SISTEMA DE DETERMINAÇÃO DE COORDENADAS CARTESIANAS BASEADO EM VISÃO ESTÉREO

GUILHERME COSTA HOLANDA¹, JÉSSYCA ALMEIDA BESSA¹, RENAN BESSA¹,
FRANCISCO ALAN XAVIER DA MOTA¹, AJALMAR REGO DA ROCHA NETO¹,
AUZUIR RIPARDO ALEXANDRIA¹

¹Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE) - *Campus Fortaleza*
<guilherm3@msn.com>, <bessajessyca@gmail.com>, <renan_bessa@hotmail.com>,
<alan.mota@hotmail.com>, <ajalmar@ifce.edu.br>, <auzuir@gmail.com>

DOI:10.21439/conexoes.v10i2.752

Resumo. Os métodos de determinação de coordenadas cartesianas são muito importantes no mapeamento de objetos em um ambiente. Devido a essa importância, estes métodos geram bastante interesse por parte de pesquisadores da área de Visão Computacional. Assim, neste trabalho será abordado um desses métodos, que é baseado em visão estéreo. Como as câmeras são sensores passivos, os métodos de medição de dimensões por visão estéreo se destacam, pois podem fornecer, além das coordenadas, muitas outras informações, por exemplo, grau de iluminação, cores, identificação e reconhecimento de objetos. Baseado nesta vantagem, o objetivo deste trabalho, portanto, é desenvolver um sistema de determinação de coordenadas cartesianas baseado em visão estéreo. Para esta finalidade, uma câmera Minoru3D foi utilizada. A estrutura do equipamento inclui duas câmeras em uma só. Além disso, um mapa artesanal foi confeccionado para testar o sistema implementado e um objeto de faces retangulares foi utilizado nos testes de determinação de coordenadas. Este objeto de teste foi disposto em algumas posições no mapa e foi feita uma comparação entre as coordenadas do mapa e do sistema estéreo. Os testes do sistema apresentaram resultado satisfatório com erro quadrático médio de 0,58. Sendo assim, o sistema estéreo implementado foi capaz de disponibilizar coordenadas confiáveis do objeto nas posições dispostas.

Palavras-chaves: Visão Computacional, Visão Estéreo, Determinação de Coordenadas Cartesianas.

Abstract. The available methods applied to determine Cartesian coordinates are very important in the mapping of objects in an environment. Due to this importance, these methods generate enough interest from researchers at the Computer Vision area. In this work will be discussed one such method, which is based on stereo vision. In this work will be discussed one such method, which is based on stereo vision. As the cameras are passive sensors, measurement methods for stereo vision stand out, because they can provide, in addition to the coordinates, many other information, such as illumination, colors, object identification and recognition. Based on this advantage, the aim of this study therefore is to develop a determination of Cartesian coordinates based on stereo vision system. This work aims to develop a system to determine the Cartesian coordinates based on stereo vision. For this purpose, a Minoru3D camera was used. The structure of the equipment includes two cameras in one. Furthermore a handmade map was made to test the implemented system and an object of rectangular faces was used in order to determine the coordinate tests. The test object was placed at some positions on the map and a comparison was made between the map coordinates and stereo system. The system tests showed satisfactory results with mean square error of 0.58. Therefore, the implemented stereo system was able to provide reliable object coordinates in the arranged positions.

Keywords: Computer Vision, Stereo Vision, Cartesian coordinates Determination.

1 INTRODUÇÃO

A visão é um dos principais sentidos do ser humano, por meio dela é possível captar cor, formas, sombra e profundidade de objetos, provendo, assim, vários dados para o cérebro, que os interpreta em informações úteis. Em uma analogia a revolução industrial, o homem sempre busca automatizar algumas de suas próprias funções para máquinas, isso também ocorre com a visão computacional, que visa emular as funções que realizamos naturalmente (PITERI; RODRIGUES, 2011).

Com o acesso à grande poder de processamento em pequenos dispositivos, muitos sistemas de captura e processamento de imagens em tempo real vêm surgindo. Porém, o grande entrave é a qualidade das câmeras, principalmente das lentes, que muitas vezes geram distorções, assim a calibração das câmeras e a correção dessa distorção por software, é o fator crucial para um bom resultado quando se busca realizar tarefas que necessitam da disparidade, como no caso detecção de distância ou obtenção de coordenadas cartesianas (MENDES, 2012).

Os métodos de obtenção de coordenadas cartesianas de um objeto e medição de distância podem ser divididos em ativos e passivos. Os métodos ativos medem a distância através do envio de algum sinal ao objeto (por exemplo, feixes de laser, sinais de rádio, ultrassom, entre outros), enquanto os métodos passivos apenas recebem informações da posição do objeto. Dentre os passivos, os mais populares são aqueles baseados em métodos de medição estereoscópica o Visão estéreo (MROVLJE; VRANCIC, 2008).

Visão estéreo é uma técnica utilizada para a construção de uma descrição tridimensional de uma cena observada de um determinado ponto de vista. É considerado um método passivo quando não se é necessário uma iluminação adicional da cena, por exemplo, por feixes de laser. A visão estéreo passiva passa a ser muito atraente para muitas aplicações em robótica, incluindo reconhecimento tridimensional de objetos e localização, bem como navegação tridimensional de robôs móveis (AYACHE; LUSTMAN, 1991).

A maioria das pesquisas sobre visão estéreo passiva tem sido dedicada à visão binocular na qual duas câmeras observam a mesma cena a partir de dois pontos de vista ligeiramente diferentes. Assim que dois pontos (um em cada imagem) são combinados, isto é, identificados como correspondentes a um mesmo ponto físico, é possível calcular as coordenadas tridimensionais deste ponto físico (PORTELA, 2015; AYACHE; LUSTMAN, 1991).

Desta forma, o objetivo deste trabalho é desenvolver um sistema de determinação de coordenadas cartesianas

baseado em visão estéreo.

A determinação de coordenadas cartesianas por métodos passivos é interessante, devido à eliminação de erros causados por problemas de resolução angular ou reflexão difusa. A estrutura desse trabalho consiste inicialmente no referencial teórico, abordando os conceitos de visão estereoscópica, calibração de câmeras, entre outros utilizados no desenvolvimento. Serão comentadas todas as ferramentas de programação e de hardware usados, a metodologia e processos desenvolvidos, assim como apreciação e comparação de resultados obtidos.

2 FUNDAMENTAÇÃO

2.1 VISÃO COMPUTACIONAL

A Visão Computacional é um campo de rápido crescimento, principalmente pelo fato da evolução de *hardware* que tornou as câmeras e processadores mais baratos e mais poderosos, também em parte devido à evolução de *software* que aumentou o poder de processamento disponível e também devido ao aperfeiçoamento dos algoritmos de visão (BRADSKI; KAEHLER, 2008).

Segundo Bradski e Kaehler (2008), a captura é a transformação de dados de uma câmera fotográfica ou de vídeo em uma nova representação. Todas essas transformações são feitas para alcançar objetivos específicos. Essa nova representação pode significar transformar uma imagem colorida em tons de cinza ou remoção de movimento da câmera a partir de uma sequência de imagens.

As ferramentas necessárias por um sistema de Visão Computacional incluem um hardware para aquisição e armazenamento de imagens digitais em um computador utilizado para processar as imagens e para comunicar os resultados aos usuários ou sistemas automatizados (GALO; TOMMASELLI, 2011).

Segundo Trucco e Verri (1998), o primeiro processo de um sistema de Visão Computacional consiste na aquisição de uma imagem, ou um par estéreo, ou uma sequência de um vídeo inteiro. Antes de serem preparadas para algum algoritmo, as imagens são pré-processadas para atenuar a presença de ruído adquirido no processo de aquisição (Ribeiro Filho, 2012).

As características das imagens podem ser extraídas para auxiliar em alguns processos, tais como:

Calibração, que determina o valor dos parâmetros internos e externos do sistema de Visão;

Reconhecimento, que determina a identidade do objeto e posição no espaço;

Análise estéreo, que explora a diferença entre duas

imagens para computar a estrutura (forma) de objetos 3-D, e suas posições no espaço;

Análise de movimento, que explora as mudanças finitas em uma sequência de imagens, ocasionadas pelo movimento relativo entre o mundo e a câmera, para estimar a estrutura 3-D e o movimento; Estimação da forma do objeto 3-D, que é possível a partir de informações contidas em apenas uma imagem (TRUCCO; VERRI, 1998).

Algumas das aplicações da visão computacional são:

- Controle na agricultura (SANTOS; YASSITEPE, 2015);
- Inspeção de embalagens, rótulos, etiquetas e tampas (LIMA; CARVALHO, 2015);
- Reconhecimento de caracteres e leitura de código de barras (MIRANDA et al., 2013);
- Detecção de obstáculos em veículos autônomos (SCHNEIDER, 2013);
- Monitoramento de avicultura (SALTORATTO et al., 2013);
- Contagem de veículos em tráfego urbano (BARCELLOS, 2014);
- Inspeção de placas eletrônicas (MELO, 2013);
- Inspeção de superfícies (KUHN et al., 2011); e
- Visão para robôs e manipuladores (COUTO, 2012).

2.1.1 MODELO DA CÂMERA DE ORIFÍCIO (PINHOLE)

O modelo de câmera mais utilizado é o *pinhole*, ou câmera de orifício. Este modelo é a mais simples representação do funcionamento de uma câmera. Imagina-se a câmera como uma caixa com um furo pequeno no centro de um dos seus lados e com uma placa opaca no seu interior. Se a caixa for posicionada de frente para uma fonte de luz, tal como uma vela, será possível ver uma imagem invertida da mesma aparecendo na placa opaca. Esta imagem é formada pelos raios de luz emitidos pela fonte em frente à caixa. Se o orifício fosse realmente reduzido a um ponto (fisicamente impossível), exatamente um único raio de luz passaria através de cada ponto na placa (plano da imagem) (Sousa Segundo, 2007). Na Figura 2 é possível se visualizar um exemplo de projeção na câmara de orifício.

O modelo de projeção da perspectiva do orifício (ou perspectiva central) é matematicamente conveniente e, apesar de sua simplicidade, geralmente proporciona uma aproximação aceitável do processo de formação da imagem.

A projeção perspectiva cria imagens invertidas, e é conveniente algumas vezes considerar uma imagem virtual associada com um plano que se encontra na frente do orifício, à mesma distância para o plano da imagem. Esta imagem virtual não é invertida, mas é estritamente equivalente à atual. Dependendo do problema, pode ser mais adequado utilizar um ou outro. Na Figura 3 é possível se visualizar a projeção perspectiva de um ponto P considerando a imagem virtual sem inversão (Sousa Segundo, 2007).

Considerando que as coordenadas X , Y e Z são relativas ao ponto P e que as coordenadas x , y e z são relativas a p , que é imagem do ponto P , pode-se obter, por semelhança de triângulos, conforme Figura 4, o seguinte:

$$Z = f \quad (1)$$

$$X = f \frac{X_c}{Z_c} \quad (2)$$

De modo análogo pode-se obter a relação para o eixo y , conforme Equação 3.

$$y = f \frac{Y_c}{Z_c} \quad (3)$$

Para converter as coordenadas (x,y) do ponto p que está localizado no plano da imagem em coordenadas em pixel (u,v) , os seguintes parâmetros das lentes são necessários:

1. Coordenadas em pixel do ponto O , que é o ponto principal localizado no plano da imagem:

$$O = (u_0, v_0) \quad (4)$$

2. Fatores de escala (α_u, α_v) para o tamanho do pixel:

$$u = u_0 + \alpha_u x \quad (5)$$

$$v = v_0 + \alpha_v y \quad (6)$$

O ponto (u_0, v_0) depende em como as lentes estão montadas sobre o sensor. Para uma câmera ideal, ele está localizado exatamente no centro do plano da imagem.

Figura 1: Diagrama de Blocos de um sistema de Visão Computacional (TRUCCO; VERRI, 1998).

Figura 2: Projeção de imagem invertida no interior da caixa de orifício (Fonte: Autor).

Figura 4: Semelhança de triângulos (Fonte: Autor).

Figura 3: Projeção perspectiva considerando imagem virtual não invertida (Fonte: Autor).

Figura 5: Ilustração da conversão do plano da imagem para pixel (Fonte: Autor).

2.1.2 CALIBRAÇÃO DE CÂMERAS

A calibração de câmeras é uma etapa necessária a qualquer aplicação em visão computacional e tem por finalidade a obtenção dos parâmetros que tornam possível o mapeamento de coordenadas tridimensionais do ambiente em coordenadas bidimensionais de imagens. Esses parâmetros são geralmente divididos em duas categorias: parâmetros extrínsecos, que fornecem a posição da origem do sistema de coordenadas da câmera em relação à origem das coordenadas do mundo, e os intrínsecos, que exprimem a geometria interna da câmera bem como suas características ópticas (Sousa Segundo, 2007).

Ainda segundo o autor, nos métodos de calibração convencional, um ponto em comum é a modelagem por meio de equações que relacionam pontos conhecidos em coordenadas de mundo com suas respectivas representações em imagem. Essas equações compõem um sistema a partir do qual são encontrados os parâmetros extrínsecos e intrínsecos. É necessária a observação de uma ou mais imagens de objetos cuja geometria no espaço tridimensional é conhecida com alta precisão, como, por exemplo, um tabuleiro de xadrez.

Os parâmetros extrínsecos representam a posição e orientação da origem do referencial da imagem em relação ao referencial de um eixo do mundo real. Ou seja, definem a posição e orientação 3D da câmera em relação a um determinado sistema de coordenadas. Estes parâmetros fazem a transformação entre o sistema de coordenadas do mundo para o sistema de coordenadas da câmera e são, portanto, translações e rotações (Sousa Segundo, 2007).

Para Sousa Segundo (2007), os parâmetros intrínsecos são aqueles que definem a geometria interna da câmera e suas características ópticas. São eles: distância focal da câmera, ponto central, constante de distorção óptica e fator de escala. A distância focal da câmera é a distância entre o ponto nodal interior do sistema de lentes e o plano da imagem. O ponto principal ou centro projetivo é definido como a intersecção do eixo óptico do sistema de lentes da câmera com o plano da imagem. A constante de distorção óptica mede a distorção causada pela imperfeição na forma geométrica das lentes da câmera.

A partir da substituição das Equações 2 e 3 nas Equações 5 e 6 são obtidas as seguintes equações:

$$u = u_0 + \alpha_u f \frac{X_c}{Z_c} \quad (7)$$

$$v = v_0 + \alpha_v f \frac{Y_c}{Z_c} \quad (8)$$

Multiplicando-se as Equações 7 e 8 em ambos os lados por Z_c , obtém-se a seguinte equação, onde s é o fator escala que é igual a Z_c .

$$\begin{bmatrix} s \cdot u \\ s \cdot v \\ s \end{bmatrix} = \begin{bmatrix} \alpha_u f_x & 0 & u_0 & 0 \\ 0 & \alpha_v f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad (9)$$

Logo, a Matriz M de parâmetros intrínsecos é dada por:

$$M = \begin{bmatrix} \alpha_u f_x & 0 & u_0 \\ 0 & \alpha_v f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (10)$$

2.1.3 VISÃO ESTÉREO

A Visão 3D baseia suas operações na aquisição de um par estéreo de imagens de uma cena e em seguida na decodificação da informação de profundidade capturada implicitamente pela disparidade do par estéreo, isto é, o deslocamento relativo do conteúdo de uma das imagens em relação à outra. Este processo é chamado de técnica estéreo-fotogramétrica, ou seja, medição a partir de um par estéreo de imagens (CYGANÉK; SIEBERT, 2011).

Como analogia, pode-se citar a visão humana. Quando um ser humano observa uma cena com os dois olhos, uma imagem da cena é formada na retina de cada olho. Entretanto, como os olhos são dispostos com um deslocamento horizontal entre si, as imagens formadas por eles não são idênticas (CYGANÉK; SIEBERT, 2011). Quando as duas imagens chegam simultaneamente na parte traseira do cérebro, são combinadas em uma única imagem (SERRA, 2012). O cérebro combina as duas imagens associando as similaridades e adicionando as pequenas diferenças. Essas pequenas diferenças entre as duas imagens são adicionadas até que se obtenha uma diferença total na imagem final. A imagem combinada é mais do que a soma de duas imagens, é uma imagem estereoscópica tridimensional (Sousa Segundo, 2007).

A Visão Computacional pode ser utilizada como um sensor de distâncias, substituindo sonares e sensores baseados em infravermelho, que apresentam problemas de resolução angular e reflexão difusa, respectivamente. Uma imagem plana não é suficiente - na maior parte dos casos - para fornecer informações sobre a distância dos objetos à câmera (TRENTINI et al., 2013). Para obter esta informação, são utilizadas, no mínimo, duas imagens, sejam elas adquiridas por uma ou mais câmeras, ou pela movimentação da(s) mesma(s) ou de um robô. Usar duas imagens, conhecendo-se o deslocamento do

observador na tomada de uma para a outra, para obter a distância dos objetos é no que consiste a visão estéreo (Sousa Segundo, 2007).

Para se implementar um programa computacional capaz de simular uma percepção estereoscópica, é preciso entender como vários pontos de vista de uma mesma cena restringem a estrutura tridimensional e as configurações correspondentes das câmeras. No caso de uma visão estéreo binocular, a primeira imagem de qualquer ponto deve estar no plano formado pela segunda imagem e pelo centro óptico das duas câmeras. Essa restrição epipolar é expressa geometricamente por uma matriz 3×3 chamada de matriz essencial, quando os parâmetros intrínsecos das câmeras são conhecidos (Sousa Segundo, 2007).

A visão estéreo envolve basicamente três processos (Sousa Segundo, 2007):

- Extração das características das imagens observadas pelo par de câmeras;
- A fusão binocular das características observadas pelas duas câmeras, ou seja, estabelecer a correspondência entre características das duas imagens;
- Utilizar a informação do estágio anterior juntamente com os dados geométricos do sistema estéreo para recuperar a posição tridimensional do ponto no mundo.

Um processo de calibração se faz necessário para se conhecer as informações geométricas a fim de se permitir a inversão do processo de formação de imagens para se descobrir as distâncias entre câmeras e objetos da cena (CYGANNEK; SIEBERT, 2011).

Observa-se na Figura 6 um sistema de visão estéreo constituído de duas câmeras de distância focal f equidistantes de uma origem, separadas pela distância b (SIEGWART; NOURBAKHSH, 2004).

Ao visualizar um objeto, cada ponto de seu contorno visível possui uma coordenada (X, Y, Z) no espaço físico em relação a uma origem localizada entre as lentes esquerda e direita. A medição da coordenada z da câmera é dada por (SIEGWART; NOURBAKHSH, 2004):

$$Z = \frac{b \cdot f}{x_l - x_r} \quad (11)$$

onde os termos X_l, X_r, b e f são correspondentes, respectivamente, a coordenada x em pixel na câmera esquerda, coordenada x em pixel na câmera direita, distância entre as duas câmeras no sistema estéreo e distância focal.

Figura 6: Geometria de um sistema de visão estéreo (SIEGWART; NOURBAKHSH, 2004).

Geometria Epipolar

O processo de localização de pontos correspondentes nas imagens de cada câmera requer bastante processamento computacional. Por este motivo é interessante haver restrições na busca de tais correspondentes para acelerar o processo, destacando-se o método baseado na geometria da linha epipolar.

Considerem-se as imagens P_l e P_r de um ponto P observado por duas câmeras com centros ópticos C_l e C_r . O ponto P_r está sobre a linha epipolar onde esse plano e a retina π_r da segunda câmera se intersectam. Esta linha epipolar é a linha associada ao ponto P_l , e passa através do ponto e_r , onde a linha base entre os centros ópticos C_l e C_r intersectam π_r . Simetricamente, o ponto P_l está sobre a linha epipolar que passa através da intersecção e da linha base com o plano π_l (Sousa Segundo, 2007).

Figura 7: A Geometria Epipolar (TRUCCO; VERRI, 1998).

Os pontos e_l e e_r são chamados de epipolos das duas câmeras. O epipolo e_r é a imagem (virtual) do centro óptico C_l da primeira câmera na imagem obser-

vada pela segunda câmera, e vice versa. Se P_l e P_r são imagens do mesmo ponto, então P_r deve estar na linha epipolar associada com P_l . Essa restrição epipolar é fundamental para visão estéreo e análise de movimento.

A restrição epipolar limita a busca por correspondência de pontos nas duas câmeras: As coordenadas do ponto P_l determinam completamente a junção do raio C_l e P_l , consequentemente determinando o plano epipolar associado $C_l C_r P$ e a linha epipolar.

A busca por correspondências pode ser restrita a essa linha em vez das imagens inteiras, como ilustrado na Figura 8. Na análise estéreo, cada câmera deve estar calibrada internamente, mas a transformação rígida separando o sistema de coordenadas das duas câmeras é desconhecida. Nesse caso, a geometria epipolar restringe o conjunto de possíveis correspondências entre as duas imagens.

Figura 8: Restrição epipolar (Sousa Segundo, 2007).

2.1.4 Retificação

Com o objetivo de reduzir a complexidade do problema de determinação de correspondência de pontos na imagem das duas câmeras, é normalmente efetuada uma transformação das imagens, de forma que se possa garantir que pontos correspondentes estejam localizados sempre na mesma linha. Esta transformação, que permite que uma busca a princípio bidimensional de pontos correspondentes seja reduzida a uma busca unidimensional, é denominada retificação (RODRIGUES; MENCARI; BAUCHSPIESS, 1999).

Para uma visão geral do processo, considera-se inicialmente a Figura 9 que exemplifica a formação de imagens de um objeto pontual por duas câmeras em posições arbitrárias. As câmeras são caracterizadas pelos centros ópticos C_l e C_r e pelos planos de retina π_l e π_r . Já na Figura 10 é mostrado o caso em que as duas câmeras apresentam o mesmo plano focal; além disso, a linha que une os pontos principais dos planos de retina é perpendicular aos dois eixos ópticos e paralela a um dos eixos da imagem. Esta é a denominada confi-

Figura 9: Correspondência entre pontos de imagens estéreo geradas por câmeras em configuração arbitrária (RODRIGUES; MENCARI; BAUCHSPIESS, 1999).

guração canônica. Neste caso, embora P_r não possa ser determinado somente a partir de P_l , sabe-se de antemão que ele se encontra ao longo de uma linha paralela a um dos eixos da imagem (no exemplo, o eixo horizontal). Esta linha destacada em azul no plano π_r na Figura 10 é formada pelos possíveis correspondentes do ponto P_l , é a denominada linha epipolar (RODRIGUES; MENCARI; BAUCHSPIESS, 1999).

Figura 10: Correspondência entre pontos de imagens estéreo geradas por câmeras em configuração canônica (RODRIGUES; MENCARI; BAUCHSPIESS, 1999).

No caso da Figura 10, observe-se que a linha epipolar de P_l não só é horizontal como também passa por P_l . Em outras palavras, os pontos P_l e P_r têm a mesma

coordenada vertical. Nesta situação, reduz-se o conjunto de pontos que possivelmente são homólogos de P_i àqueles que na imagem da direita situam-se na mesma linha horizontal que esse ponto. Durante a retificação, o que se deseja é calcular a partir das imagens fornecidas por um sistema em configuração arbitrária aquelas que seriam fornecidas por câmeras na configuração canônica, de forma a garantir que pontos correspondentes em imagens distintas passem a se situar na mesma linha. Este processo se baseia no fato de que, dispondo-se da imagem fornecida por uma câmera, pode-se calcular a imagem da mesma cena que seria fornecida por uma outra câmera qualquer, desde que mantido o centro focal da configuração original (RODRIGUES; MENCARI; BAUCHSPIESS, 1999).

2.2 MÉTODO DE CORRESPONDÊNCIA DE PONTOS (TEMPLATE MATCHING)

A correspondência de pontos através do método de *Template Matching* é baseada em uma função que busca em uma determinada área de busca em uma imagem, qual o quadro que possui maior semelhança com o quadro que se deseja encontrar na imagem. Um critério de análise de correspondência é utilizado para definir o grau de semelhança entre o quadro que se deseja encontrar (*template*) e a imagem de busca (BRADSKI; KAEHLER, 2008). Na Figura 11 é possível se visualizar um exemplo de busca utilizando *Template Matching*.

A *OpenCV* oferece os seguintes critérios de análise de correspondência para o método de *Template Matching* (BRADSKI; KAEHLER, 2008):

- Correspondência por diferença quadrática (CV_TM_SQDIFF): Uma perfeita correspondência resulta no valor 0 e uma péssima correspondência resulta em um valor alto, dependendo da semelhança dos quadros.
- Correspondência por Correlação (CV_TM_CCORR): Uma perfeita correspondência resulta em valor alto e uma péssima correspondência resulta em um valor baixo ou zero.
- Correspondência por coeficiente de Correlação (CV_TM_CCOEFF): Uma perfeita correspondência resulta em 1 e uma péssima correspondência resulta em -1 . O valor zero simplesmente significa que não há nenhuma correlação.
- Métodos Normalizados: Para cada um dos três métodos acima descritos, há também uma versão normalizada. Os métodos normalizados são bastante

usuais porque podem auxiliar na redução de efeitos de diferenças de iluminação entre o *template* e a imagem

Figura 11: Exemplo de aplicação do *Template Matching* (BRADSKI; KAEHLER, 2008).

3 METODOLOGIA

A justificativa para a produção deste trabalho está na necessidade, para a área de Visão Computacional, de se localizar objetos em um ambiente. Para a realização deste mapeamento existem dois tipos de métodos, classificados, pois, em ativos e em passivos. No caso desta obra, é abordado um método passivo baseado em visão estéreo, que é uma técnica utilizada para a construção de uma descrição tridimensional de uma cena observada sob um determinado ponto de vista. Assim, como as câmeras são sensores passivos, os métodos de medição de dimensões por visão estéreo se destacam, já que podem fornecer, além das coordenadas, muitas outras informações, por exemplo, grau de iluminação, cores, identificação e reconhecimento de objetos.

Para que seja alcançado o objetivo deste trabalho, que consiste em desenvolver um sistema de determinação de coordenadas cartesianas baseado em visão estéreo, foi utilizada uma câmera Minoru3D. A estrutura do equipamento inclui duas câmeras em uma só. Por causa disso a Minoru3D é uma boa solução para ser utilizada como uma ferramenta poderosa na área de visão computacional.

Além da câmera, para compor o sistema proposto foi confeccionado um mapa artesanal com a finalidade de testar o sistema implementado. Os testes consistem em determinar as coordenadas de um objeto de testes, que possui as faces retangulares.

Inicialmente as câmeras são calibradas utilizando uma *toolbox* para MATLAB, fornecendo assim os parâmetros das câmeras. Em seguida, utilizando a biblioteca OpenCV, são adquiridas imagens das duas câmeras, as quais são retificadas, buscando-se pontos correspondentes entre as duas imagens para então se determinar as coordenadas cartesianas de um determinado ponto no mundo.

O objeto de teste foi disposto em diferentes posições no mapa e para cada posição foi feita uma comparação entre as coordenadas do mapa e do sistema estéreo. Estes dados estão apresentados na Tabela 1.

4 MATERIAIS E MÉTODOS

4.1 EQUIPAMENTOS UTILIZADOS

4.1.1 Câmera 3D do tipo Webcam

A câmera escolhida para o desenvolvimento do projeto foi uma Minoru3D. O equipamento possui duas câmeras montadas em uma só estrutura, facilitando assim a montagem do sistema estéreo. A seguir são apresentadas as especificações técnicas da câmera (MINORU3D):

- Sensor VGA CMOS
- Resolução de 320x240 a 800x600
- Distância de 6cm entre as duas câmeras
- Taxa de aquisição: 30 fps
- Fácil conexão por meio de um cabo USB

Na Figura 12 é apresentada a Minoru3D.

Figura 12: Minoru3D Webcam (MINORU3D).

Um mapa de dimensão 1,00 m x 1,25 m foi confeccionado utilizando cartolinas e marcações com resolução de 2 mm para o teste do sistema de visão implementado. Na Figura 13 é apresentado o mapa, com a câmera em uma posição fixa.

Figura 13: Mapa Artesanal desenvolvido (Fonte: Autor).

4.1.2 Objeto utilizado nos testes

O objeto da Figura 14 foi utilizado nos testes do sistema. O objeto possui faces de 7 cm x 6,5 cm e o mesmo possui marcação nos vértices de uma das faces para auxílio no processo de obtenção de coordenadas nas imagens das duas câmeras.

Figura 14: Objeto utilizado nos testes (Fonte: Autor).

4.2 BASES PARA O DESENVOLVIMENTO DA PESQUISA

4.2.1 Estudos preliminares

Conforme discutido nas seções anteriores, primeiramente é necessária a realização de uma calibração das câmeras para a obtenção dos parâmetros das mesmas. Em seguida, é preciso que as imagens sejam retificadas, de modo a facilitar a correspondência de pontos em duas imagens através da teoria da geometria epipolar. Teoria essa que permite realizar a busca de um ponto em uma imagem na mesma linha na outra imagem, ou seja, transformando uma tarefa bidimensional em uma unidimensional. Os próximos itens abordam as etapas necessárias para o desenvolvimento e implementação prática do sistema.

4.2.2 Calibração das Câmeras

A calibração das câmeras foi realizada a partir de uma *toolbox* desenvolvida por BOUGHET e que está a disposição para *download*. A *toolbox* disponibiliza interface em MATLAB.

Primeiramente, é necessário que se utilize um padrão para calibração. É oferecido aos usuários no sítio da *toolbox* um tabuleiro de xadrez para impressão, o qual deve ser fotografado pelas câmeras para que em seguida as fotos sejam utilizadas no processo de calibração das câmeras. Na Figura 15 é possível se visualizar o tabuleiro de xadrez utilizado como padrão de calibração. Cada quadrado do padrão possui dimensão 30 mm x 30 mm.

Figura 15: Padrão utilizado no processo de calibração (BOUGHET).

Vinte e quatro fotos de cada câmera foram obtidas para serem utilizadas no processo de calibração (no lado esquerdo as fotos da câmera esquerda e no lado direito as fotos da câmera direita), conforme a Figura 16.

Todo o procedimento solicitado na interface da *toolbox* foi cumprido e ao fim, foram obtidas as matrizes de calibração:

- Matriz de Parâmetros intrínsecos - Câmera Esquerda:

$$M1 = \begin{bmatrix} 863,80 & 0 & 309,15 \\ 0 & 865,77 & 250,27 \\ 0 & 0 & 1 \end{bmatrix} \quad (12)$$

- Matriz de Parâmetros intrínsecos - Câmera Direita:

$$M2 = \begin{bmatrix} 862,39 & 0 & 300,41 \\ 0 & 863,00 & 257,05 \\ 0 & 0 & 1 \end{bmatrix} \quad (13)$$

Para utilização no sistema, foi utilizada a distância focal média das duas câmeras, por se tratarem de câmeras semelhantes:

$$f = \frac{863,80 + 865,77 + 862,39 + 863,00}{4} \quad (14)$$

$$f = 863,74 \text{ pixels.} \quad (15)$$

A distância b é fornecida pelo fabricante da câmera. Assim, $b = 6\text{cm}$.

4.3 ETAPAS DO DESENVOLVIMENTO

Após a obtenção dos parâmetros das câmeras, é necessário primeiramente definir o posicionamento do sistema de visão no ambiente.

Primeiramente deve-se obter uma imagem de cada câmera, realizar a retificação de cada imagem, de modo que o sistema possa ser representado na forma canônica e as Equações 2, 3 e 11 sejam válidas. Em seguida, através de clique do mouse os pontos que se deseja determinar as coordenadas no mundo podem ser selecionados e através do algoritmo de correspondência de pontos (*Template Matching*) o ponto correspondente da outra imagem pode ser obtido e então as coordenadas relativas ao sistema estéreo podem ser determinadas. Por fim as coordenadas destes pontos são possíveis de serem determinadas em relação ao mundo, ou seja, ao mapa utilizado nos testes. A Figura 17 apresenta um diagrama explicativo.

4.3.1 Posicionamento do Sistema

A câmera estéreo foi posicionada na posição (52,0;15,6;7,7) do mapa com rotação de -90° em relação ao eixo x do mapa, conforme Figura 18.

4.3.2 Retificação das imagens

Após adquiridas as imagens de cada câmera, as mesmas foram retificadas para posterior escolha de pontos para cálculo de coordenadas cartesianas. Nas Figuras 19 e 20 são apresentadas imagens antes e depois da retificação.

4.3.3 Escolha e correspondência de pontos

O código implementado em C++ permite a predefinição de quantos pontos se deseja selecionar para determinação de coordenadas cartesianas. Os pontos são então obtidos através de cliques com o mouse na imagem da câmera esquerda. Na Figura 21(a) é exibida uma imagem da câmera esquerda com quatro pontos definidos através do clique do *mouse*.

Após selecionados os pontos na câmera esquerda, o método utilizado para correspondência dos pontos com a câmera direita foi o *Template Matching*. O critério

Figura 16: Fotos obtidas para serem utilizadas no processo de Calibração (Fonte: Autor). (a) Câmera esquerda (b) Câmera direita

Figura 17: Diagrama explicativo do sistema implementado (Fonte: Autor).

Figura 18: Posicionamento da Câmera no mapa (Fonte: Autor).

utilizado para análise de correspondência foi o da diferença quadrática, o qual apresentou bom resultado, conforme Figura 21. Os demais critérios não foram testados neste trabalho.

4.3.4 Determinação de coordenadas em relação ao sistema estéreo

Considerando que um ponto foi obtido nas imagens das duas câmeras: (x_l, y_l) da câmera esquerda e (x_r, y_r) da câmera direita. Por meio da Equação 16 é possível se obter a coordenada do ponto no eixo Z do sistema estéreo:

$$Z_c = \frac{b \cdot f}{x_l - x_r} \quad (16)$$

Segundo a Equação 12, o ponto O da câmera esquerda tem coordenadas $(309, 15; 250, 27)$, então a partir das Equações 2 e 3 e considerando o deslocamento do eixo da câmera esquerda de $b/2$ em relação ao eixo x do sistema de coordenadas do sistema estéreo, tem-se:

$$X_c = \frac{(x_l - 309, 15) \cdot Z_c}{f} - \frac{b}{2} \quad (17)$$

$$Y_c = \frac{(y_l - 250, 27) \cdot Z_c}{f} \quad (18)$$

4.3.5 Determinação de coordenadas em relação ao mapa (mundo)

Como o eixo de coordenadas do sistema estéreo está na posição $(52, 0; 15, 6; 7, 7)$ e com rotação de -90° em relação ao eixo x do mapa, temos:

$$X_w = X_c + 52, 0 \quad (19)$$

Figura 19: Retificação Câmera esquerda (Fonte: Autor). (a) Sem retificação. (b) Com retificação.

Figura 20: Retificação Câmera Direita (Fonte: Autor). (a) Sem retificação. (b) Com retificação.

Figura 21: Correspondência de pontos nas duas imagens (Fonte: Autor). (a) Câmera esquerda. (b) Câmera direita.

$$Y_w = Z_c + 15,6 \quad (20)$$

$$Z_w = -Y_c + 7,7 \quad (21)$$

Na Figura 22 é mostrado o mapa, objeto utilizado para testes e a câmera posicionada em uma posição fixa.

Figura 22: Sistema Estéreo (Fonte: Autor).

5 RESULTADOS: APRESENTAÇÃO E ANÁLISE

Após a câmera ser posicionada conforme descrição no capítulo anterior, o objeto utilizado no teste foi disposto em seis diferentes posições e as coordenadas das quatro (4) arestas de uma das faces do objeto foram determinadas e comparadas com as coordenadas obtidas através do mapa, conforme Tabela 1.

Na Figura 23 são apresentadas as localizações reais do objeto nas diferentes disposições obtidas através das coordenadas do mapa (linha tracejada) e também são apresentadas as posições obtidas a partir do sistema de visão estéreo (linha contínua). O Eixo W representa as coordenadas do mundo e o eixo de coordenadas C representa o eixo do sistema estéreo.

O erro quadrático médio apresentou valor de 0,58. Os testes apresentaram um bom resultado. Pode-se observar que os pontos obtidos pelo sistema de visão estéreo se apresentaram bem próximos aos obtidos no mapa. Porém testes com equipamentos de medição mais precisos a fim de comparar com o sistema desenvolvido se fazem necessários, uma vez que um mapa artesanal não é o modo ideal para se testar o sistema. Devem-se realizar testes utilizando mapas ou equipamentos que forneçam melhor precisão que o mapa confeccionado. Mesmo com a limitação do mapa, o sistema foi capaz de apresentar posições confiáveis para o objeto testado, mostrando o quão eficiente é o sistema de Visão Computacional. Além do mais, o sistema pode fornecer diversas outras informações, como cores, reconhecimento de objetos, formas, entre outras.

6 CONCLUSÕES

O problema da determinação de coordenadas cartesianas utilizando técnicas de visão computacional é algo que requer análises mais profundas de precisão e de custo computacional. O problema da correspondência de pontos em duas imagens é algo que ainda é bastante estudado na busca de técnicas que não exijam alto poder computacional.

Quanto aos objetivos, no que diz respeito aos procedimentos adotados, a etapa que mais demandou tempo foi a etapa preliminar, que englobou o processo de calibração. Etapa esta responsável por fornecer os parâmetros das câmeras e que por esse motivo deve ser encarada com bastante atenção para obtenção de valores mais próximos dos reais.

Os testes do sistema implementado forneceram um bom resultado (erro quadrático médio de 0,58). Porém é sugerida como trabalho futuro a utilização de mapas ou equipamentos mais precisos para obtenção das coordenadas para comparação com o sistema desenvolvido. Além disso, é necessário o teste em ambientes mais amplos para análise do alcance na determinação de coordenadas de pontos/objetos no ambiente.

Mesmo com bons resultados, este trabalho teve algumas limitações, como o mapa artesanal que foi confeccionado e a não aquisição e correspondência automática de pontos para determinação de coordenadas.

No geral, este trabalho é importante para mostrar a aplicabilidade das técnicas de visão computacional na determinação da posição de objetos em uma determinada cena. Este sistema pode ser aplicado na área da Robótica, onde não é adequado o uso de sensores ativos de medição de distância para determinação de posição de objetos.

Seguem algumas sugestões para trabalhos futuros:

- Confeção de um mapa que proporcione maior precisão das leituras de coordenadas ou utilizar equipamentos capazes de disponibilizar coordenadas tridimensionais com boa precisão para teste do sistema implementado;
- Aplicação do sistema desenvolvido em ambientes maiores, para testar o alcance e precisão;
- Utilização do algoritmo para aquisição e correspondência automática de pontos desejáveis nas imagens;
- Utilizar outros métodos de *Template Matching*;
- Construção de mapas de profundidade;
- Análise de custo computacional ao se executar determinação de coordenadas de vários pontos.

Tabela 1: Coordenadas obtidas a partir do Sistema de Visão Estéreo.

Posição	Ponto	Localização Real no Mapa (cm)			Localização a partir do Sistema de Visão Estéreo (cm)			Erro Quadrático
		x	y	z	x	y	z	
1	1	45,00	90,00	6,50	44,94	90,12	6,41	0,16
	2	52,00	90,00	6,50	51,72	90,10	6,41	0,31
	3	45,00	90,00	0,00	45,00	90,00	0,13	0,13
	4	52,00	90,00	0,00	51,64	90,00	-0,03	0,36
2	1	38,00	72,00	6,50	38,13	72,19	6,11	0,45
	2	45,00	72,00	6,50	44,81	72,50	6,15	0,64
	3	38,00	72,00	0,00	38,21	71,39	-0,23	0,68
	4	45,00	72,00	0,00	44,96	71,98	-0,45	0,45
3	1	60,00	76,00	6,50	60,59	76,46	6,80	0,81
	2	66,00	72,00	6,50	66,02	71,39	5,71	1,00
	3	60,00	76,00	0,00	60,25	75,38	-0,58	0,89
	4	66,00	72,00	0,00	66,02	71,39	-0,69	0,92
4	1	34,00	105,00	6,50	33,82	105,30	6,30	0,40
	2	40,40	108,00	6,50	40,10	107,70	6,64	0,45
	3	34,00	105,00	0,00	33,72	105,40	0,46	0,67
	4	40,40	108,00	0,00	39,83	108,40	0,28	0,77
5	1	49,00	115,00	6,50	49,11	115,10	6,67	0,25
	2	56,00	115,00	6,50	55,67	115,20	6,44	0,41
	3	49,00	115,00	0,00	48,87	115,40	0,23	0,48
	4	56,00	115,00	0,00	55,55	115,70	0,11	0,82
6	1	77,00	117,00	6,50	76,46	116,70	6,40	0,63
	2	84,00	117,00	6,50	84,79	117,40	6,35	0,90
	3	77,00	117,00	0,00	76,64	116,70	0,00	0,48
	4	84,00	117,00	0,00	83,98	117,80	-0,50	0,94
ERRO QUADRÁTICO MÉDIO								0,58

Figura 23: Representação gráfica da localização dos pontos no mundo (Fonte: Autor).

REFERÊNCIAS

- AYACHE, N.; LUSTMAN, F. Trinocular stereo vision for robotics. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, IEEE, v. 13, n. 1, p. 73–85, 1991.
- BARCELLOS, P. R. M. Detecção e contagem de veículos em vídeos de tráfego urbano. 2014.
- BOUGUET, J. Y. *Camera calibration toolbox for Matlab*. 2013. Disponível em: <http://www.vision.caltech.edu/bouguetj/calib_doc> Acesso em: 08 jul. 2013.
- BRADSKI, G.; KAEHLER, A. *Learning OpenCV: Computer vision with the OpenCV library*. [S.l.]: O'Reilly Media, Inc., 2008.
- COUTO, L. N. *Sistema para localização robótica de veículos autônomos baseado em visão computacional por pontos de referência*. Tese (Doutorado) — Universidade de São Paulo, 2012.
- CYGANEK, B.; SIEBERT, J. P. *An introduction to 3D computer vision techniques and algorithms*. [S.l.]: John Wiley & Sons, 2011.
- GALO, M.; TOMMASELLI, A. Calibração de câmaras. *Visão Computacional*, p. 53–112, 2011.
- KUHN, V. N. d.; JUNIOR, P. L. J. D.; GOMES, S. C. P.; CUNHA, M. A. B.; BOTELHO, S. S. d. C. Controle automático de um rov para inspeção de estruturas submersas. 2011.
- LAZZARI, F. d. S.; OLIVEIRA, H. R. d.; PAIXÃO, T. S. B. d. Q. Desenvolvimento de um robô paralelo tipo delta associado com visão computacional para aplicações pick and place. Curitiba, 2013.
- LIMA, R. P. S.; CARVALHO, C. S. S. Implantação de um sistema de visão computacional aplicada ao controle de qualidade de agulhas em uma indústria farmacêutica. Curitiba, 2015.
- MEDEIROS, M. D. d. *Correspondência estéreo usando imagens em multiresolução com profundidade variável*. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Norte, 2006.
- MELO, D. F. F. de. *Desenvolvimento de máquina automática para inspeção óptica de placas de circuito impresso em pequenas séries*. Tese (Doutorado) — Universidade Federal de Santa Catarina, 2013.
- MENDES, C. C. T. *Navegação de robôs móveis utilizando visão estéreo*. Tese (Doutorado) — Universidade de São Paulo, 2012.
- MINORU3D. *MINORU 3D WEBCAM*. 2014. Disponível em: <<http://www.minoru3d.com>>.
- MIRANDA, R. A. R.; SILVA, F. A. d.; PAZOTI, M. A.; ARTERO, A. O.; PITERI, M. A. Algoritmo para o reconhecimento de caracteres manuscritos. In: *Colloquium Exactarum*. [S.l.: s.n.], 2013. v. 5, n. 2, p. 109–127.
- MROVLJE, J.; VRANCIC, D. Distance measuring based on stereoscopic pictures. In: *9th International PhD Workshop on Systems and Control, Young Generation Viewpoint*. Izola, Slovenia. [S.l.: s.n.], 2008.
- PITERI, M. A.; RODRIGUES, J. C. Fundamentos de visão computacional. *Presidente Prudente: FCT/UNESP-PP*, 2011.
- PORTELA, T. d. C. Arquitetura de hardware dedicada para o cálculo da distância frontal baseado em segmentação por cores usando visão estéreo aplicada em robótica móvel. 2015.
- Ribeiro Filho, P. Visão computacional: Um novo campo de pesquisa em cognição visual. *Psicologia: Teoria e Pesquisa*, v. 3, n. 2, p. 138–150, 2012.
- RODRIGUES, C. J. M.; MENCARI, M. de A.; BAUCHSPIESS, A. Sistema de visão estéreo para formas polimétricas. 1999.
- SALTORATTO, A. Y. K.; SILVA, F. A. d.; CAMARGO, A. C. A. C.; SILVA, P. C. G. d.; SOUZA, L. F. A. d. Monitoramento de avicultura a partir de técnicas de visão computacional. In: *Colloquium Exactarum*. [S.l.: s.n.], 2013. v. 5, n. 2, p. 47–66.
- SANTOS, T. T.; YASSITEPE, J. E. d. C. T. Fenotipagem de plantas em larga escala: um novo campo de aplicação para a visão computacional na agricultura. *Embrapa Informática Agropecuária-Capítulo em livro técnico-científico (ALICE)*, In: MASSRUHÁ, SMFS; LEITE, MA de A.; LUCHIARI JUNIOR, A.; ROMANI, LAS (Ed.). Tecnologias da informação e comunicação e suas relações com a agricultura. Brasília, DF: Embrapa, 2014. Cap. 5., 2015.
- SCHNEIDER, I. R. Uso de visão estereoscópica para detecção de obstáculos na locomoção de um veículo autônomo. 2013.

SERRA, R. M. S. Sistema de visão stereo ativo aplicado aos robôs do iseporto. Instituto Politécnico do Porto. Instituto Superior de Engenharia do Porto, 2012.

Sousa Segundo, J. S. A. d. *Um método para determinação da profundidade combinando visão estéreo e autocalibração para aplicação em robótica móvel*. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Norte, 2007.

TRENTINI, V. B.; RODRIGUES, V.; MARANA, A. N.; LULIO, L. C.; TRONCO, M. L. Stereo matching of natural scenes on external agricultural environment. In: *IX Workshop de Visão Computacional*. [S.l.: s.n.], 2013.

TRUCCO, E.; VERRI, A. *Introductory techniques for 3-D computer vision*. [S.l.]: Prentice Hall Englewood Cliffs, 1998. v. 201.