

Um Dispositivo Automático para Obtenção da Forma 3D de Objetos

John Kennedy Schettino de Souza
Mário Fernando Montenegro Campos

{jkss, mario}@dcc.ufmg.br

Departamento de Ciência da Computação
Instituto de Ciências Exatas
Universidade Federal de Minas Gerais
Av. Antônio Carlos, 6627, 31270-010 Belo Horizonte, MG- Brasil

Abstract. This article presents an automatic device capable of acquiring the three-dimensional shape of surfaces of arbitrary objects. It is a low-cost equipment with an excellent performance in terms of precision and repeatability. It uses a structured light plane and a CCD camera to determine, through triangulation, the coordinates of each point of the object, producing a dense depth map. It can be used in many applications such as quality control in assembling lines, reverse engineering (e. g. analysis of worn out mechanical components), as well as in obtaining models to be applied in virtual reality. The functioning and assembling methodologies as well as a performance analysis of the device are presented in this article.

Keywords. Three-dimensional shape, Reconstruction, Structured light, Computer Vision.

1 Introdução

A competitividade a nível mundial traz consigo a demanda por serviços, processos e produtos de alta qualidade. Muitas são as aplicações onde se faz necessária a obtenção ágil e automática da forma tridimensional de objetos com o objetivo de se ajustar uma determinada linha de produção para controle rígido da qualidade. Objetos como peças automotivas, implantes dentários, próteses e órteses para citar apenas alguns, podem ter seu processo produtivo grandemente melhorado ao se fazer uso desses dispositivos. Outras aplicações de grande importância são abordadas pela área denominada de engenharia reversa, onde o propósito é o de se determinar, por exemplo, o desgaste de uma peça após um ciclo de trabalho, podendo-se assim aprimorá-la em implementações futuras. Nesse sentido, o número de sistemas desenvolvidos para obtenção automatizada de informações de uma cena, tem crescido em ritmo acelerado e diversas modalidades, utilizando técnicas distintas tem sido disponibilizadas.

Tendo em vista solucionar problemas como os mencionados, foi projetado um dispositivo de baixo custo e de fácil operação, que possibilita a extração da forma tridimensional de objetos. No protótipo desenvolvido apenas a superfície frontal é obtida, mas a extensão para obtenção de todo o objeto é imediata e se constituirá em um dos próximos passos no desenvolvimento dessas técnicas. O dispositivo é composto fundamentalmente de um emissor de luz estruturada (laser) e uma câmera CCD (Charged Couple Device) que são deslocados linearmente em relação ao objeto. A câmera captura cada uma das interseções do plano de luz com o objeto, e por meio de uma das técnicas mostradas a seguir,

determina a profundidade de cada ponto sobre a linha. O modelo geométrico adotado calcula a profundidade por meio de triangulação simples, assumindo a projeção perspectiva. Outro método com características interessantes é o que se baseia na invariância da razão-dupla (cross-ratio) [Richard O. Duda (1973)]. Em ambos os casos existe a necessidade de correção das dimensões Y e X do objeto.

A utilização de luz no espectro visível traz consigo dificuldades como os problemas relacionados a superfície do objeto como a sua cor, reflectância e mesmo a sua forma. Em particular, concavidades presentes na superfície do objeto (oclusão) continuam a ser problemáticas [D. H. Ballard (1982)] e não solucionadas pelo dispositivo. Entretanto, o mapa de profundidade obtido é bastante preciso e atende a aplicações a nível de reconstrução.

O problema de obtenção da forma tridimensional e as soluções existentes são discutidos na seção 2. A solução proposta e implementada neste trabalho é discutida em mais detalhes na seção 3 seguido dos conceitos básicos sobre técnicas de projeção perspectiva e invariância da razão-dupla. Resultados experimentais são apresentados na seção 4 e as conclusões na seção 5.

2 Aquisição da forma 3D

O problema mais geral a ser resolvido consiste na obtenção da estrutura tridimensional de objetos existentes na cena, a partir de imagens bidimensionais. Muitos são os trabalhos na literatura que descrevem as várias abordagens existentes utilizadas para a obtenção da estrutura 3D de objetos. Dentre os diversos métodos, realçamos aqueles que utilizam luz estruturada. Nesses casos,

uma fonte de energia externa é projetada sobre a cena [Stuart C. Shapiro (1987)] e um sensor capta a interação desta fonte de energia com o objeto. Esta interação da luz com o objeto é capturada por um detector (sensor ou câmera) e transformada em informação de profundidade. Isso é possível, pois em geral a geometria (ângulos e distância da fonte e do detector) é conhecida.

Dentre os métodos que fazem parte desta categoria estão as técnicas Moiré [R. A. Jarvis (1983)] e [Z. Jason Geng (1996)] que projetam padrões de luz estruturada sobre a cena. Este padrão é capturado por uma câmera posicionada com um deslocamento lateral em relação ao projetor de luz. A informação de profundidade é extraída através de um movimento conhecido entre a câmera e objeto ou através do deslocamento de fase. As dificuldades encontradas neste método são: a correspondência do contorno gerado pelo padrão, a limitação da resolução lateral que não pode exceder o espaço da grade geradora do padrão e a variação da reflexividade da superfície do objeto, que interfere na formação do contorno. Uma outra técnica empregada é a de absorção de luz [Denis Laurendeau (1992)] que utiliza uma câmera 2D trabalhando com absorção diferencial de luz através de um líquido colorido. Embora este sistema forneça uma alta densidade de dados e seja de baixo custo, ele necessita que o objeto inspecionado esteja imerso no líquido.

Uma das técnicas mais populares para extração de profundidade é a que utiliza um plano de luz projetado perpendicularmente sobre a superfície do objeto e captura a interseção deste plano com o objeto através de uma câmera [R. A. Jarvis (1983)]. Uma das limitações deste método é a velocidade de obtenção da informação, pois para se obter um mapa denso de profundidades é necessário um número maior de projeções do plano de luz sobre a superfície do objeto.

Visando superar a limitação de velocidade, outros trabalhos foram apresentados em [K. L. Boyer (1987)] e [Z. Jason Geng (1996)] onde a codificação da luz estruturada permite a extração de profundidade da cena, podendo o processamento ser feito em tempo real. Entretanto, estas técnicas ficam limitadas a objetos com superfícies neutras (branca) pois a coloração da superfície do objeto pode produzir informação incorreta de profundidade.

Uma outra abordagem utilizando a variação de foco é apresentada em [Shree K. Nayar (1996)], onde a informação é obtida em tempo real e o problema de cena com baixa textura é resolvido utilizando luz estruturada. Esta técnica, entretanto, está restrita a ambientes estruturados.

3 Abordagem utilizada

Como discutido anteriormente, cada uma das abordagens descritas apresentam peculiaridades distintas. Algumas delas requerem dispositivos ópticos mais avançados, outros demandam maior tempo computacional. Como um dos objetivos é a construção de um dispositivo de baixo custo, de fácil implementação, que utilize apenas componentes off-the-shelf sem comprometer a precisão, optamos pelo método da luz estruturada.

O sistema proposto neste trabalho utiliza-se de uma fonte de luz estruturada e de um sensor que captura a luz refletida pelo objeto presente na cena. A Figura 1 apresenta a estrutura básica do sistema.

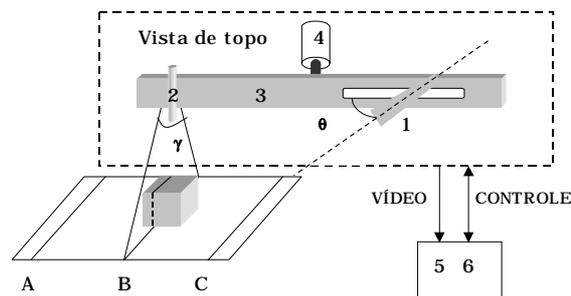


Figura 1 Configuração do Sistema.

Como pode ser observado na Figura 1, o sistema é composto fundamentalmente de dois módulos: um de varredura e um de controle. O módulo de varredura (vista de topo) é constituído de uma câmera (1) e um emissor de laser (2), montados sobre um mesmo suporte (3) rígido. Este conjunto tem liberdade de movimento horizontal; produzido por um sistema de translação (4).

O módulo de controle é constituído de um microcomputador (6) equipado com uma placa digitalizadora de vídeo (frame grabber) (5) responsável pela captura das imagens provenientes da câmera. Através da interface paralela do microcomputador são enviados os comandos para movimentação do motor de passos.

A informação de profundidade é obtida baseado na cena da Figura 2a onde é apresentada a projeção do plano de laser sobre a superfície do objeto. Já a Figura 2b apresenta a respectiva imagem formada no CCD após captura e aplicação do filtro binário. É importante observar que os eixos H e V correspondem as dimensões horizontal e vertical do CCD.

Como pode ser observado o plano de laser refletido pela superfície de sustentação do objeto está sendo projetado no CCD formando uma coluna descontínua (J) próxima da origem do eixo horizontal do CCD. Já a reflexão do plano de laser relativa a superfície do objeto formou uma coluna (K) no CCD deslocada da origem do eixo horizontal.

que o processo acima seja repetido até que toda a superfície do objeto tenha sido capturada. Ao término desta etapa o Sistema apresenta o mapa de profundidade (Figura 6b) associado ao objeto (Figura 6a), onde a profundidade de cada ponto do objeto é expressa através de níveis de cinza. Como pode ser observado, os níveis de cinza mais claros representam maiores profundidades.

A altura do objeto corresponde ao eixo Y do mapa de profundidade gerado pelo Sistema. Na Figura 6, embora o objeto apresente a mesma altura ao longo de seu comprimento, o mapa de profundidade apresenta diferentes alturas (Y) para diferentes posições de X.

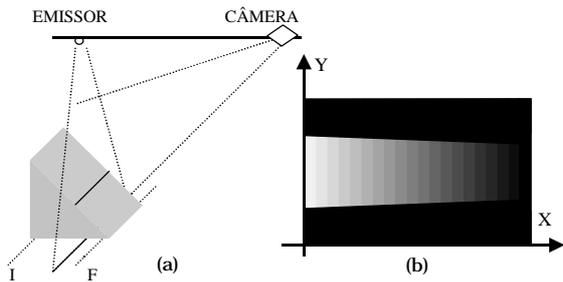


Figura 6 (a) Objeto (b) mapa de profundidade.

Para corrigir o problema apresentado é necessário que seja feito um ajuste da dimensão Y. A geometria deste problema pode ser observada na Figura 7.

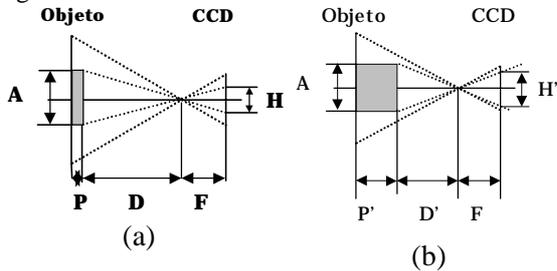


Figura 7 Projeção Y.

Um objeto de altura A e com uma profundidade P gera uma imagem de altura H (Figura 7a). Já no segundo caso (Figura 7b) um objeto de mesma altura A e com profundidade P', gera sobre o plano de imagem uma projeção H', sendo que H' > H.

Para corrigir este problema, basta aplicar a relação extraída da Figura 8.

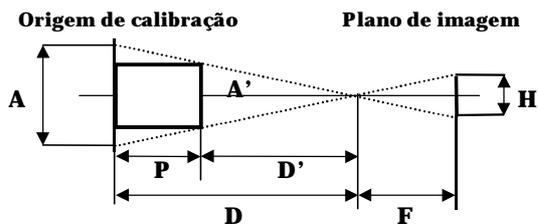


Figura 8 Correção Y.

Supondo A a máxima altura possível de um objeto tratado pelo Sistema e H sua projeção no plano de imagem, então: $\frac{A}{D} = \frac{H}{F}$. Para corrigir as

medidas dos pontos da superfície de um objeto que estejam deslocados da posição de calibração (D) é necessário que uma nova relação seja traçada. Sendo H e F constantes do Sistema e P a profundidade extraída pelo Sistema, basta estabelecer esta nova relação: $\frac{A'}{D-P} = \frac{H}{F}$ que apresentará o valor real da altura A'.

A dimensão X do objeto é extraída de forma bem simples. De acordo com Figura 9a é possível observar quatro marcações distintas A, B, C, e D. O intervalo de [A,B] indica a região que antecede ao objeto. O intervalo]B,C[é a região onde o objeto encontra-se posicionado e finalmente o intervalo [C,D] indica a região posterior ao objeto. Neste exemplo, o Sistema faz o imageamento partindo da marcação A até atingir a marcação D. No intervalo [A,B] as projeções do plano de laser não interceptam o objeto. A cada uma destas projeções, é associado um Z=0 como indica a Figura 9a.

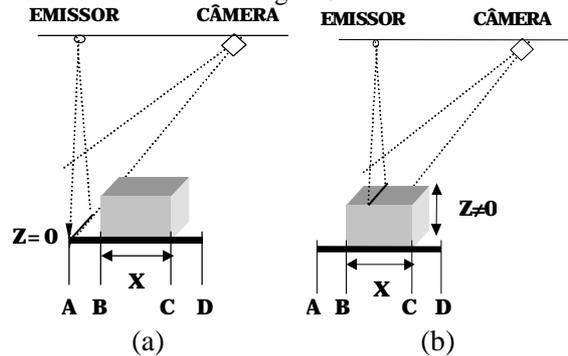


Figura 9 Extraído a dimensão X.

A partir da marcação B, as projeções passam a interceptar o objeto e Z assume valores $\neq 0$ (Figura 9b). O imageamento estende-se até a marcação D, mas somente no intervalo definido entre]B,C[é observado Z $\neq 0$. Ao término do imageamento o Sistema processa esta informação e indica o número de capturas (N) com Z $\neq 0$. Considerando L (mm) o deslocamento do módulo de varredura a cada captura, basta fazer NL para se obter a dimensão X.

4 Integração do Sistema

O sistema foi implementado com componentes disponíveis no mercado, e dessa forma o esforço concentrou-se na integração de cada uma dessas partes. Uma imagem do Sistema pode ser vista na Figura 10. O emissor de laser utilizado no sistema é o Lasiris SNF-X-670-10. O laser é do tipo "estado

sólido”, com a frequência de 690 nm e potência de 10mW que é responsável pela geração do plano de laser projetado sobre a cena.



Figura 10 Sistema de aquisição 3D.

Como mecanismo de translação foi utilizado o mecanismo de transporte de uma impressora matricial de 132 colunas. O acionador é o próprio motor de passos da impressora original, porém o seu controle é realizado diretamente pelo microcomputador de controle do sistema. O frame grabber o DT-3155 da Data Translation apresenta capacidade máxima de geração de 30 quadros por segundo com numa resolução de 640X480 com 256 níveis de cinza. A câmera de vídeo utilizada é monocromática, modelo XCM 77 da Sony O microcomputador é genérico, e na implementação atual foi um Pentium 75, com 16 Mbytes de RAM. Ele é utilizado para controle e integração do sistema. O Borlandc 5.0 foi utilizado como ferramenta de desenvolvimento rodando sobre o sistema operacional Windows95.

5 Resultados

A fim de se poder verificar a precisão do sistema, foi usinado um gabarito em alumínio com uma concavidade em forma de "V". Esse gabarito possui dimensões 100x160x100 mm. Três vistas do gabarito foram selecionadas para se obter medidas de dimensões e características de reflexão distintas.

Na Vista 1 (Figura 11) são analisadas altura, largura e alguns pontos de profundidade.

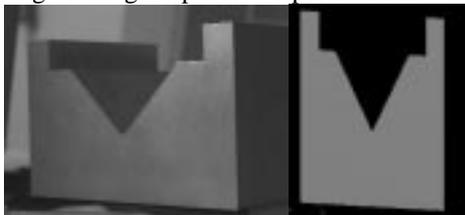


Figura 11 Vista 1 e mapa profundidade.

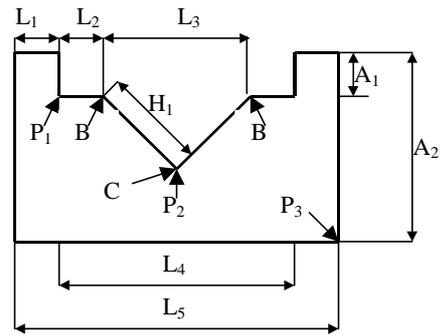


Figura 12 Vista 1 Dimensões analisadas.

Como pode ser observado nas Figuras 14, 15 e 16 os valores percentuais de L_1 , L_2 e A_1 sofreram maiores variações. Isto ocorreu pois nos experimentos realizados cada deslocamento do módulo de varredura corresponde a 0,5mm o que representa uma coluna do mapa de profundidade. Desta forma se, no momento da obtenção das medidas, qualquer falha for cometida pelo operador na ordem de duas colunas (dois pixels na dimensão horizontal) corresponderá a um erro de 5% em relação a seus valores. É importante observar que L_3 e H_1 são mais suscetíveis a erros. Isto pode ser entendido uma vez que os pontos indicados por B e C apresentam maiores dificuldades para suas obtenções.

A faixa de erro médio percentual das medidas ficou em $\pm 3,3\%$. O desvio padrão no máximo foi 1,8% com exceção da medida L_2 onde o desvio padrão é de 2,35%, o que indica uma boa repetibilidade

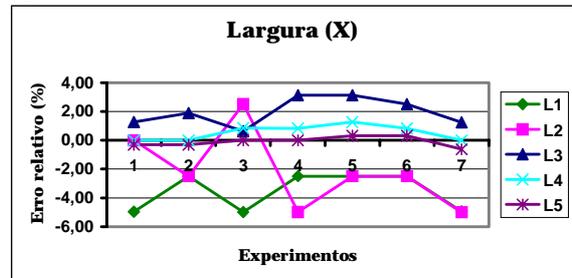


Figura 13 Erro % nas medidas de largura.

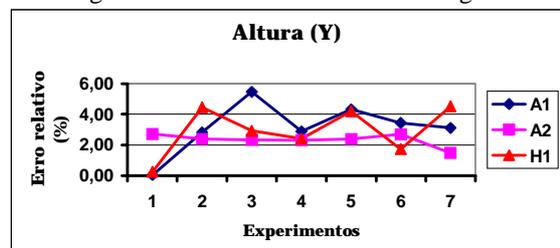


Figura 14 Erro % nas medidas de altura.

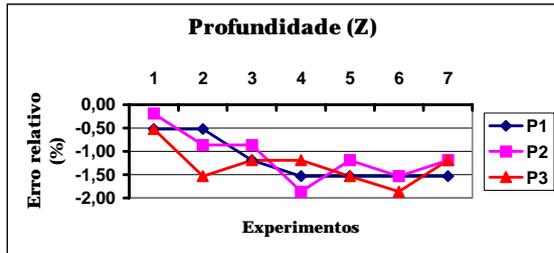


Figura 15 Erro % nas medidas de profundidade.

Na Vista 2 (Figura 16) são analisadas altura, largura e alguns pontos de profundidade.



Figura 16 Vista 2 mapa profundidade.

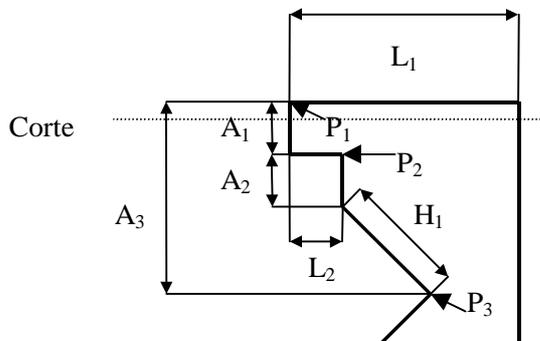


Figura 17 Vista 2 Medidas analisadas.

Com a Vista 2 os resultados (Figuras 18, 19, 20) obtidos ficaram próximos dos resultados encontrados no experimento com a Vista 1. Os maiores erros percentuais estão associados as medidas H_1 e A_2 que apresentaram as mesmas características das medidas H_1 e A_1 da Vista 1. Como na Vista 1 as medidas de profundidade mantiveram com um baixo erro percentual.

Para ilustrar as variações de profundidade em uma superfície plana foi traçado um gráfico contendo os níveis de cinza em uma linha da Vista 2. O resultado obtido pode ser visto na Figura 21. Os níveis de cinza da região de corte apresentam valores aproximadamente constantes, pois até a coluna 77 apresenta nível de cinza de 127 e as demais colunas assumem o valor de 128

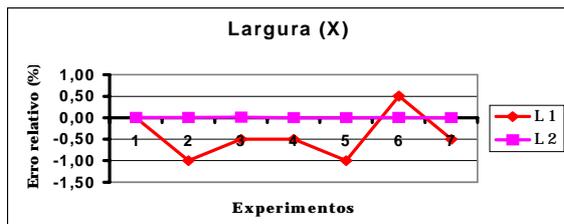


Figura 18 Erro % nas medidas de largura.

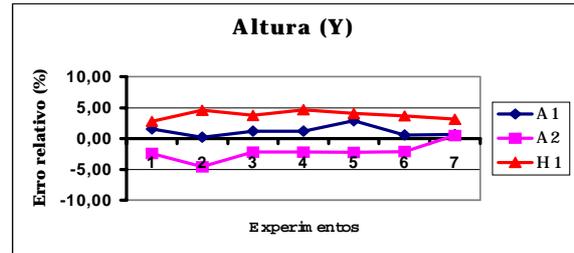


Figura 19 Erro % nas medidas de altura.

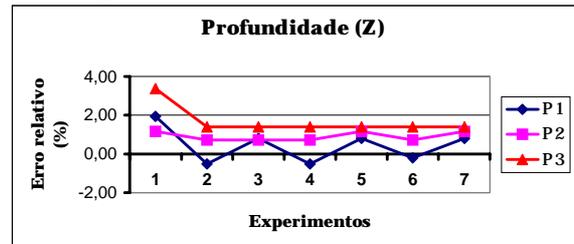


Figura 20 Erro % nas medidas de altura.

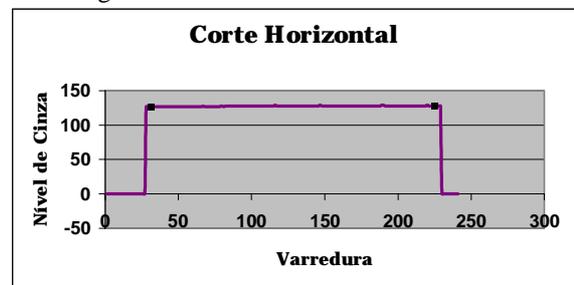


Figura 21 Corte horizontal na Vista 2

A Vista 3 apresenta regiões com profundidades distintas e superfícies com reflexão diferenciada, devido suas inclinações, o que torna o Sistema mais suscetível a erros. Com esta vista será dado um maior enfoque nas medidas de profundidade. Comparando os mapas de profundidade da Vista 3 com as vistas anteriores é possível observar que a imagem obtida para esta vista apresenta mais detalhes e conseqüentemente um grau de dificuldade maior para extração dos pontos de interesse.

Neste experimento existe a necessidade de um tratamento diferenciado com relação aos experimentos anteriores. A luz do ambiente e a abertura do diafragma influenciam na informação capturada pelo Sistema. Para estes experimentos foi utilizada uma abertura de 2,5 (f-stop).

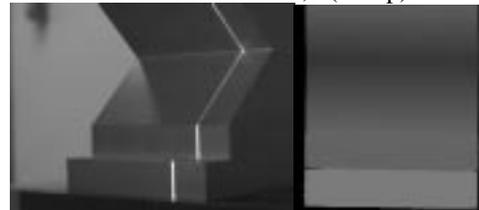


Figura 22 Vista 3 e mapa profundidade.

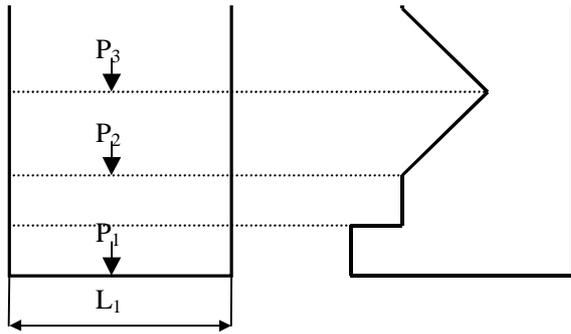


Figura 23 Vista 3 Dimensões analisadas.

Neste experimento o maior erro percentual encontrado foi sobre o ponto P3 (Figura 25) sendo este da ordem de 1,68%. O ponto P₃ foi definido na interseção dos planos inclinados da peça. Neste ponto é gerada uma dupla reflexão proveniente destas superfícies. Como pode ser observado na Figura 22 existe um espalhamento dos feixes refletidos o que ocasiona um valor de profundidade um pouco maior que o real.

Os demais erros de profundidade e largura ficaram abaixo de 1%. Neste experimento pode ser observada uma alta repetibilidade, uma vez que, o desvio padrão ficou na faixa de 0,47%.

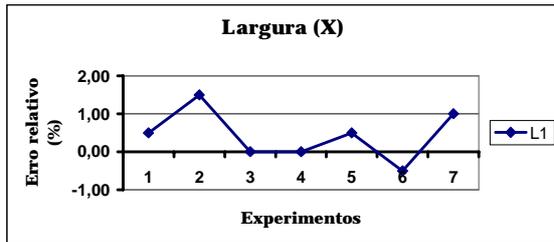


Figura 24 Erro % nas medidas de largura.

Para a análise de desempenho do Sistema foram realizados experimentos e levantados os tempos relativos aos seguintes processamentos: MP (Tempo de movimentação do motor de passos), SO_{MP} (Tempo consumido pelo S.O. durante o movimento do motor de passos), Captura (Tempo relativo a captura de imagem), SO_{CI} (Tempo consumido pelo S.O. durante o processo de captura de imagens), Tratamento (Tempo relativo ao tratamento da imagem capturada), SO_{TI} (Tempo consumido pelo S.O. durante o processo de tratamento das imagens).

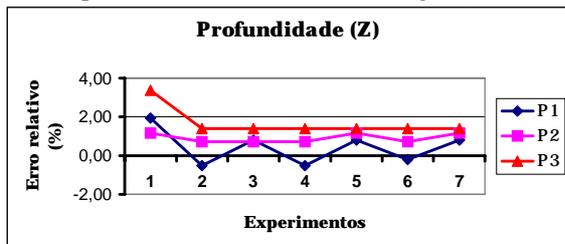


Figura 25 Erro % nas medidas de profundidade.

Nesta análise, foram realizados 10 experimentos para cada varredura. Para cada um dos experimentos, três varreduras diferentes foram utilizadas sendo de 5, 10 e 15mm o que corresponde respectivamente em 100, 200 e 300 capturas.

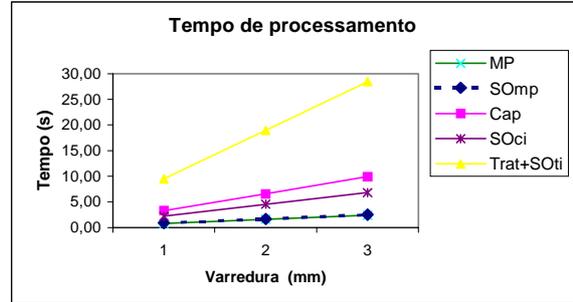


Figura 26 Gráfico de tempo de processamento.

Como pode ser observado na Figura 26 a variação de tempo é linear e dependente do número de capturas. A maior parte do tempo é consumida no tratamento da imagem, sendo assim, uma melhora na resposta do Sistema pode ser obtida utilizando uma máquina de maior velocidade ou substituindo etapas de software por elementos de hardware.

Já o tempo mínimo de captura, fica limitado ao número de quadros capturados por segundo, pelo frame grabber.

Os tempos do sistema operacional crescem proporcionalmente ao volume de processamento. Esta carga pode ser suprimida desde que seja utilizado um dispositivo dedicado (hardware/software).

6 Conclusões e Direções Futuras

O Sistema implementado baseou-se em técnicas de Visão Computacional para obtenção da estrutura tridimensional da superfície de objetos. Os resultados obtidos, no aspecto repetibilidade, apresentaram um erro percentual máximo de 2,35% e no aspecto precisão, um erro médio percentual $\pm 3,3\%$. Entretanto, estes erros foram observados em medidas onde haviam maiores dificuldades para obtenção das coordenadas, uma vez que, elas são obtidas a partir de informações visuais fornecidas pelo mapa de profundidade. Valores mais precisos podem ser alcançados aplicando-se técnicas de detecção de bordas, onde, os ruídos presentes nesta região serão minimizados. Os erros informados acima foram obtidos experimentalmente, entretanto, um estudo analítico deve ser desenvolvido.

Pode-se também utilizar dispositivos de captura e digitalização de imagem com maiores resoluções, motores de passos com menor ângulo de passo e componentes com padrão óptico já disponíveis no mercado. É importante frisar que mesmo na

configuração atual, o Sistema é bastante útil para tarefas como manipulação robótica do tipo pick and place atendendo também aplicações voltadas à reconstrução de objetos.

A utilização de componentes de maior resolução permite um aumento significativo no volume útil do Sistema. Na situação atual o *frame grabber* apresenta-se como limitador da dimensão Z, pois sua resolução horizontal (640 pixels) é menor que a da câmera (756 pixels).

Para o processo de calibração é possível construir uma base de sustentação (mesa com marcações métricas) para o dispositivo de varredura que permitirá o levantamento mais preciso dos parâmetros a serem informados ao Sistema, como, a distância do emissor ao objeto e do emissor a câmera de vídeo.

O Sistema atual permite apenas a obtenção da superfície do objeto que esteja voltada para a câmera e o laser. Outras vistas podem ser obtidas desde que se movimente o objeto consistentemente, com a manutenção de marcas fiduciárias que permitam a combinação das diversas imagens. Pretende-se resolver esse problema em um próximo protótipo onde uma mesa giratória servirá para suportar o objeto a ser imageado. Com aquisições sucessivas poder-se-á ter um arquivo com a descrição completa do objeto. Entretanto, alguns problemas de oclusão, como concavidades presentes no objeto, não poderão ser resolvidos pelo Sistema.

O mapa de profundidade é disponibilizado em formato texto com as coordenadas X, Y e Z para cada ponto dispostas de forma sequencial. O mapa de profundidade deverá ser formatado possibilitando sua transferência para outros sistemas como as máquinas CNC e outras já bastante difundidas na indústria nacional.

Com relação a iluminação é possível trabalhar em ambientes não estruturados, como o caso dos experimentos realizados. Nestas situações basta regular a abertura da lente evitando a saturação parcial ou total do CCD.

O tempo de resposta do Sistema cresce linearmente em função do número de capturas realizadas. Do tempo total de processamento cabe ao algoritmo de tratamento da imagem a maior parcela. Devido à simplicidade deste algoritmo, pretende-se implementá-lo em hardware por meio de FPGA's permitindo assim a aquisição em frame rate.

Finalmente, acredita-se que tal dispositivo possa ser facilmente fabricado no país e várias indústrias poderão usufruir dessa técnica para melhoria na qualidade final de seus processos e produtos, aumentando desta forma a sua competitividade a nível mundial.

7 Agradecimentos

Este trabalho contou com apoio parcial da FAPEMIG TEC609/96 e CNPq 522618/96.

8 Referências

D. H. Ballard & C. M. Brown. Computer Vision. Prentice Hall, Englewood cliffs, 1982.

Denis Laurendeau, Régis Houde, Marc Samson and Denis Poussart. 3D Range Acquisition Through Differential Light Absorption. IEEE Transactions on Instrumentation and Measurement, vol 4, n° 5. October 1992.

J. K. S. Souza. Um dispositivo automático para aquisição de forma 3D de objetos. Dissertação de Mestrado. UFMG, 1998.

K. L. Boyer and A. C. Kak. Color-Encoded Structured Light for Rapid Active Ranging. Pattern Analyses and Machine Intelligence, vol 9, n° 1, p. January 1987.

R. A. Jarvis. A Perspective on Range Finding Techniques for Computer Vision. Pattern Analysis and Machine Vision, vol 5, n° 2, p. 122-139, march 1983.

Rafael C. Gonzalez, e, Paul Wintz. Digital Image Processing. Addison Wesley, 1987.

Richard O. Duda and Peter E. Hart. Pattern Classification and Scene Analysis. Wiley-Interscience Publication, 1973.

Shree K. Nayar, Masahiro Watanabe e Minori Noguchi. Real Time Focus Range Sensor. IEEE Transactions on Pattern Analysis and Intelligence, Vol. 18, n° 12, 1996.

Stuart C. Shapiro. Encyclopedia of Artificial Intelligence. Vol. I e II. Wiley-Interscience Publication, 1987.

Yoshiaki Shirai. Three Dimensional Computer Vision. Springer-Verlag, 1987.

Z. Jason Geng. Rainbow three-dimensional camera: new concept of high-speed three-dimensional vision systems. Optical Engineering, p.376-383, Feb. 1996. Conference on Robotics and Automation. April 1991.