tavares, 95a, b, 95c). Assim, quando uma nova entidade se torna visível, ela é inserida no modelo com um factor de confiança de valor médio. Sempre que uma entidade é novamente visível, este factor é aumentado de uma unidade até a um valor máximo. Se, pelo contrário, não é visível, este factor é diminuído em uma unidade até a um valor mínimo, caso em que é removida do modelo.

Com a gestão adoptada, dispõe-se de um modelo cujos elementos estão continuamente actualizados, não existindo a possibilidade de os recursos utilizados crescerem de forma não controlada.

## 2.4 – Obtenção de informação 3D

Após o emparelhamento das entidades consideradas em imagens sucessivas, torna-se possível extrair as coordenadas 3D dos seus pontos.

Um ponto no sistema 3D mundo é transformado num ponto 2D no plano imagem por uma transformação global determinada através da calibração da câmara. Esta transformação para um ponto  $P$ , em coordenadas homogéneas  $(x_w, y_w, z_w, 1)$ , para uma câmara na posição e orientação  $k$ , é traduzida como:

$$\begin{bmatrix} w \cdot x_u^k \\ w \cdot y_u^k \\ w \end{bmatrix} = M_{pp_k} \cdot \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix},$$

com

$$M_{pp_k} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix},$$

onde  $(w \cdot x_u^k, w \cdot y_u^k, w)$  são as coordenadas homogéneas não distorcidas do ponto  $P$  no plano imagem e  $M_{pp_k}$  é a matriz não

invertível da transformação global da câmara na posição e orientação  $k$ .

Como se pode observar da equação anterior, não é possível determinar as coordenadas 3D do ponto  $P$  a partir unicamente das coordenadas do mesmo no plano de uma imagem. Assim, são necessárias as coordenadas do mesmo ponto no plano imagem para outra posição e orientação da câmara, obtendo-se:

$$\begin{bmatrix} w \cdot x_u^{k+1} \\ w \cdot y_u^{k+1} \\ w \end{bmatrix} = M_{pp_{k+1}} \cdot \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix},$$

com

$$M_{pp_{k+1}} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \end{bmatrix},$$

onde

$$(w \cdot x_u^{k+1}, w \cdot y_u^{k+1}, w)$$

são as coordenadas homogêneas não distorcidas do ponto  $P$  no plano imagem da câmara na posição e orientação  $k + 1$ , e  $M_{pp_{k+1}}$  é a matriz não invertível da transformação global da câmara associada.

Utilizando as equações anteriores, obtém-se um sistema de equações lineares sobredeterminado, designado por sistema de projecção estéreo (Tavares, 95b; Coelho, 03):

$$\begin{bmatrix} a_{11} - a_{31} \cdot x_u^k & a_{12} - a_{32} \cdot x_u^k & a_{13} - a_{33} \cdot x_u^k \\ a_{21} - a_{31} \cdot y_u^k & a_{22} - a_{32} \cdot y_u^k & a_{23} - a_{33} \cdot y_u^k \\ b_{11} - b_{31} \cdot x_u^{k+1} & b_{12} - b_{32} \cdot x_u^{k+1} & b_{13} - b_{33} \cdot x_u^{k+1} \\ b_{21} - b_{31} \cdot y_u^{k+1} & b_{22} - b_{32} \cdot y_u^{k+1} & b_{23} - b_{33} \cdot y_u^{k+1} \end{bmatrix} \cdot \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = \begin{bmatrix} a_{34} \cdot x_u^k - a_{14} \\ a_{34} \cdot y_u^k - a_{24} \\ b_{34} \cdot x_u^{k+1} - b_{14} \\ b_{34} \cdot y_u^{k+1} - b_{24} \end{bmatrix}.$$

Trata-se de um sistema com quatro equações lineares para três incógnitas que pode ser resolvido utilizando, por exemplo, o método dos mínimos quadrados com decomposição em valor único ou de Cholesky (Press, 92).

### 3 – RESULTADOS

Na Fig. 8, estão representadas as imagens par, de uma sequência de oito imagens, de um cubo obtidas com rotação incremental de  $3^\circ$  de uma câmara em torno do eixo  $z$ .

As entidades a considerar em cada imagem foram determinadas utilizando-se o detector de Deriche (Deriche, 87; Tavares, 95b), e seguimento com histerese, Fig. 9. Seguidamente, as entidades de cada imagem foram simplificadas, Fig. 10.

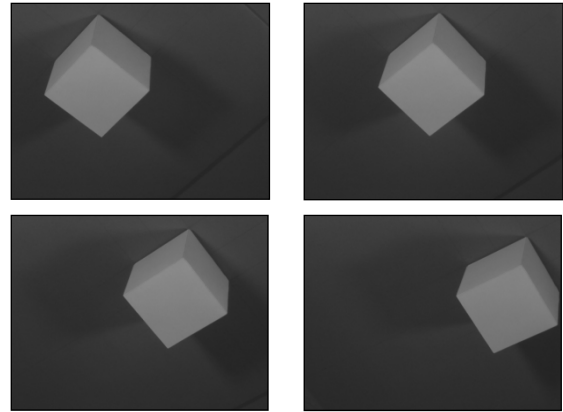


Fig. 8 – Imagens par obtidas com rotação de uma câmara em torno do eixo  $z$ .

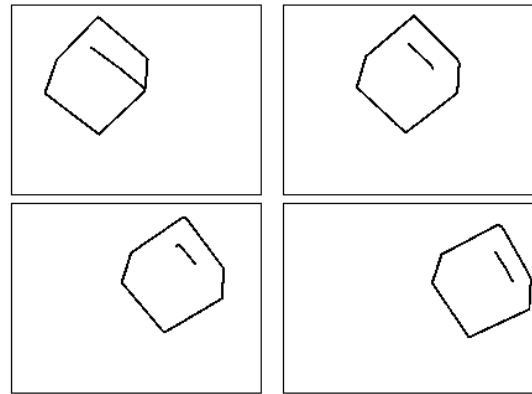


Fig. 9 – Entidades determinadas.

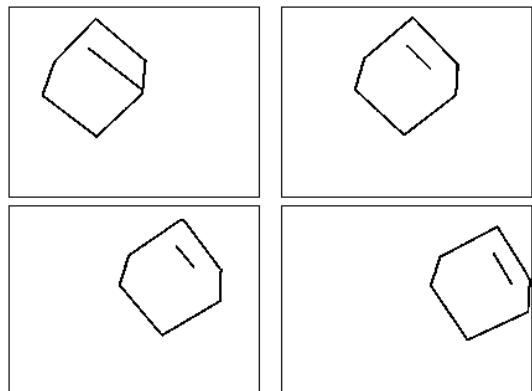


Fig. 10 – Entidades simplificadas.

O seguimento das entidades foi realizado na próxima etapa. Na Fig. 11, pode-se observar o movimento 2D das entidades ao longo da sequência, e as entidades consideradas na primeira imagem e a sua identificação.

Na Tab. 1, estão apresentados os resultados obtidos no seguimento e também descrito o modo como este decorreu. Como

é possível depreender dos resultados apresentados, o processo de seguimento processou-se com sucesso. Dos mesmos, é também verificável que a utilização de restrições geométricas tem como vantagens o contínuo êxito do emparelhamento e, conseqüentemente, o respectivo seguimento, e uma fase de inicialização mais simples e flexível.

Tabela 1 – Descrição do processo de seguimento.

1ª Imagem					2ª Imagem						
Ref.	P. Inic.		P. Fin.		Emp.	Ref.	P. Inic.		P. Fin.		Emp.
1	46	288	223	399	I	1	99	267	265	392	KRG
2	194	42	359	162	I	2	262	35	415	163	F
3	175	131	348	252	I	3	242	126	404	250	F
4	223	399	353	251	I	4	405	262	265	392	KRG
5	192	44	72	196	I	5	135	166	261	36	F
6	72	196	46	288	I	6	99	267	135	166	F
7	353	251	358	164	I	7	405	262	415	165	KRG
3ª Imagem					4ª Imagem						
1	155	255	317	389	KM	1	221	249	371	387	KM
2	328	31	472	165	F	2	399	30	537	174	F
3	309	120	464	253	F	3	383	118	452	185	F
4	463	265	318	388	KM	4	372	387	532	265	KM
5	192	155	328	31	F	5	398	31	249	156	F
6	155	255	192	155	KRG	6	249	156	221	249	KM
7	463	265	473	168	KM	7	532	265	537	175	KM
5ª Imagem					6ª Imagem						
1	274	242	415	388	KM	1	335	239	465	391	KM
2	461	32	590	181	F	2	532	41	648	195	KM
3	441	117	516	198	KRG	3	562	184	511	123	KRG
4	580	280	416	388	KRG	4	643	289	467	392	KM
5	460	33	303	150	KRG	5	524	41	363	150	KM
6	303	150	274	242	KRG	6	363	150	335	239	KM
7	580	280	590	183	KRG	7	643	289	647	196	KM
7ª Imagem					8ª Imagem						
1	396	238	516	397	KM	1	454	240	567	406	KM
2	598	49	706	210	KM	2	659	62	762	226	KM
3	581	132	642	219	KRG	3	649	145	702	235	KRG
4	696	307	517	397	KM	4	567	406	754	320	KM
5	597	50	425	148	KM	5	656	64	482	155	KM
6	425	148	396	238	KM	6	482	155	454	240	KM
7	697	307	706	212	KM	7	754	320	760	229	KM

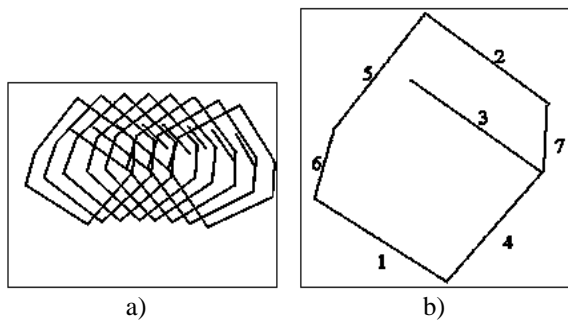
**Legenda:** Ref. - Referência da entidade, I - Inicialização da entidade, P. Inic. - Coordenadas 2D do ponto inicial, P. Fin. - Coordenadas 2D do ponto final, Emp. - Tipo de emparelhamento: KM - Filtros de Kalman e distâncias de Mahalanobis, KRG - Filtros de Kalman e restrições geométricas e F - Não obtido com êxito.

#### 4 - CONCLUSÕES E PERSPECTIVAS FUTURAS

Foi apresentada uma metodologia que permite a obtenção de informação 3D a partir do movimento de uma câmara. Assim, foram descritas as fases: calibração da câmara, detecção e simplificação das

entidades, seguimento das entidades e obtenção das coordenadas 3D.

A calibração permite obter os parâmetros intrínsecos e extrínsecos da câmara, sendo adoptado uma abordagem que utiliza como pontos de calibração vértices de quadrados definidos no plano de calibração.



**Fig. 11** – a) Entidades consideradas; b) ID de cada entidade da primeira imagem da sequência

Para a determinação das entidades, é utilizado um filtro de detecção de orlas de intensidade. Posteriormente, é executado o seguimento e subsequente aproximação poligonal das orlas, utilizando um algoritmo de faixas dinâmicas. Finalmente, é realizada a fusão dos segmentos que presumivelmente pertencem a uma mesma entidade. Desta forma, cada imagem fica reduzida a um conjunto de segmentos de recta de características fortes, representando as entidades presentes mais relevantes.

A fase de seguimento, e consequente obtenção de informação 3D, é solucionada considerando: entidades parametrizadas pelo ponto médio, comprimento e direcção; introdução de ruído na modelação do sistema; utilização de três filtros de Kalman independentes, cada um com um modelo cinemático de aceleração constante localmente; utilização como medida do emparelhamento de distâncias normalizadas de Mahalanobis, ou, quando desta forma o emparelhamento não é obtido, restrições geométricas; utilização do princípio da triangulação estereoscópica para a obtenção das coordenadas 3D, utilização de um modelo de entidades 2D continuamente actualizado. A filosofia utilizada na medição e posterior actualização revelou-se acertada, pois mesmo nos casos em que uma dada entidade não é visível numa imagem e é visível na seguinte, o seu seguimento é obtido com êxito. Também se verificou que, quando os filtros utilizados começam a divergir ou a convergir em demasia, a utilização de restrições geométricas origina que os filtros se auto-ajustem de forma mais rápida e mantenham emparelhamentos com êxito ao longo do

seguimento. A introdução de ruído na modelação, resolveu de forma satisfatória o problema de os filtros poderem convergir em demasia. Também se verificou que a utilização de restrições geométricas teve como consequência tornar mais fácil e flexível a fase de inicialização. A utilização do factor de confiança para a gestão das entidades revela-se de grande utilidade, nomeadamente para o manter sempre actualizado, evitando assim que este crescesse para valores desnecessários.

A metodologia global apresentada possibilita resultados bastante satisfatórios, essencialmente quando os objectos em causa podem ser adequadamente representados por segmentos de recta, tendo-se revelada robusta e eficiente.

Como tarefas futuras podem ser consideradas a adaptação da metodologia para os casos em que o movimento não é conhecido, a consideração de procedimentos de autocalibração da câmara e a obtenção de informação 3D densa.

## REFERÊNCIAS

- Azevedo T.C.S., Tavares J.M.R.S., Vaz M.A.P., 3D Object Reconstruction from Uncalibrated Images using an Off-the-Shelf Camera, *Advances in Computational Vision and Medical Image Processing: Methods and Applications* 117/136, Springer, 2008
- Coelho C.P., Tavares J.M.R.S., Introdução às Técnicas para Levantamento de Forma Tridimensional sem Contacto, *Revista Mecânica Experimental* 9:51/65, 2003
- Crowley J.L., Stelmazyk P., Skordas T., Puget P., Measurement and Integration of 3-D Structures by Tracking Edge Lines, *Int. Journal of Computer Vision* 8(1):29/52, 1992
- Deriche R., Using Canny's Criteria to derive a Recursively Implemented Optimal Edge Detector, *Int. Journal of Computer Vision* 1(2):167/187, 1987
- Deriche R., Faugeras O., Tracking line segments, *Image and Vision Computing* 8(4):261/270, 1990
- Giai-Checa B., Deriche R., Viéille T., Faugeras O., Suivi de segments dans une séquence

- d'images monoculaire, Institut National de Recherche en Informatique et en Automation, 1993
- Jähne B., Haubecker H., Computer Vision and Applications, Academic Press, 2000
- Lenz R.K., Tsai R.Y., Techniques for Calibration of the Scale Factor and Image Center for High Accuracy 3-D Machine Vision Metrology, Transactions on Pattern Analysis and Machine Intelligence 10(5):713/720, 1988
- Leung M.K., Yang Y.-H., Dynamic Strip Algorithm in Curve Fitting, Computer Vision, Graphics, and Image Processing 51(146):146/165, 1990
- Maybeck P.S., Stochastic Models, Estimation, and Control, Academic Press, 1979
- Pinho R.R., Tavares J.M.R.S., Tracking Features in Image Sequences with Kalman Filtering, Global Optimization, Mahalanobis Distance and a Management Model, Computer Modeling in Engineering & Sciences 46(1):51/75, 2009
- Press W.H., Teukolsky S.A., Vetterling W.T., Flannery B.P., Numerical Recipes in C - The Art of Scientific Computing, Cambridge University Press, 1992
- Silva J.A.D., Tese de Doutoramento, Aquisição e Processamento de Informação Tridimensional, Faculdade de Engenharia, Universidade do Porto, Porto, 1994
- Tavares J.M.R.S., Algumas Ferramentas para Visão Tridimensional por Computador, Faculdade de Engenharia da Universidade do Porto, 1995a
- Tavares J.M.R.S., Dissertação de Mestrado, Obtenção de Estrutura Tridimensional a Partir de Movimento de Câmara, Faculdade de Engenharia, Universidade do Porto, 1995b
- Tavares J.M.R.S., Padilha A.J., Matching Lines in Image Sequences with Geometric Constraints, 7th Portuguese Conf. on Pattern Recognition, Aveiro, Portugal, 1995c
- Tavares J.M.R.S., Padilha A.J., A new Approach for Merging Edge Line Segments, 7th Portuguese Conf. on Pattern Recognition, Aveiro, Portugal, 1995d
- Tsai R.Y., A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology using Off-the-Shelf TV Cameras and Lenses, Journal of Robotics and Automation RA-3(4):323/344, 1987