



## SISTEMA DE RECONSTRUÇÃO 3D DE BAIXO CUSTO

Marcelo Archanjo José  
Roseli de Deus Lopes

USP – Universidade de São Paulo, Departamento de Engenharia de Sistemas  
Eletrônicos da Escola Politécnica

marcelo.jose@poli.usp.br, roseli.lopes@poli.usp.br

### RESUMO

O acesso às tecnologias de Reconstrução 3D é afetado pelo custo dos equipamentos necessários para a aquisição e processamento da geometria 3D. Este trabalho apresenta um sistema que utiliza equipamentos convencionais em todo o processo de Reconstrução 3D, promovendo a redução de custos, o que trás como consequência, maior acesso a esta tecnologia bem como a difusão do conhecimento sobre ela.

**Palavras-chave: Reconstrução 3D, Visão Estéreo, Correspondência Estéreo, Visão Computacional.**

### ABSTRACT

The access to 3D Reconstruction technologies is affected by expensive costs of the equipments of acquisition and processing of 3D geometry. This work presents a system which uses conventional equipments in whole 3D Reconstruction phases, this make the costs low, that promote a better access to the technology and the knowledge spread.

**Keywords: 3D Reconstruction, Stereo Vision, Stereo Matching, Computer Vision.**

## 1 Introdução

A Reconstrução 3D é um ramo da Visão Computacional que tem por objeto prover informações tridimensionais de ambiente reais, estas informações podem ser do volume como um todo incluindo seu interior ou somente da superfície. A Reconstrução 3D é aplicada em:

- Medicina - Melhor visualização e diagnóstico
- Arquitetura – Reconstrução 3D de fachada e ambientes
- Cartografia – Mapas 3D

- Robótica
- Automação industrial
- Reconhecimento de padrões
- Navegação

A grande aplicabilidade da Reconstrução 3D a torna uma área bastante pesquisada, importante e promissora. Existem diversos desafios nesta área dentre eles um que se destaca é o custo alto dos aparatos necessários para a pesquisa muitas vezes inacessíveis para pesquisadores independentes e até para algumas universidades. Existem métodos baseados em volume e baseados em superfície. Sendo que para a reconstrução baseada em superfície temos: sistemas de laser scanning, sistemas baseados em imagens.

Atualmente (2007) um Laser Scanner Cyberware Hires Color 3D Scanner 3030/sRGB/MS custa em torno de US\$87.000,00; nos EUA, sem considerar o software e aparato necessário para suportar o scanner.

Sistemas baseados em imagens podem ser assistidos, onde se tem fotos em diferentes ângulos e o usuário do sistema introduz pontos de referência nas bordas e/ou pontos de destaque para orientar a reconstrução 3D. Por outro lado os sistemas baseados em imagens podem ser automáticos como ocorre com Visão Estéreo.

Visão Estéreo sofre diversas dificuldades para encontrar as correspondências entre as imagens de entrada. Para conseguir resultados satisfatórios são impostas restrições, como a restrição epipolar, a qual é amplamente utilizada, pois facilita o trabalho de procura de correspondências. Na prática ela garante que um determinado ponto na imagem da esquerda se encontra na mesma linha na imagem da direita, mudando somente a coluna. Esta restrição exige que as câmeras para Visão Estéreo sejam bastante precisas e estejam rigidamente calibradas, o que geralmente deixa os sistemas de reconstrução 3D baseados em imagens complexos e por conseqüência caros, por exemplo somente uma câmera digital métrica custa atualmente (2007) algo entorno de US\$ 3.000,00, esta câmera tem seus parâmetros internos conhecidos e estáveis. Outra questão importante é o uso intensivo de memória, tanto algoritmos de procura de correspondência tradicionais quanto os mais modernos exigem computadores não convencionais (que são caros) ou trabalham com imagens de baixa resolução.

A estratégia para este trabalho é realizar a reconstrução 3D com:

Hardware: Trabalhar com câmeras digitais e computadores convencionais;

Software: Utilizar Visão Estéreo para obtenção de informações 3D:

- Técnica das Faixas – Técnica que reduz drasticamente o consumo de memória (80% a 95%) durante o processo de procura de correspondências, permitindo o uso de imagens de alta resolução, como por exemplo: 2592x1944;
- Procura 2D - Permite trabalhar sem a restrição epipolar, calibragem básica (duas fotos paralelas, com a mesma câmera), maior flexibilidade para os parâmetros:
  - Extrínsecos - Disposição física: posição e orientação;
  - Intrínsecos - Características ópticas e geometria interna da câmera.

- Opção do uso de Operador Gradiente – Liberdade para a variação de luminosidade no momento de retirar as fotos.

O objetivo destas estratégias é permitir o uso de sistemas mais baratos e simples de implementar, do ponto de vista do usuário.

Utilizando uma abordagem menos rígida, isto é, o uso de câmeras digitais comuns respeitando somente que as imagens estéreo sejam superfícies Lambertianas (superfícies cuja aparência não possui variação significativa com a variação do ponto de vista) temos o barateamento dos custos para a Reconstrução 3D.

## 2 Visão Estéreo

Para descrever o funcionamento de um algoritmo de Correspondência Estéreo tradicional, vamos supor que precisamos realizar a procura de correspondência entre duas imagens com resolução de  $x=320$  e  $y=200$  cada, e que o valor máximo de disparidade ( $d$ ) esperado seja de 10 pixels.

### 2.1 Custo de Correspondência

Reservamos 10 planos com  $320 \times 200$  posições, criando um espaço  $x,y,d$  (veja a figura 1) e realizamos o processo de cálculo de correspondência (soma das diferenças absolutas, soma dos quadrados das diferenças, ou qualquer outro) para cada pixel, para cada disparidade, isto é, comparamos o pixel  $(60,50)$  da imagem da esquerda com o pixel  $(60,50)$  da imagem da direita e armazenamos seu resultado na posição  $(60,50)$  do plano 0 (equivalente à disparidade 0), em seguida comparamos o pixel  $(60,50)$  da imagem da esquerda com o pixel  $(59,50)$  da imagem da direita e armazenamos seu resultado na posição  $(60,50)$  do plano 1 (equivalente a disparidade 1), e assim por diante. Este processo é conhecido como “Matching Cost” [1][2] aqui traduzido como Custo de Correspondência.

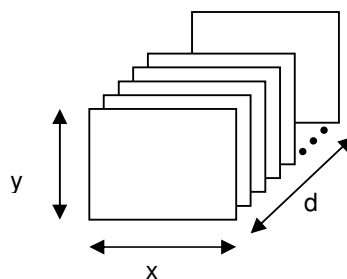


Figura 1: Planos que armazenam os Custos de Correspondência.

### 2.2 Agregação de Evidências

Em seguida são alocados mais 10 planos com  $320 \times 200$  posições onde serão registrados os resultados das somas das janelas. Trata-se de somar (considerando uma janela quadrada  $3 \times 3$ ) os nove resultados de cada plano de Custo de Correspondência na posição central da janela neste novo plano, se por exemplo estivermos no plano 3 de Custo de Correspondência e

estamos na posição (60,70) iremos somar o resultados que estão nas posições descritas na figura 2.

(59,69)	(60,69)	(61,69)
(59,70)	(60,70)	(61,70)
(59,71)	(60,71)	(61,71)

Figura 2:Posições de uma janela 3x3 centrada em (60,70).

É armazenado o resultado desta soma na posição (60,70) do plano 3 recém alocado. Este processo é conhecido como Agregação de Evidências [1][2]. Ao invés de alocar mais 10 planos pode-se alocar somente 1 plano que armazena temporariamente o resultado da agregação, quanto o processo termina de preencher o plano inteiro ele sobrepõe o plano de Custo de Correspondência original.

### 2.3 Seleção de Disparidades

Por último é criado mais um plano com 320x200 nele é armazenado em cada posição o valor da disparidade que obteve a melhor correspondência, isto é, o menor valor. Suponhamos que estamos analisando o ponto (40,50) e que nos planos de Agregação de Evidências temos os resultados da figura 3 para este ponto:

Plano	0	1	2	3	4	5	6	7	8	9
Valor	101	232	184	609	900	143	7	240	78	93

Figura 3:Simulação de valores nos diversos planos para a posição (40,50).

O menor resultado encontra-se no plano 6, o que significa que a melhor correspondência foi encontrada com disparidade de 6 pixels. Neste plano recém criado armazenamos na posição (40,50) o valor 6. Este processo é realizado para os 320x200 pontos e é conhecido como Seleção de Disparidades [1][2]. Deste plano é criada uma imagem em escalas de cinza onde, para o nosso exemplo de disparidade máxima de 10 pixels o 0 seria (0,0,0) (preto em RGB) e o 9 seria (255,255,255) (branco em RGB) e os demais valores os níveis intermediários distribuídos uniformemente. Escolher o menor como melhor correspondência é conhecido como "winner-take-all" (WTA) ou aqui traduzido como "vencedor-leva-tudo".

## 3 Técnica das Faixas

A forma tradicional de procura de correspondências tem sido amplamente pesquisada, mas duas questões não costumam ser abordadas: consumo de memória, não usar a restrição epipolar.

Isto ocorre porque a maioria das pesquisas procura à melhoria da qualidade ou a melhoria do desempenho. Nos laboratórios o uso da restrição epipolar somente facilita o trabalho técnico apesar de exigir maior precisão física. Por outro lado o consumo de memória também não é problema uma vez que se trabalha com imagens de baixa resolução ou com equipamentos especiais com memória suficiente para a pesquisa realizada. O fato é que a questão do consumo de memória não está resolvida. Na técnica tradicional o maior consumo de memória vem da alocação dos planos de Custo de Correspondência, pois trata-se de um volume  $x,y,d$ . A figura 4 apresenta uma simulação do consumo de memória para diferentes resoluções das imagens de entrada.

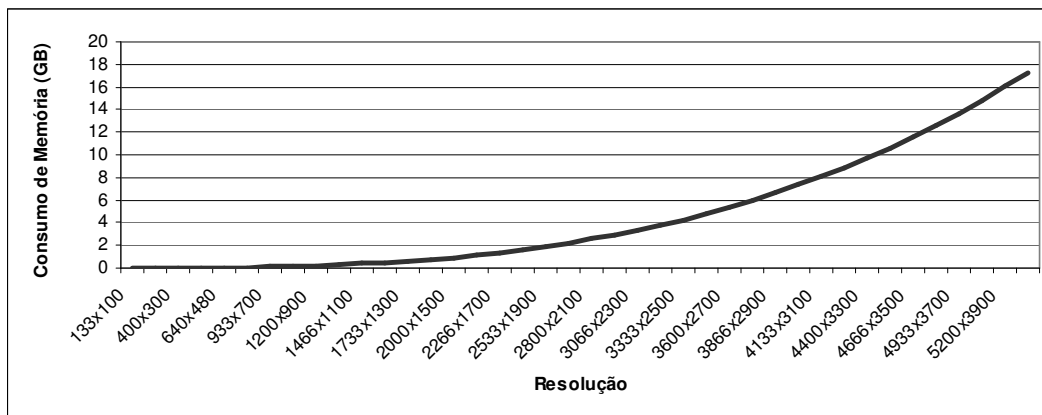


Figura 4: Consumo de memória para técnica tradicional versus resolução, para busca 1D.

Com a Técnica das Faixas as imagens de entrada já carregadas são tratadas em faixas e o processo de procura de correspondências ocorre somente na faixa, estas devem ter o menor tamanho possível sem deixar de manter os resultados intermediários para reuso durante o processo.

Foi estudado qual seria a menor altura desta faixa chegando à conclusão de que ela deve ser igual à altura da janela de procura mais 1, pois durante o processo de agregação é necessário ter os resultados de Custo de Correspondência disponíveis, sendo que uma altura menor não permitiria isso. Também outra conclusão é que a faixa deve ter a mesma largura da imagem de entrada, pois larguras menores que esta eliminaria dados para reuso durante o deslocamento lateral da janela.

O processo de Custo de Correspondência somente trata as faixas das imagens de entrada, preenchendo as *Faixas de Agregação* com a altura da janela (uma linha abaixo ficará livre na primeira rodada). Estas são trabalhadas no processo de Agregação de Evidências produzindo as *Linhas de Seleção de Disparidade*, que são tratadas no processo de Seleção de Disparidade produzindo uma linha no mapa de disparidades, veja a figura 5.

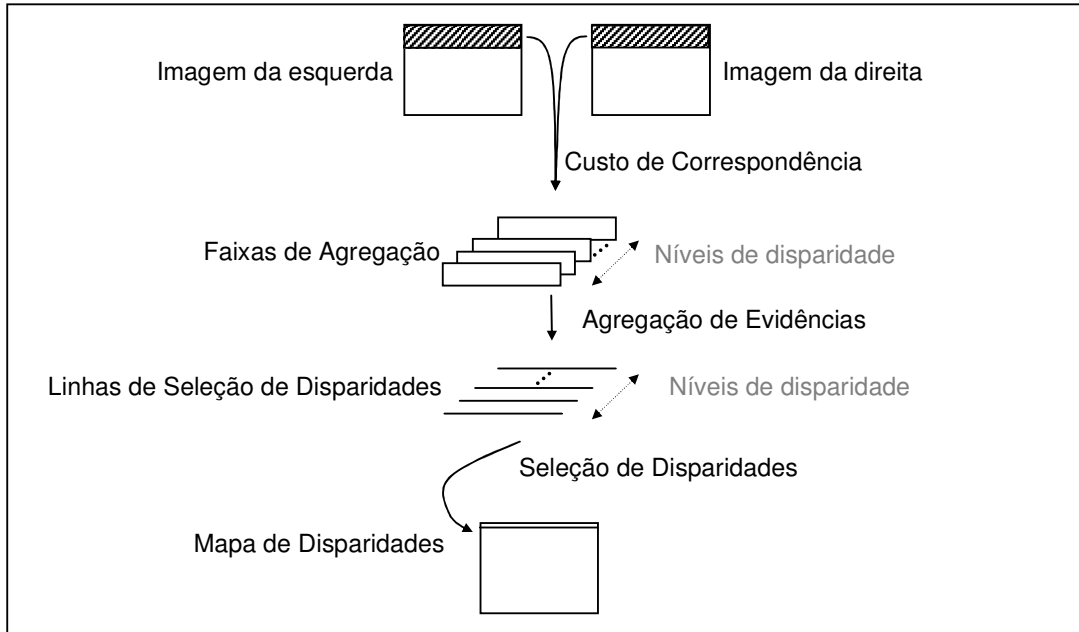


Figura 5: Diagrama da Técnica das Faixas.

Depois da primeira rodada é realizado o cálculo do Custo de Correspondência somente em uma linha (imediatamente abaixo) das imagens de entrada, o resultado é armazenado naquela linha que sobrou nas *Faixas de Agregação*. O processo de Agregação de Evidências agora ignora a primeira linha das *Faixas de Agregação*, produzindo as *Linhas de Seleção de Disparidades* e seguindo normalmente o processo de Seleção de Disparidades. Depois da segunda rodada são eliminados os dados da primeira linha das *Faixas de Agregação* e no lugar são colocados os dados de Custo de Correspondência da linha imediatamente inferior à faixa nas imagens de entrada, e assim por diante até o fim da imagem. Este processo reaproveita os dados que já haviam sido calculados nas rodadas anteriores, somente remove os desnecessários e adiciona os novos.

#### 4 Procura 2D

A forma tradicional de trabalho com Visão Estéreo vinculada à restrição epipolar está bem estabelecida e apresenta resultados muito bons no âmbito de desempenho na construção do mapa de disparidades. Como já descrito, a definição do uso da restrição epipolar exige a construção de aparatos fotográficos com precisão tanto em sua disposição física, nos ajustes ópticos e também nas características internas da câmera, o que faz com que a pesquisa sobre Visão Estéreo ou Reconstrução 3D se torne cara e em alguns caso até inviável. Não utilizar a restrição epipolar provoca um aumento da complexidade e perda de desempenho do algoritmo de busca de correspondências. Por outro lado reduz os custos dos equipamentos e facilita a disposição dos aparatos fotográficos na obtenção do par de imagens, devido a não exigência de precisão e ao fato de permitir trabalhar somente com uma câmera, o que preserva os

mesmos ajustes ópticos em ambas as fotos.

Como alternativa a não utilização da restrição epipolar (que permite uma busca 1D) será necessária uma procura mais sofisticada que exigirá maior tempo de execução. É realizada uma busca 2D numa região de maior probabilidade de encontrar a correspondência, uma vez selecionado o ponto na imagem da esquerda é definida uma janela de procura na imagem da direita.

Nesta procura o volume de dados a serem tratados aumenta drasticamente. Se para um valor de disparidade máxima de 10 pixels for feita a procura por correspondência em cinco linhas iremos varrer 50 pixels que é 5 vezes mais que no processo normal, não esquecendo que para cada pixel existe um conjunto de cálculos que envolvem sua vizinhança.

A abordagem de procura 2D usada neste trabalho não deixa de considerar que a disparidade é horizontal, o fato de procurar a correspondência em outras linhas somente considera que houve três possíveis efeitos, causados pela falta de rigidez na calibragem física da(s) câmera(s) durante o processo de obtenção das imagens:

- Deslocamento de linha – Uma possível linha horizontal que na imagem da esquerda estaria na linha 12 na imagem da direita é encontrada na linha 15;
- Rotação da câmera – Uma possível linha horizontal que na imagem da esquerda estaria na linha 12 na imagem da direita começa na linha 10 passa pela 11, 12 e termina na linha 14;
- Distorções na lente – Uma possível linha horizontal que na imagem da esquerda estaria na linha 12 na imagem da direita começa na linha 14 sobe para linha 13 vai até a 12 e depois volta para 13 e por fim na 14.

A procura 2D não resolve a distorção causada por fotos tiradas por câmeras não paralelas.

## 5 Operador Gradiente

Algoritmos de Visão Estéreo geralmente são muito sensíveis a variações na luminosidade, pois utilizam a intensidade das cores para realizar a procura de correspondências, pequenas variações entre as duas fotos podem apresentar resultados bastante distintos no mapa de disparidades.

Para realizar o processamento é utilizado o operador Sobel [3], através da passagem das máscaras da figura 6, sobre as imagens de entrada:

1	0	-1
2	0	-2
1	0	-1

1	2	1
0	0	0
-1	-2	-1

Figura 6: Máscaras utilizadas para implementação do operador gradiente.

## 6 Reconstrução 3D

Uma vez construído o mapa de disparidades é realizada a criação de uma malha 3D aberta,

supondo uma malha uniforme com 64 vértices em X, 64 vértices em Y e um mapa de disparidades com 320x200, logo há uma relação entre 64 vértices distribuídos uniformemente pelos 320 pixels horizontais e outros 64 para os 200 pixels verticais. Onde o Z da malha será definido pelo valor máximo da disparidade menos a disparidade do pixel do mapa de disparidades mais próximo do vértice em questão, lembrando que a disparidade é inversamente proporcional à profundidade, veja os resultados deste processo na figura 7.

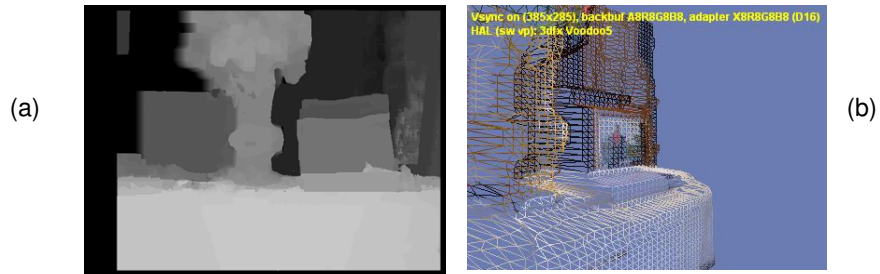


Figura 7:(a) mapa de Disparidades retificado e (b) geometria 3D Reconstruída.

## 7 Resultados

Os resultados obtidos com a Técnica das Faixas, com a Procura 2D e com o Operador Gradiente são apresentados, nesta seção.

### 7.1 Resultados Técnica das Faixas

Usando a abordagem tradicional (com restrição epipolar) e imagens de 2592 x 1944 são necessários mais de 2GB de memória, para imagens maiores como: 5200 x 3900 o consumo de memória sobe para 16GB o que realmente inviabiliza uma ferramenta de baixo custo. A técnica das faixas apresenta uma redução drástica no consumo de memória algo entre 80% a 95% em relação à abordagem tradicional, por exemplo, para a mesma imagem de 2592 x 1944 o consumo de memória é de apenas 124MB e para a imagem de 5200 x 3900 é de 723MB. O gráfico da figura 8 apresenta o consumo de memória da técnica das faixas frente a diversas resoluções.

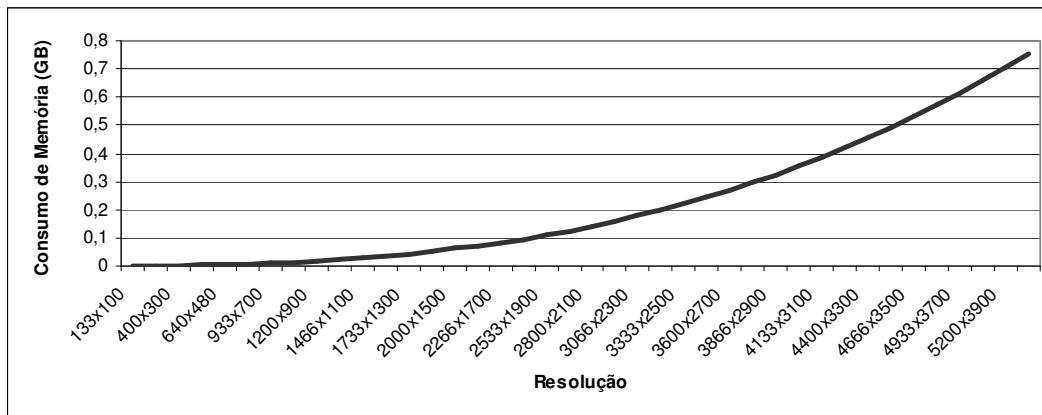


Figura 8:Consumo de memória para técnica das faixas versus resolução, para busca 1D.



## 7.2 Resultados da Procura 2D

A figura 9 apresenta no alto uma referência [4] para a produção do mapa de disparidades usando as imagens de entrada originais e realizando uma busca 1D. Abaixo cada tipo de distorção e o mapa de disparidades gerado tanto para a busca 1D quanto para a busca 2D, fica claro que a busca 1D sofre muito com as distorções simuladas, mas a busca 2D trata muito bem as distorções.

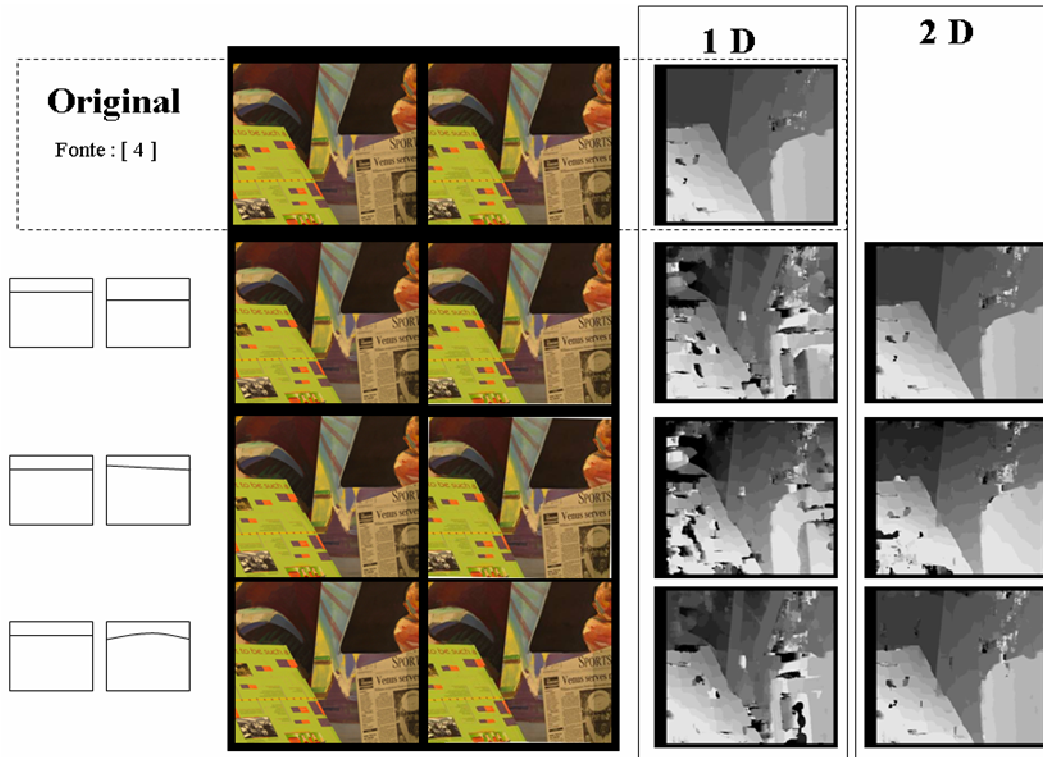


Figura 9: Simulação de distorções e os mapas de disparidades.

## 7.3 Resultados Operador Gradiente

A figura 10 apresenta os efeitos causados no mapa de disparidades: No alto temos imagens originais como referência [1]. No centro temos as imagens de entrada e o mapa de disparidades produzido utilizando a intensidade das cores, visando demonstrar as distorções provocadas pelo fato da imagem da direita (do par estéreo) estar mais clara. Abaixo o resultado usando o operador gradiente, pode-se notar que mesmo com a imagem da direita (do par estéreo) mais clara, o mapa de disparidades apresentou resultado bastante similar ao das imagens de referência.



Figura 10: Efeitos causados pela variação de luminosidade entre as imagens da esquerda e direita.

## 8 Conclusões

A solução aqui apresentada endereça aspectos importantes para a redução de custos. A Técnica das Faixas permite uma redução significativa do consumo de memória, o que possibilita o uso de computadores convencionais. A procura 2D flexibiliza a calibração das câmeras e não exige o uso de câmeras precisas e finalmente o uso do Operador Gradiente permite que pequenas variações na luminosidade não afetem a qualidade do mapa de disparidades não exigindo que o ambiente fotografado esteja igualmente iluminado em ambas as fotos.

## Referências

- [1] Scharstein, Daniel & Szeliski, Richard. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. **International Journal of Computer Vision**, 47(1/2/3):7-42, April-June 2002.
- [2] Scharstein, Daniel. View Synthesis Using Stereo Vision. **Springer**, 1999.
- [3] Gonzalez, Rafael C. & Woods, Richard E. **Processamento de Imagens Digitais**. Editora Edgard Blücher Ltda, 2000.
- [4] Scharstein, Daniel & Szeliski, Richard. High-accuracy stereo depth maps using structured light. **IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003)**, volume 1, pages 195-202, Madison, WI, June 2003.