

Controle de uma Plataforma Móvel Utilizando Exclusivamente Realimentação Visual

WALTER FETTER LAGES
JACQUES WALDMANN

Instituto Tecnológico de Aeronáutica
CTA - ITA - IIEEE
12228-900 - São José dos Campos - SP
w.fetter@ieee.org

Abstract: A prototype mobile platform for experimentation on computer vision-based navigation is presented. A line on the workspace floor is tracked by the vision system. The identification of the line parameters is accomplished by use of the Hough transform. Results obtained so far shows that provided the initial conditions are favorable, the line can be tracked. Whenever the initial conditions are not so favorable, the line is missed. The reasons for this failure are discussed and some solutions are suggested.

Introdução

Em ambientes de automação industrial existe o problema de movimentação de peças entre as máquinas para, por exemplo, alimentação de matéria prima. Para que esta movimentação possa ser feita de forma automática poderia ser utilizado um veículo sobre trilhos com um CLP para determinação dos pontos de parada. No entanto, esta solução exige a instalação de trilhos, que representam problemas, no caso de necessidade de alteração do lay-out das máquinas.

Estes problemas poderiam ser evitados se a movimentação fosse feita por um veículo sem trilhos. Porém, este tipo de veículo não possui uma trajetória natural. Para tratar este problema, este artigo apresenta uma plataforma móvel, capaz de seguir uma linha marcada no chão de sua área de trabalho. O sistema é capaz de adquirir imagens e processá-las, obtendo os erros de posicionamento e orientação que serão utilizados pelo controlador da plataforma.

A Plataforma Móvel

Na Figura 1 tem-se um esboço da plataforma móvel e as definições dos sistemas de coordenadas utilizados para a modelagem cinemática da mesma.

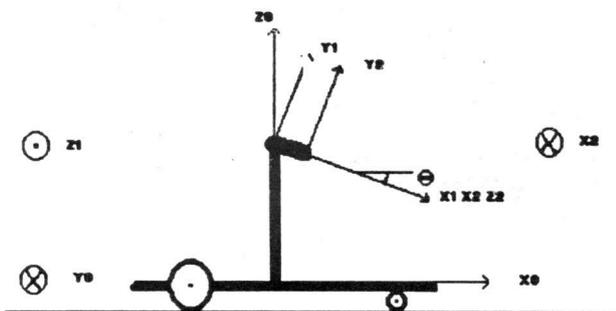


Figura 1: Esboço da plataforma móvel.

A plataforma móvel utilizada movimentar-se apoiada em três rodas.

Modelagem Cinemática

Os pontos (x_i, y_i) na imagem serão então mapeados em pontos no sistema de coordenadas (X_2, Y_2, Z_2) fixo no plano de imagem através das expressões:

$${}^2x_i = K_x \frac{x_i}{\lambda} (\lambda - {}^2z_i) \quad {}^2y_i = K_y \frac{y_i}{\lambda} (\lambda - {}^2z_i) \quad (1)$$

onde:

(x_i, y_i) = coordenadas do ponto na imagem (pixels);

λ = distância focal (m);

$({}^2x_i, {}^2y_i, {}^2z_i)$ = coordenadas do ponto no sistema de coordenadas (X_2, Y_2, Z_2) (m);

K_x = fator de escala de conversão de pixels para metros;

K_y = fator de escala de conversão de pixels para metros; e

2z_i = profundidade do ponto 3-D com imagem no ponto (x_i, y_i) (m).

Para o sistema em questão, os fatores de escala K_x e K_y , desprezando-se as distorções provocadas pelas lentes, são $K_x = 8,7786 \times 10^{-5}$ m/pixel e $K_y = 1,1220 \times 10^{-5}$ m/pixel. A distância focal é $\lambda = 11$ mm [David (1992)].

O ponto (x_i, y_i) da imagem expresso em coordenadas homogêneas no sistema de coordenadas (X_2, Y_2, Z_2) será então ${}^2p_i = [{}^2x_i, {}^2y_i, {}^2z_i, 1]^T$. Por sua vez, pontos no sistema de coordenadas (X_2, Y_2, Z_2) serão descritos no sistema de coordenadas (X_1, Y_1, Z_1) através da seguinte transformação:

$${}^1P_2 = {}^1T_2 {}^2P = \begin{bmatrix} 0 & 0 & 1 & a_2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} {}^2P \quad (2)$$

Similarmente, entre pontos no sistema de coordenadas (X_1, Y_1, Z_1) e pontos no sistema de coordenadas (X_0, Y_0, Z_0) tem-se a seguinte transformação:

$${}^0P_1 = {}^0T_1 {}^1P = \begin{bmatrix} \cos(\theta_1) & -\sin(\theta_1) & 0 & 0 \\ 0 & 0 & -1 & 0 \\ \sin(\theta_1) & \cos(\theta_1) & 0 & d_1 \\ 0 & 0 & 0 & 1 \end{bmatrix} {}^1P \quad (3)$$

Conseqüentemente, um ponto no sistema de coordenadas (X_2, Y_2, Z_2) será mapeado no sistema de coordenadas (X_0, Y_0, Z_0) através da transformação:

$${}^0P_2 = {}^0T_1 {}^1T_2 {}^2P = {}^0T_2 {}^2P \quad (4)$$

Os parâmetros são $d_1 = 20$ cm, $a_2 = 6,5$ cm e $\theta_1 = 45^\circ$, definidos conforme as convenções de Denavit-Hartenberg [Fu et al 1987]

Processamento da Imagem

A imagem é obtida com resolução de 512x480, sendo então sub-amostrada para obter-se uma imagem com resolução de 128x120, à semelhança do que foi feito em [Kak-Meng (1993)]. Este artifício aumenta a velocidade de processamento da imagem e não prejudica sensivelmente a capacidade do sistema de detectar a fita.

A detecção da fita na imagem é feita via transformada de Hough [Kak-Meng (1993)][Fu et al. (1987)] do gradiente de intensidade da imagem. Foi utilizado um acumulador com 32 valores para θ (ângulo entre a reta que passa pela origem e é normal à reta a ser identificada e o eixo horizontal) e 64 valores para ρ (distância da origem à reta a ser identificada normal), ambos uniformemente distribuídos, com $\theta \in [0, \pi)$ e $\rho \in (-128\sqrt{2}, 128\sqrt{2})$ para uma imagem com 128x120 pixels. O fato de θ ter sido discretizado em um número relativamente pequeno de valores foi explorado montando-se uma tabela para os valores de $\sin(\theta)$ e $\cos(\theta)$ ao invés de calculá-los a todo instante.

Após a montagem do acumulador de Hough, procura-se os pontos correspondentes às arestas direita e esquerda da fita sobre o chão. Para tanto, são detectados os dois maiores picos do acumulador e é testado se estes picos são maiores do que um certo limiar de detecção. Com isto procura-se evitar a detecção de picos gerados pelo ruído presente na imagem. Se não for possível achar dois picos acima do limiar, a detecção da fita é dada como falhada e a plataforma interrompe o seu movimento imediatamente e aguarda a intervenção de um operador humano. O valor de limiar para detecção de picos foi determinado experimentalmente como 100. Este valor é um compromisso entre a capacidade de detectar picos que realmente existem e a capacidade de rejeitar picos devido ao ruído na imagem. Quando este limiar é excessivamente alto, os picos gerados pelas bordas da fita são perdidos. Por outro lado, quando este limiar é baixo demais, o ruído torna-se capaz de gerar picos que não correspondem a arestas na imagem real.

Obtidos os parâmetros ρ e θ das retas correspondentes às duas arestas da fita, são escolhidos dois pontos para representar cada reta. Estes pontos são escolhidos como sendo os pontos de intersecção da reta com os limites da imagem. Desta forma os pontos estarão, em geral, afastados o bastante para não distorcer a reta. Os quatro pontos são então mapeados do sistema de coordenadas da imagem para o sistema de coordenadas (X_0, Y_0, Z_0) , fixo na base da plataforma, utilizando-se as expressões (1-4). Obtidas as duas retas neste novo sistema de coordenadas, é calculada a reta média, isto é, a reta que está equidistante das duas arestas detectadas. Esta reta é utilizada para calcular os erros que serão utilizados pelo sistema de controle. Ao invés de detectar as duas arestas e calcular a reta média poder-se-ia executar um *thinning* da fita e detectar uma única aresta. No entanto, o método utilizado além de ser mais rápido do que o *thinning*, é potencialmente mais poderoso [Rosenfeld-Kak (1976)], pois utiliza informações mais globais do que o *thinning*.

Modelagem Dinâmica e Controle

Como a plataforma deve apenas seguir a fita, não existe nenhum referencial natural para o sistema. Pode-se optar então por fazer o controle no sistema de coordenadas (X_0, Y_0, Z_0) , que está fixo na plataforma. Neste sistema de coordenadas, o único grau de liberdade do sistema é rotação. Tem-se então

$$\theta = \frac{v}{r} \tan(\phi) \quad (6)$$

onde:

θ = velocidade angular
 v = velocidade linear na direção do eixo X_0 ;
 ϕ = ângulo da roda de direção em relação ao eixo X_0 ; e
 r = distância do eixo traseiro à roda de direção.

Para controlar o sistema é utilizada uma estratégia semelhante à que é utilizada quando se está dirigindo um automóvel: seleciona-se uma velocidade linear e atua-se apenas sobre o ângulo da roda de direção. Utilizando-se um atuador com três níveis para ϕ tem-se:

$$\phi = \phi_{\text{máx}} \quad \text{se } e_y < 0; \quad (7)$$

$$\phi = 0 \quad \text{se } e_y = 0; \quad (8)$$

$$\phi = -\phi_{\text{máx}} \quad \text{se } e_y > 0. \quad (9)$$

onde:

e_y = erro lateral.

Implementação

O sistema implementado para controle da plataforma móvel utiliza como *host* um microcomputador 80386SX25. Aos slots deste microcomputador estão conectadas uma placa de digitalização de imagem Data Translation DT2851 com resolução de 512x480x256 e uma placa aceleradora SuperCard SC-1/AT da CSPI, com processador 80860 a 33 MHz [Lages (1993)]. A plataforma móvel é conectada através de uma interface própria em qualquer porta de impressora padrão IBM-PC. Uma câmera de vídeo Sony modelo HVM-302 é montada na plataforma móvel e é conectada à placa digitalizadora DT2851.

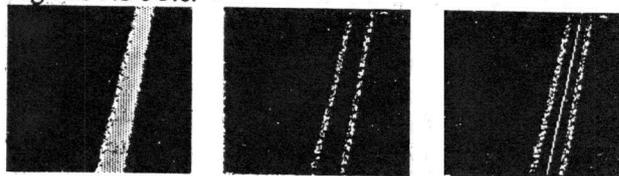
A interface para acionamento da plataforma móvel aciona os motores utilizando a técnica de PWM (*Pulse Width Modulation*), com uma frequência de chaveamento de 30 KHz.

Embora os algoritmos para processamento da imagem e controle em tempo real sejam executados no 80860, este processador está montado em uma placa coprocessadora e não se tem acesso aos dispositivos de I/O do IBM-PC. Conseqüentemente, o processamento de entrada e saída (aquisição de imagem e controle dos motores) é executado no 80386SX25.

Resultados

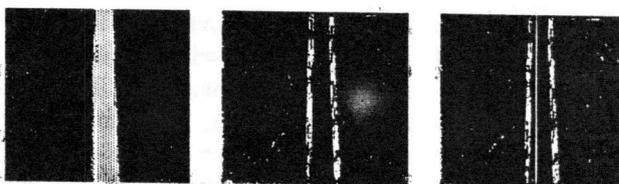
Os experimentos realizados em laboratório revelaram que, em geral, a plataforma consegue efetivamente seguir a linha marcada no chão. Na Figura 2.a tem-se a imagem de uma situação típica de operação do sistema. A Figura 2.b apresenta o gradiente de intensidade desta imagem. Na Figura 2.c é mostrada a linha identificada sobreposta ao gradiente de intensidade da imagem original. Para esta imagem foi calculado um erro

angular de $0,217 \text{ rd} = 12,413^\circ$ e um erro lateral de $-3,3 \text{ cm}$. A Figura 3.a mostra uma imagem obtida alguns instantes após a obtenção da imagem da Figura 2.a. Para esta imagem foi obtido um erro angular de 0 e um erro lateral de $0,1 \text{ cm}$. O gradiente de intensidade desta imagem e a linha identificada podem ser vistos nas Figuras 3.b e 3.c.



(a) (b) (c)

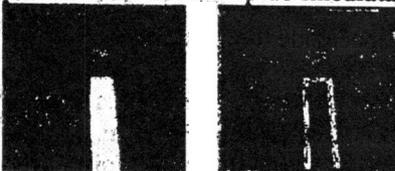
Figura 2: Imagens com erro angular e erro lateral.



(a) (b) (c)

Figura 3: Imagens sem erro angular e erro lateral.

Quando a plataforma alcança o final da linha, obtém-se a imagem mostrada na Figura 4.a. O gradiente de intensidade desta imagem pode ser visto na Figura 4.b. Note-se que nesta situação a área da imagem ocupada pela fita é menor do que a área ocupada nas figuras anteriores. Devido a este fato, os picos gerados no acumulador de Hough terão uma intensidade menor do que o limiar de detecção. Neste caso, a plataforma não consegue identificar a linha, conforme já discutido. Conseqüentemente, o sistema assume que a fita foi perdida e o movimento da plataforma é interrompido imediatamente.



(a) (b)

Figura 4: Fita não identificada pelo sistema.

Idealmente, esta situação de perda da fita deveria ocorrer somente no final da trajetória da plataforma. No entanto, foi verificado que a perda da fita também ocorre quando o erro lateral da plataforma é excessivo. Aparentemente o que ocorre nesta situação é que o controlador faz com que o ângulo entre o eixo X_0 (ver Figura 1) da plataforma e a fita seja grande, fazendo

com que a plataforma mova-se rapidamente em direção à fita. Geralmente, este movimento faz com que a fita saia do campo de visão da câmera, sendo então gerada uma imagem onde a fita não aparece e portanto não pode ser identificada. É importante notar que esta situação de erro lateral excessivo ocorre apenas devido a condições iniciais desfavoráveis. Ou seja, quando a condição inicial da plataforma apresenta um erro lateral pequeno, o controlador mantém este erro lateral dentro de limites que não causam a perda da fita.

Este problema poderia ser solucionado utilizando-se um controlador que manobrasse a plataforma de modo mais suave. Porém, um controlador como este permitiria um erro de rastreamento da fita maior.

Por outro lado, a perda da fita deve-se ao fato da câmera estar fixa na plataforma, fazendo com que um movimento da plataforma implique um movimento da câmera. A utilização de visão ativa, ou seja, de um sistema de visão que permita a movimentação da câmera independentemente da plataforma, permitiria movimentar a plataforma sem que a fita saia do campo de visão. Além disso, se eventualmente a fita saísse do campo de visão, a câmera poderia ser movimentada de forma a obter novamente uma imagem onde a fita possa ser identificada.

Conclusões

O sistema de aquisição de imagem possui uma resolução relativamente alta. Isto pode parecer uma vantagem a princípio, mas em geral processar todos estes pixels consome muito tempo. Neste trabalho, este problema foi tratado através de uma sub-amostragem da imagem. No entanto, esta solução nem sempre é satisfatória, pois são descartados vários pixels adjacentes, que poderiam conter detalhes importantes da imagem. Seria mais interessante reduzir o tamanho da imagem através de uma média entre os valores de pixels adjacentes, porém este procedimento toma tempo. Um sistema de aquisição de imagem com uma resolução menor realizaria esta média no momento de captação da imagem.

A detecção de retas utilizando-se a transformada de Hough é um método bastante robusto, pois tolera bem relações sinal/ruído relativamente baixas [Fu et al. (1987)], [Duda-Hart (1972)]. No entanto o seu uso implica em uma série de considerações de implementação, como limiar de gradiente para incrementar o acumulador, resolução do acumulador e limiar para detecção de picos no acumulador. A escolha de valores adequados para estes parâmetros depende de experimentação [Ballard (1981)]. A determinação apenas destes parâmetros consiste na procura de um ponto em um espaço de dimensão 4. Difícilmente se

encontrará o ponto ótimo. Felizmente, em geral, a determinação destes valores não é muito crítica.

Atualmente, a principal limitação do sistema é devido à possibilidade da fita escapar do campo de visão da câmera. Dotando-se o sistema de visão ativa pretende-se minimizar este problema. A inclusão de visão ativa no sistema cria uma série de possibilidades para desenvolvimento do sistema, principalmente quanto à estratégias para coordenar o movimento da plataforma e os movimentos da câmera, de forma a manter a fita sempre no campo de visão.

Embora os motores utilizados no robô sejam de baixa potência, a técnica utilizada para acioná-los (PWM) é bastante adequada para acionamento de motores de grande potência. Esta característica confere ao circuito uma grande versatilidade, pois trocando-se apenas o circuito de potência pode-se controlar motores bem maiores. O PWM implementado é totalmente controlado digitalmente, o que dispensa o uso de conversores AD-DA, reduzindo o custo do sistema.

Uma vantagem importante de um sistema de navegação baseado em visão é a sua mínima interação com o ambiente, pois o sensor é basicamente passivo. Outros tipos de sensores como sonares ou radares têm maior potencialidade de gerar interferências indesejáveis em outros equipamentos.

Referências

- D. H. Ballard, Generalizing the Hough Transform to Detect Arbitrary Shapes, *Pattern Recognition*, vol. 13, pp. 111-122, 1981.
- M. B. David, Controle Dinâmico de Manipuladores Robóticos com Realimentação Visual, Tese de Mestrado, ITA, 1992.
- R. O. Duda & P. E. Hart, Use of Hough Transformation to Detect Lines and Curves in Pictures, *Comm. ACM*, vol. 15, pp. 11-15, 1972.
- K. S. Fu, R. C. Gonzalez, C. S. G. Lee, Robotics: Control, Sensing, Vision and Intelligence - McGraw-Hill, 1987.
- A. C. Kak & M. Meng, Mobile Robot Navigation Using Neural Networks and Nonmetrical Environment Models, *IEEE Control Systems* vol. 13 n. 5, October, 1993.
- W. F. Lages, Ambiente para Controle em Tempo Real e Supervisão de Manipuladores Robóticos, Tese de Mestrado, ITA, 1993.
- A. Rosenfeld & A. C. Kak, Digital Picture Processing, Academic Press, New York, 1976.