

# OBTENDO MAPA DE PROFUNDIDADE DE IMAGENS ESTÉREO

MANOEL VIEIRA COELHO NETO\*

\*SQS 203 Bloco J  
Brasília, DF, Brasil

Email: vieiranetoc@gmail.com

## 1 Objetivos

Este relatório tem como objetivo extrair informação de distância entre os objetos da imagem e a câmera. Para isso foi necessário o uso da biblioteca OpenCV, linguagem C++ e alguns algoritmos explicitados ao longo do texto (como busca por SAD). Para que seja possível a obtenção do mapa, é necessário que as duas imagens estejam retificadas e sejam epipolares (linhas coincidentes), tais imagens são chamadas de imagens estéreo. A imagem resultante do processo é o mapa de profundidade, com ele podemos fazer reconstrução de ambientes em 3D, por exemplo.

## 2 Introdução

A visão humana consegue medir distâncias com precisão graças à livre movimentação dos olhos em direção a um ponto de foco, a distância entre os olhos e o ponto é dada por  $d = \frac{c}{2\text{tg}(\alpha/2)}$ , onde  $c$  é a distância entre os olhos e  $\alpha$  o ângulo entre o ponto e os olhos.

Associado ao foco  $f$ , temos o ponto  $P$  interno à câmera. Sabendo a diferença no eixo  $x$  entre o ponto  $P$  de cada um dos olhos ( $D, E$ ) e o ponto  $P_D$  e  $P_E$  podemos calcular a distância entre a câmera e o ponto do mundo por meio da fórmula:

$$Z = \frac{bf}{(X_L - X_R)}$$

Essa fórmula pode ser obtida através de semelhança de triângulos, a imagem segue no final do texto como anexo

Imagens estéreo são obtidas a partir de duas diferentes angulações da câmera em torno do objeto, tal qual os olhos humanos, ao fazer isso sem mudar o eixo  $y$  da câmera teremos imagens epipolares, ou seja, as linhas da imagem à esquerda são coincidentes com as linhas da imagem à direita. Assim, há um pixel em cada imagem na mesma linha associado a um mesmo ponto da imagem. Aqueles pontos que não tiverem coincidência entre si são, portanto, nosso *background* e não pertencem ao(s) objeto(s) de interesse.

## 3 Metodologia

O algoritmo usado neste trabalho é conhecido como *Window-based matching*, ou correspondência baseada em janela em tradução livre. E para

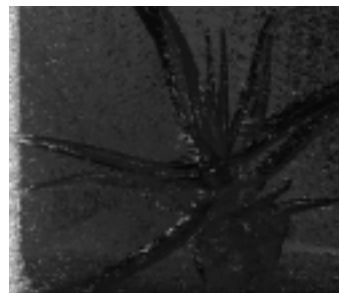
qualquer algoritmo de profundidade estéreo é necessário a retificação das duas imagens de forma que possa se achar os pixels correspondentes nas duas imagens.

Para determinar os pixels coincidentes nas imagens foi feito uso do algoritmo SAD, onde cria-se um *template* de tamanho  $w$  que será uma matriz quadrada a ser comparada com uma região de mesma ordem em torno do pixel da outra imagem, onde cada elemento da matriz de comparação resultante será a diferença absoluta entre os elementos  $(i,j)$ , ao final, somamos todos os elementos da matriz de comparação e chamamos esse processo de SAD (*sum of absolut differences*), aquele pixel com menor sad é dito o pixel correspondente na imagem à direita, iteramos cada pixel de uma linha da imagem da esquerda com todos os pixels da mesma linha na imagem à direita e o processo se repete para todas as linhas das imagens.<sup>1</sup>

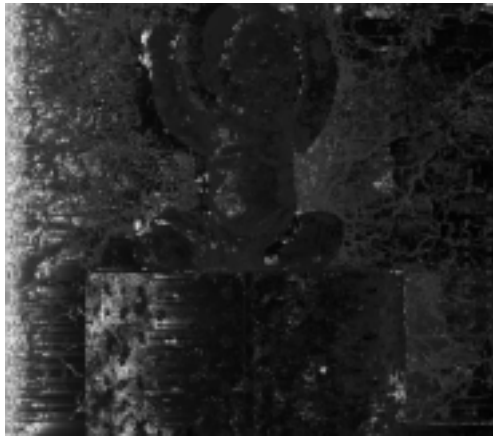
Calcula-se então a distância  $Z$  para cada pixel da imagem usando da fórmula de  $Z$  mostrada acima, para aqueles pontos em que  $(X_L - X_R) = 0$  associamos um número absurdamente grande ao pixel de  $Z$ , para indicar que há correspondência total naquele valor de  $(x,y)$  da imagem  $Z$ . Então nossa imagem normalizada entre 0 e 255 deve ser nosso mapa de profundidade.

## 4 Resultados

Como resultado temos as imagens abaixo, as imagens originais foram upadas no site do imgur pois sua resolução é alta e não cabe nesse documento:



<http://imgur.com/MzQWSYo>



<http://imgur.com/3RuP5I7>

## 5 Conclusão e Discussões

É visível a imprecisão do algoritmo e a quantidade de ruídos presentes na imagem, há outros algoritmos que fazem o mapa de forma otimizada como o *Ground Truth* ou *SGBM* (*Semi-Global Block-Matching*)<sup>[2]</sup>, que talvez tivessem trazido resultados mais satisfatórios, a quantidade de ruído presente na imagem deve-se ao fato que há uma probabilidade de que haja um outro ponto na imagem com o "*sad*" menor e devido a isso o valor daquele pixel seja discrepante em relação aos pontos, métodos de morfologia aplicados sobre a imagem também não trouxeram bons resultados quanto à diminuição do ruído, assim como método de *threshold* binário também não foi favorável ao resultado almejado. Mas, ainda com a alta taxa de ruído, o objeto é distinguível na imagem de profundidade e pode-se dizer que dentro de uma boa margem de aceitação o resultado esperado foi obtido.

## Referências

Hirschmüller, H. (2011). Stereo processing by semi-global matching and mutual information.

*OpenCV Docs* (n.d.).  
<http://docs.opencv.org/2.4/>

*Sum of absolute differences (SAD)* (n.d.).  
[https://en.wikipedia.org/wiki/Sum\\_of\\_absolute\\_differences](https://en.wikipedia.org/wiki/Sum_of_absolute_differences)

<sup>1</sup>(*Sum of absolute differences (SAD)*, n.d.)

<sup>2</sup>(Hirschmüller, 2011) (*OpenCV Docs*, n.d.)