

ANÁLISE DE DESEMPENHO DE TÉCNICAS DE VISÃO COMPUTACIONAL

Alfred Makoto Kabayama

Instituto Tecnológico de Aeronáutica, Divisão de Engenharia Mecânica - Aeronáutica
12228-900, São José dos Campos, SP, Brasil
e-mail: alfred@mec.ita.cta.br

Luís Gonzaga Trabasso

Instituto Tecnológico de Aeronáutica, Divisão de Engenharia Mecânica - Aeronáutica
12228-900, São José dos Campos, SP, Brasil
e-mail: gonzaga@mec.ita.cta.br

RESUMO. *Este trabalho apresenta a implementação e comparação de três técnicas distintas de visão computacional tridimensional a saber:*

- *Visão estéreo: correlação de imagens bidimensionais (2D);*
- *Fusão sensorial - uso de sensores distintos: câmera 2D + sensor de ultra-som unidimensionais (1D);*
- *Iluminação estruturada.*

Os algoritmos de visão computacional apresentados são basicamente implementações de trabalhos anteriores, realizados a partir de técnicas atualizadas de comprovada eficiência. A principal contribuição deste trabalho é disponibilizar para a comunidade científica técnicas distintas de visão computacional tridimensional (3D), bem como um método de avaliação de quais das técnicas disponíveis é a mais adequada para determinada aplicação. As principais características avaliadas durante o estudo comparativo entre as três técnicas são:

- *Influência do meio (ruído devido a uma iluminação não uniforme, ofuscamento, sombras);*
- *Levantamento da infra-estrutura para cada técnica;*
- *Análise de incertezas, precisão (repetibilidade) e acuidade.*

Optou-se pelo uso do software Matlab versão 5.1 para a implementação dos algoritmos, dada a facilidade de programação e depuração, além de ser muito difundido e utilizado no meio acadêmico, permitindo que os resultados desse trabalho sejam reproduzidos e verificados. São apresentados exemplos de visão tridimensional aplicados a tarefas de montagem robotizada.

Palavras-chave: Visão computacional, Sensores de Distância, Robótica, Mecatrônica.

1. INTRODUÇÃO

Com o aumento do número de robôs industriais, a robótica está se tornando uma área da engenharia com identidade própria. O avanço e barateamento de tecnologia de sensores, viabiliza cada vez mais a construção de robôs com maior capacidade de realimentação (*feedback*) de dados sobre o ambiente de trabalho, aumentando, desta maneira, a acuidade de posicionamento, a velocidade de operação e flexibilidade dos mesmos. É consensualmente aceito que a visão computacional é o mais poderoso e flexível meio de realimentação de

informações disponíveis sobre o ambiente e que tem canalizado um considerável esforço em pesquisa e desenvolvimento para a área.

A determinação dos dados tridimensionais a partir das imagens bidimensionais é de grande importância no campo de visão computacional e uma das formas mais diretas de se obtê-la é através da visão estéreo. Na reconstrução de uma cena tridimensional a partir de um par de imagens estéreo dois problemas básicos devem ser resolvidos. O primeiro e mais complexo é o problema de correspondência ou correlação de imagens, ou seja, considerando um ponto da cena projetado na primeira imagem, determinar o ponto na segunda imagem que corresponde à projeção deste mesmo ponto da cena. O segundo problema é o cálculo geométrico da posição tridimensional da parte da cena a partir de suas projeções, usando-se a técnica de triangularização, como exposto em Fu (1987).

A visão estéreo pode ser baseada em áreas ou regiões, que utiliza o domínio métrico de similaridade de imagem no processo de correspondência. Nesta abordagem, divide-se a cena em sub-regiões e aplica-se a medida de similaridade fotométrica em todas as sub-regiões para encontrar quais regiões das duas imagens que possuem maior semelhança.

A técnica de fusão sensorial consiste em utilizar um sistema de visão para realizar uma série de medidas bidimensionais e combiná-las com os dados provenientes de um sensor de distância independente, de maneira a completar a informação tridimensional necessária do objeto. Exemplos de medidas bidimensionais são: área, comprimento e largura, centro de área e a inclinação do eixo principal do objeto em relação à horizontal. De posse destas informações, é possível determinar a atitude e abertura das garras de manipuladores robóticos e posiciona-los espacialmente de forma correta em relação ao centro de área do objeto imageado.

As aplicações da iluminação estruturada tem sido feitas principalmente nas áreas de inspeção e controle de qualidade. Este método é baseado no uso da informação geométrica contida na iluminação mostrada em imagem 2D. Ela é adquirida pela projeção de um padrão de luz apropriado (usando um projetor de luz de alta potência ou um laser) sobre o objeto e pela observação das deformações que a forma do objeto produz no padrão de luz. O padrão de luz pode ser uma série de pontos geometricamente arranjados, linhas paralelas, ou uma simples linha de luz, dependendo da forma do objeto e da aplicação.

2. VISÃO ESTÉREO

O sistema de visão estéreo deste trabalho foi idealizado para aplicações robóticas do tipo *pick-and-place*, mostrado esquematicamente na Fig.1

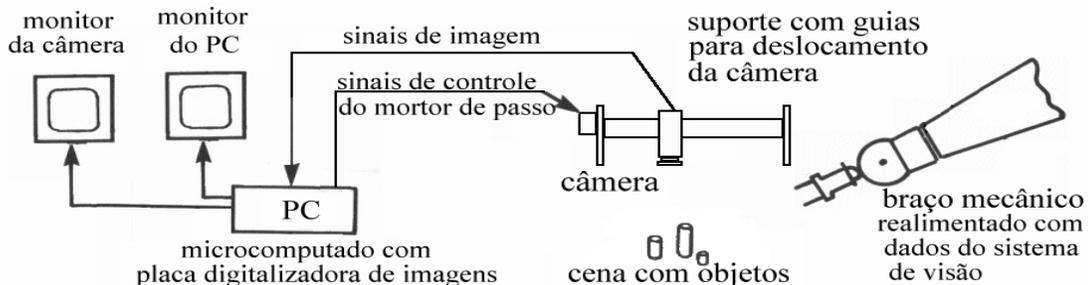


Figura 1: Diagrama do sistema de visão estéreo

A captura das imagens estéreo é feita em duas etapas, com a câmera posicionada na direção vertical, em relação à cena. Após a captura da primeira imagem, a câmera é deslocada lateralmente (curso de 0,5 a 1,5 cm), através de suporte mecânico por um motor de passo. Efetuado o deslocamento, a segunda imagem da cena é então capturada.

2.1 Implementação da visão estéreo

O desenvolvimento e os testes do algoritmo foram realizados em quatro etapas.

A primeira foi a programação das rotinas de correlação de imagens no software MatLab® versão 5.1, usando imagens sintéticas, geradas no software de computação gráfica 3D Studio versão 4.0, da Autodesk®. Trabalhou-se inicialmente com imagens de resolução reduzida (200x300 pixels) para acelerar processo de desenvolvimento das rotinas devido a grande carga computacional envolvida nesta técnica. Com a otimização dos parâmetros do algoritmo de correlação de imagens de maneira a melhorar o desempenho, iniciou-se a segunda etapa do procedimento, que foi a aplicação do algoritmo em imagens reais, capturadas por câmeras, com resolução de 512x512 pixels e 64 níveis de cinza.

A terceira etapa foi a realização de uma calibração intrínseca da câmera, cujo parâmetro de maior interesse é a distância focal, que foi utilizado para a recuperação da informação de profundidade.

A quarta etapa foi a recuperação de dados tridimensionais sobre a cena a partir das imagens 2D. Através da primeira imagem, que é capturada em uma posição inicial pré estabelecida, obtém-se informações em pixels sobre os objetos em cena, tais como comprimento e largura. Para recuperar a informação métrica dos objetos, é necessário realizar uma calibração da relação entre as distâncias em escala métrica e a escala medida em pixels. A informação sobre a altura do objeto é obtida a partir da técnica de triangularização que é dependente da modelagem geométrica da formação da imagem na câmera na configuração estéreo.

2.2 Processo de correlação de imagens por rotulação relaxada

O processo de rotulação relaxada é um processo iterativo desenvolvido heurísticamente para resolver problemas que satisfazem certas restrições de ordem óptica e geométrica e tem se tornado uma técnica padrão em sistemas de visão computacional e reconhecimento de padrões, segundo Pelillo (1994). A busca do resultado começa com uma solução inicial grosseira e que passará por um processo de refinamento iterativo até que se obtenha um resultado estável. Frequentemente, estes algoritmos estão associados a uma função de "energia" que quantifica o grau de violação das restrições e que são minimizados ao longo do processo iterativo.

O processo de correlação de imagens, ou *stereo matching*, implementado foi seguindo o algoritmo proposto por Feris (1998).

2.3 Modelo de câmera e calibração

O modelo de câmera considerado foi o "*pinhole*", como mostrado em Fu (1987) e em Nalwa (1993): a projeção de um objeto tridimensional para uma superfície bidimensional é feita através de linhas retas que passam por um único ponto chamado centro de projeção.

A calibração de câmeras no contexto das máquinas de visão computacional tridimensionais é a determinação das características ópticas e geométricas internas da câmera, ou seja, os chamados parâmetros intrínsecos. Consiste também a determinação da posição e orientação da câmera em relação a um certo de sistema de coordenada do mundo, os chamados parâmetros extrínsecos. Laudares (1997) apresenta com detalhes a calibração extrínseca de câmeras para as aplicações robóticas propostas neste trabalho.

O parâmetro intrínseco mais importante a ser determinado nesta aplicação é a distância focal λ , ou seja, a distância do ponto central da lente até o plano sensor. Este parâmetro foi estimado utilizando-se parte da técnica implementada por Vianna (1999). Neste artigo é

apresentada uma técnica de calibração de parâmetros intrínsecos e extrínsecos de câmeras composta por dois estágios. O primeiro é composto pela aquisição de uma série 4 imagens de um padrão quadriculado cujas posições dos pontos no mundo são conhecidas e então usadas como dados em um processo de otimização linear considerando o modelo da câmera ideal, isto é, sem distorções de lentes.

2.4 Recuperação do perfil 3D do objeto (triangularização)

Considerando-se as seguintes hipóteses:

- As câmeras são idênticas.
- Os sistemas de coordenadas das duas câmeras estão perfeitamente alinhadas, diferindo apenas na localização das suas origens;
- A coordenada Z é mesma para ambas as câmeras;

Segundo Fu (1987), a recuperação da informação de profundidade, ou seja, a coordenada Z, é obtida pela seguinte equação:

$$Z = \lambda - \frac{\lambda \cdot B}{x_2 - x_1} \quad (1)$$

onde:

- λ é a distância focal estimada pelo processo de calibração implementado experimentalmente.
- x_2 e x_1 são as coordenadas em metros, que são determinadas à partir da multiplicação do valor destes, em pixels, pela razão entre a dimensão do sensor na direção x, em metros, e a resolução do sensor, em pixels, nesta mesma direção,.

As distâncias X e Y podem ser determinadas em relação ao sistema de coordenadas localizado em uma das câmeras, bastando realizar uma calibração de um fator de escala que relacione o deslocamentos em metros pelo correspondente deslocamento em pixels. As distâncias X e Y serão o resultado do produto da coordenada em pixels pelo seu respectivo fator de escala. O procedimento de calibração extrínseco da posição do sistema de coordenadas da câmera em relação ao sistema de coordenadas do robô permite que este adquira a informação 3D de que necessita para a operação de localizar e pegar o objeto na cena.

3. FUSÃO SENSORIAL

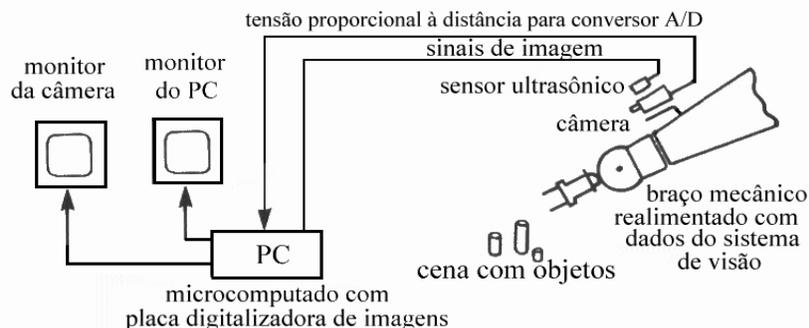


Figura 2: Diagrama do sistema de visão combinado com o sensor à ultra-som.

O conjunto de fusão sensorial concebido para obtenção da visão 3D é mostrado na Fig. 2. O sensor utilizado foi o modelo 946-A4V-2D-2CO-175E da Honeywell® e que possui as seguintes características principais:

- Alcance mínimo: 200 mm (programável)
- Alcance máximo: 2000mm (programável)

Peso: 210 gramas

O sensor fornece como medida uma tensão analógica proporcional à distância a ser medida e esta proporção pode ser direta ou inversa (*rising* ou *falling edge*), dependendo de como o sensor for programado. Foram levantados as curvas do comportamento da variação da tensão em função da variação da distância, que não foram fornecidos pelo fabricante.

Devido às características do perfil do feixe de ultra-som, além de uma calibração do sensor para determinar a sua resolução em distância, foi necessário levantar o perfil radial deste feixe.

Tabela 1: medidas do levantamento do perfil radial do feixe de ultra-som

Distâncias (cm)	Diâmetro do feixe (cm)
20	3
30	8
40	16
90	13
140	11
180	6

As distâncias na Tabela 1 são referentes ao topo do objeto.

O cálculo do diâmetro do feixe de ultra som foi feito através de um método experimental, aproximando-se se o objeto do ponto central onde o sensor apontava na superfície da bancada de testes. Sobre a bancada de testes foi colado um papel milimetrado e nele foram registrados vários pontos de detecção em cada nível medido e em várias direções no plano horizontal, da periferia para o centro do raio de ação do sensor. Os níveis foram ajustados movimentando-se o sensor até que atingissem as distâncias mostradas na Tabela 1.

A partir do levantamento das características do sensor, é possível estimar as dimensões dos objetos que podem ser manipulados utilizando esta técnica. O tamanho dos objetos não devem ser menores que os limites radiais do feixe de ultra som, no nível correspondente, para que tenham sua altura medida eficientemente. O material não pode ter propriedades que absorvam as ondas de ultra som e o topo do objeto deve ser perpendicular ao feixe que o ultra som incide.

4. ILUMINAÇÃO ESTRUTURADA

Para este trabalho, foi feito um estudo utilizando dois padrões simples para uma avaliação da acuidade métrica da técnica e de sua viabilidade e aplicação em sistemas de operação do tipo *pick-and-place* com braços mecânicos.

O primeiro padrão estudado que foi utilizado foi da luz proveniente de um dispositivo laser do tipo usado em “pointers” comerciais. Foi adaptado uma fonte de alimentação DC externa devido ao desgaste das pilha e conseqüente degradação da intensidade da luz.

A linha cheia e a linha pontilhada mostradas na Fig.3 representam o feixe laser em duas cenas distintas na aquisição de dados para a recuperação da informação da altura h do objeto.

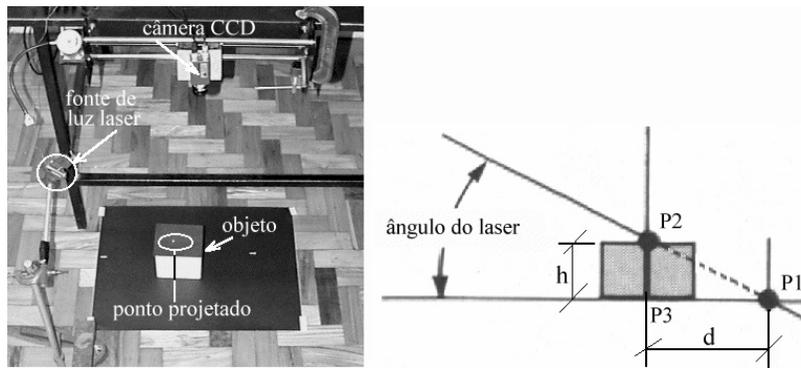


Figura 3: Montagem da experiência utilizando o laser como fonte de luz característica e respectivo modelo

Na primeira cena, representado pela linha pontilhada, o objeto não é filmado e a posição do centro de área da projeção do feixe laser quando atinge o chão na posição **P1** é determinado na imagem pelo sistema de visão. Na segunda cena, representado pela linha cheia, o objeto é filmado e a posição do centro de área da projeção do feixe laser quando atinge o topo do objeto na posição **P2** é determinado na imagem pelo sistema de visão.

O laser atinge o chão e o objeto com um ângulo θ e d é distância, em pixels, entre os centros de área dos pontos **P1** e **P2**.

Obtém-se a seguinte relação para a determinação da altura h do objeto a partir das relações trigonométricas do triângulo retângulo formado por **P1**, **P2** e **P3**, como pode ser observado no modelo da Fig. 3.

$$h = d \cdot s \cdot \text{tg}\theta \quad (2)$$

onde:

$d \in \mathbf{Z}^+$

$s \in \mathbf{R}^+$, é um fator de conversão experimental que relaciona centímetros com pixels.

$\theta \in \mathbf{Z}^+$, é o ângulo entre os segmentos de reta **P1P2** e **P1P3**.

4.1 Implementação

Em primeiro lugar, realizou-se um procedimento para determinação da razão s . Com um objeto padrão, em centímetros, realizou-se cinco medidas do comprimento do objeto em pixels. Este comprimento em pixels é medido na mesma direção onde ocorre a variação da distância d . Este detalhe é muito importante pois o formato dos pixels não é quadrado e a razão s obtida em outro sentido produz resultados diferentes. A razão s é o resultado da divisão entre a medida em centímetros do objeto e a média das medidas de comprimento em pixels. Neste caso, utilizou-se a largura de uma folha A4 em cm dividido pela média de 5 medidas da distância vertical correspondente em pixels na imagem binária feita pelo próprio sistema de visão, obtendo-se:

$$s = 21.1 / 387 \text{ (cm / pixel)}$$

Em segundo lugar, realizou-se a calibração para a determinação do ângulo θ da seguinte maneira. Utilizou-se cinco objetos de alturas h conhecidos para servirem como padrão de calibração. Foram então determinados a posição dos centros de área dos pontos laser filmados nas posições **P1** e **P2** cinco vezes, para cada objeto. Foram calculados os ângulos θ para cada par **P1** e **P2** medido a partir da Eq.(3), obtida a partir da Eq.(2):

$$\theta = \arctg\left(\frac{h}{d.s}\right) \quad (3)$$

Esses dados são mostrados na Tabela 2 abaixo.

Tabela 2 : Dados e resultados obtidos do processo de calibração de s e θ .

objeto	Altura h (cm)	Média de 5 ângulos calculados em graus
O1	1,5	44,9280
O2	4,5	45,2576
O3	6,9	46,3332
O4	11,7	45,6472
O5	14,7	42,9280
	Média dos ângulos →	45,0188
	Desvio padrão →	1,2805

Após a calibração de s e de θ , realizou-se uma série de medidas de alturas de objetos utilizando os parâmetros determinados através da Eq. (2). Os resultados das medições são mostrados na Tabela 3.

Tabela 3 : Medição de objetos utilizando dados obtidos no processo de calibração de s e θ .

objeto	Altura h (cm)	Medida da altura (cm) Média de 5 medidas	Erro (%)
O1	15,9	16,7929	5,62
O2	9,3	9,3187	0,20
O4	6,1	5,9795	1,97

O segundo padrão utilizado foi uma linha reta projetada por um retroprojetor como mostrado nas Fig. 4 e 5.

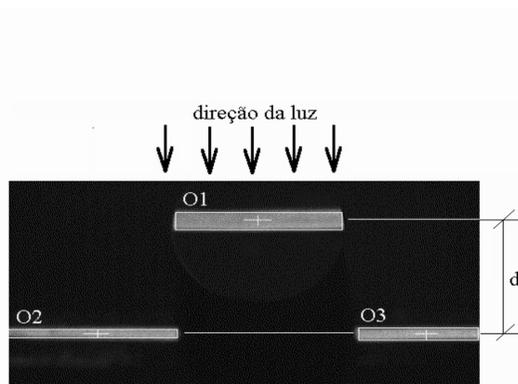
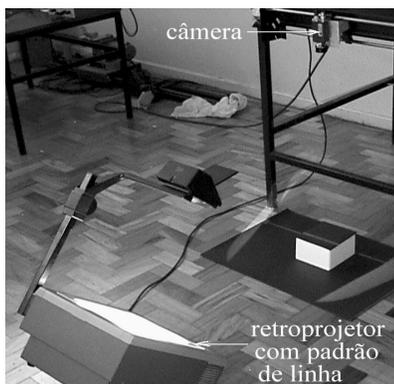


Figura 4: Vista geral do experimento Figura 5: Detalhe do padrão projetado

A recuperação da informação da altura do objeto é similar ao caso do laser, utilizando o mesmo princípio e equações. A diferença é que as linhas projetadas formam três objetos que são reconhecidos pelo sistema de visão após a obtenção da imagem binária como mostra a

Fig.5. Devido a erros de digitalização, nem sempre o alinhamento entre os objetos **O2** e **O3** é obtido. Por este motivo, a distância **d** é resultado da diferença da média entre as coordenadas no sentido da variação da deformação dos objetos **O2** e **O3**, pela respectiva coordenada do objeto **O1**. O objeto e o fundo mostrados na Fig. 4 e 5 são pretos.

Aproveitou-se o resultado da calibração do s da experiência anterior. Procedeu-se da mesma maneira para a calibração do ângulo θ , utilizando-se de cinco medidas de cada um dos cinco objetos padrão.

Tabela 4 : Resultados obtidos do processo de calibração de s e θ .

Altura h (cm)	Média de 5 ângulos calculados em graus
Média dos ângulos→	38,0321
Desvio padrão→	1,5887

Após a calibração de s e de θ , realizou-se uma série de medidas de alturas de objetos utilizando os parâmetros determinados através da Eq.(2)

Tabela 5 : Medição de objetos utilizando dados obtidos no processo de calibração dos parâmetros s e θ .

Objeto	Altura h (cm)	Medida da altura (cm) Média de 5 medidas	Erro (%)
O1	6,2	6,0345	2,67
O2	11,2	12,0865	7,91
O3	7,8	7,9260	1,61

5 - Análises e conclusões

A seguir são apresentadas algumas análises e conclusões feitas à partir dos dados obtidos experimentalmente e das dificuldades observadas durante o decorrer da implementação de cada uma das técnicas apresentadas.

Técnica	Aplicações, problemas e recomendações
Visão estéreo	<ul style="list-style-type: none"> • Qualidade do sistema de iluminação influencia na qualidade do processo de correlação de imagens, uma vez que a captura das duas imagens digitais ocorre em momentos diferentes. Ocorrência de cintilações nas lâmpadas são prejudiciais, assim como a variação da iluminação ambiente podem ser fatores interferentes. • A carga computacional desta técnica é pesada e o tempo de processamento dos algoritmos é considerável. Poder-se-ia utilizar microcomputadores com processamento mais veloz, procurar otimizar a estrutura de processamento dos códigos ou converte-los em uma linguagem compilada (C++) , cuja execução é mais rápida que em linguagens interpretadas (Matlab)

Fusão sensorial	<ul style="list-style-type: none"> • O sensor de ultra som possui restrições em relação ao tamanho máximo dos objetos, materiais e formato. O tamanho dos objetos não devem ser menores que os limites radiais do feixe de ultra som para que tenham sua altura medida eficientemente. O material não pode ter propriedades que absorvam as ondas de ultra som e o topo do objeto deve ser perpendicular ao feixe que o ultra som incide. • O uso de um sistema de iluminação alternativo, que gere alto contraste entre fundo e objeto é recomendável porque torna o processo de limiarização mais robusto à interferências luminosas do meio ambiente e diminuiria a necessidade de uma calibração freqüente dos níveis dos limiares. Uma técnica de iluminação que poderia ser usado é o <i>backlighting</i> [Galbiati(1990)] • A necessidade de calibração da relação cm/pixel (fator de escala) é um detalhe importante na implementação do sistema.
Iluminação estruturada	<ul style="list-style-type: none"> • É a técnica mais sensível à interferências da iluminação externa devido à constante necessidade de calibração dos níveis de limiarização; especialmente no caso do uso do projetor de padrão de linha, cuja intensidade da luz é menor que a do laser. • Calibração do ângulo de aplicação do padrão e do fator relação cm/pixel (fator de escala), cada vez que ocorre mudanças de posição do projetor • Um isolamento do ambiente da cena é recomendável para melhor desempenho do sistema e propiciar mais robustez à interferências da iluminação ambiente. • A aplicação é limitada em função da diferença de contraste entre objeto e fundo, que não pode ser muita alto, pela dificuldade de calibração dos valores dos níveis de limiarização.

Infraestrutura para implementação dos experimentos

Técnica	Componentes
Equipamentos comuns em todas as técnicas	<ul style="list-style-type: none"> • Microcomputador PC • Sistema de visão (Frame grabber, monitor do sistema de visão, câmera e software)
Visão estéreo (utilizando uma câmera)	<ul style="list-style-type: none"> • Mecanismo de deslocamento da câmera.
Fusão sensorial	<ul style="list-style-type: none"> • Sensor ultra som • Placa conversora analógico/ digital • Sistema backlight de iluminação (sugestão de melhoria).
Iluminação estruturada	<ul style="list-style-type: none"> • Projetor de padrões • Laser • Isolamento da iluminação ambiente (sugestão de melhoria)

REFERÊNCIAS

Feris, R.S. & Lages, W.F., 1998, Casamento de Imagens utilizando correlação e rotulação

relaxada, Anais do IV ENCITA- Encontro de Iniciação Científica e Pós-Graduação do ITA

Fu, K.S. & Gonzales, R.C. & Lee C.S.G., 1987, Robotics, Mc Grall Hill,

Galbiati, L. J., 1990, Machine vision and digital image processing fundamentals
, Prentice Hall

Laudares, D., 1997, Procedimento automático para calibração de sistemas de visão robótica
para operações *pick-and-place*, Tese de mestrado ITA

Nalwa, V.S., 1993, A guided tour of computer vision, Adson-Wesley Publishing Company

Pelillo, Marcello, 1994, On Dynamics of Relaxation Labeling Processes, IEEE Transactions
on pattern analisys and machine intelligence

Ruocco S.R., 1987, Robot Sensors and transducers, Open University Press Robotics Series
, Edited by P.G. Davey

Viana, S. A. & Waldmann, J. & Caetano, F.C., 1999, Non-linear optimization based batch
balibration with accuracy evaluation (in print)

Performance evaluation of 3D computer vision techniques

SUMMARY. This work presents the implementation and comparison of three different techniques of three-dimensional computer vision as follows:

- Stereo vision - correlation between two 2D images
- Sensorial fusion - use of different sensors: camera 2D + ultrasonic sensor (1D);
- Structured light

The computer vision algorithms presented in this work are basically implementations of previous works, executed with actual techniques of proven efficiency. The main contribution of this work is to make available for the scientific community different 3D computer vision techniques, as well as an evaluation of which method is more appropriate for certain applications. The comparative study among the three techniques took into consideration the following characteristics:

- Computational effort (elapsed time for obtain the 3D information);
- Influence of the environment conditions (noise due to an non uniform lightning, overlightning, shades);
- The cost of the infrastructure for each technique;
- Analysis of uncertainties, precision (and accuracy

The option of employing the Matlab software, version 5.1, for the implementation of the of computer vision algorithms was due to the simplicity of commands, programming and debugging. Besides, this software is known and used in the academic community, allowing the results of this work to be reproduced and verified. Examples of three-dimensional vision presented are applied to robotics assembling tasks.

Key words: computer vision, range finders, robotics, mechatronics.