

Visão Computacional e Realidade Aumentada

Trabalho 3 – Reconstrução 3D

Pedro Sampaio Vieira

Prof. Marcelo Gattass

1 Introdução

Reconstrução tridimensional é um assunto muito estudado na área de visão computacional. O objetivo desta técnica é reconstruir no mundo virtual um objeto ou cena do mundo real.

Um meio de se realizar tal processo é utilizar um *scanner 3D*, dispositivo que analisa um objeto (ou cena) capturando sua forma e aparência. Tal dispositivo é amplamente usado na indústria cinematográfica, no desenvolvimento de jogos eletrônicos, na automação industrial, engenharia, etc. Porém esses tipos de aparelhos possuem um custo muito elevado.

Uma alternativa de menor custo financeiro para esse processo, é a fotografia 3D. Tal abordagem consiste em realizar a reconstrução tridimensional a partir de imagens digitais capturadas com uma câmera fotográfica.

Com isso, é importante salientar que o princípio de todo esse processo é a *estereoscopia*, onde um mesmo objeto (na mesma posição) é analisado de pontos de vista diferente. Ela é responsável por criar a sensação de profundidade, recuperando a informação tridimensional do objeto a ser reconstruído.

Este trabalho consiste em apresentar as técnicas e o processo de reconstrução através da fotografia 3D. Na seção seguinte, é descrita a forma em que foi realizada a aquisição das imagens que farão parte da análise; a seção 3 apresenta e aborda a utilização dos padrões de luz estruturada; na seção 4 é explicado o processo de recuperação das informações tridimensionais (profundidade) para se criar uma nuvem de pontos; e por último

são apresentadas técnicas responsáveis pelo alinhamento das nuvens de pontos e criação da malha do objeto.

2 Aquisição e preparação das imagens

Antes de sair tirando fotos do objeto a ser reconstruído, deve-se primeiramente, montar toda uma estrutura para que as imagens capturadas tenham uma boa qualidade. Essa é uma parte muito importante do processo, pois qualquer coisa que passe despercebido ou alguma mudança, mesmo que pequena, tenha acontecido durante a aquisição das imagens, afeta todo o resultado final.

A estrutura para a captura das imagens contém: 2 câmeras, 1 projetor e uma mesa giratória. Cada câmera foi posicionada de forma que fotografasse uma parte do objeto. Sendo assim, fica definido, por questão de notação, que temos a câmera da direita e a câmera da esquerda. Já o projetor foi posicionado entre as duas câmeras, de forma que a projeção cubra todo o objeto. Por sua vez, o objeto se encontra em cima de uma mesa giratória, para que fosse possível realizar somente rotações em seu eixo, evitando qualquer movimentação do objeto que poderia ocorrer caso fosse feito manuseando diretamente ele. A figura 1 mostra a configuração desse cenário.

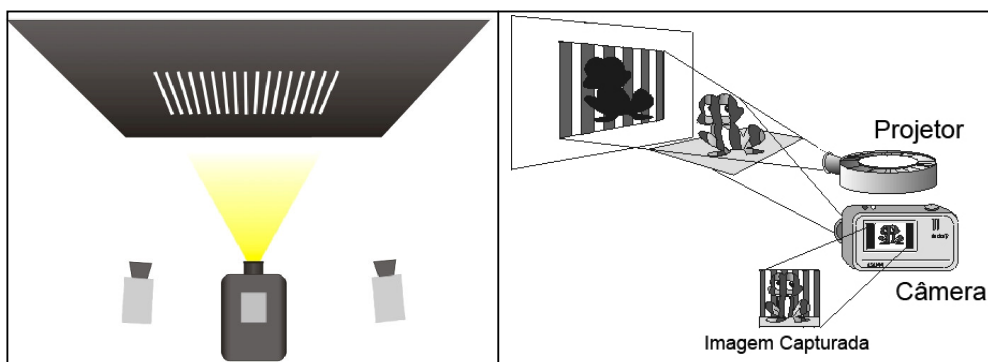


Figura 1: Configuração estrutural para a captura das imagens.

Para atingir o objetivo de criar totalmente um modelo virtual é necessário fotografar, de forma panorâmica, todas as faces do objeto, diminuindo qualquer problema de oclusão que possa ocorrer. Para isso, se faz necessário o uso da

mesa giratória, já que o *setup* das câmeras e do projetor não deve ser modificado no meio do processo. Assim, o objeto foi fotografado, por ambas as câmeras, nos ângulos de 0° , 20° , 40° , 60° , 80° , 100° , 120° e 140° . Para cada rotação foram tiradas 8 fotos, que serão explicadas mais detalhadamente na próxima seção. É importante destacar que todo o processo explicado nas seções 3 e 4 são referentes apenas a um conjunto de imagens do objeto em uma rotação, logo devendo ser repetido para os demais conjuntos. A figura 2 mostra as várias fotos do objeto em rotações diferentes. Nota-se também que as imagens foram pré-processadas para eliminar o máximo de “sujeira” referente ao fundo.

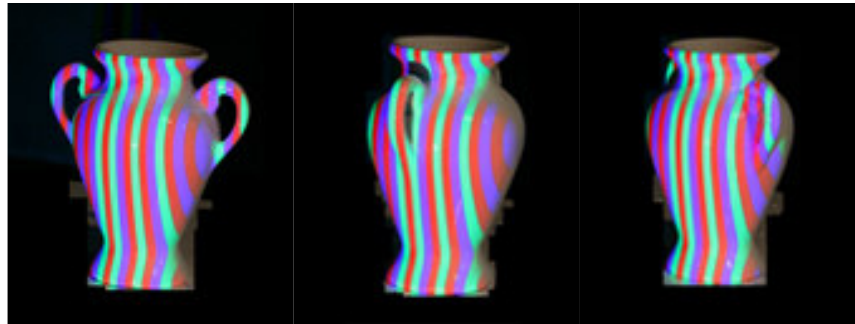


Figura 2: Fotos do objeto em diferentes rotações, já com um padrão de luz estruturada projetado. O objeto se encontra em sua posição inicial, rotacionado em 80° e rotacionado em 140° , respectivamente.

Como nas próximas etapas é necessário saber os parâmetros extrínsecos e intrínsecos da câmera, foi feita uma foto em cada câmera de um padrão de calibração para realizar o processo de calibração de câmera. As técnicas relacionadas a esse procedimento são explicadas com mais detalhes em [2].

3 Processamento das imagens

Nesta etapa, o objetivo é extrair as curvas de nível das imagens. Para isso, foi utilizado o padrão de luz estrutura (3,2)-BCSL criado por [3], onde temos 3 cores (vermelho, verde e azul) em cada slide para 2 slides projetados por vez. Cada slide é projetado em cima do modelo 3D, e possui uma certa

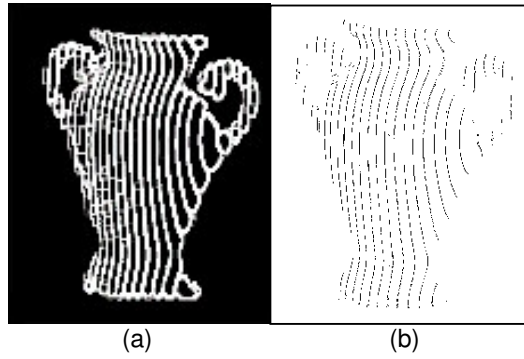


Figura 4: (a) Filtro laplaciano aplicado à imagem. (b) Curvas de nível.

Sendo assim, foram usados mais 4 slides iguais aos anteriores, porém com as faixas na direção horizontal. Isso para que se tenham cruzamentos entre as transições, possibilitando ter, agora, uma identificação de cada ponto (referente a um cruzamento qualquer) através da composição dos códigos providos pela análise das transições verticais e horizontais. Com isso, tal ponto se torna único na imagem, pois não há outro ponto que possua a mesma codificação (ver figura 5c). Por isso, torna-se necessário ter essas 8 fotos tiradas por cada câmera, como foi mencionado na seção anterior.

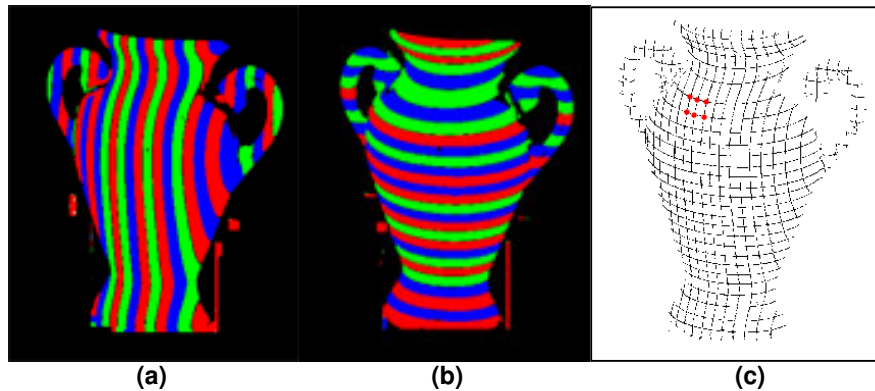


Figura 5: (a), (b) Padrão com faixas verticais e horizontais projetado no objeto. (c) Composição de (a) e (b) destacando (em vermelho) pontos referentes aos cruzamentos das faixas.

Desta forma, com base nas duas câmeras pode-se identificar e criar um conjunto formado por pares de pontos que são vistos por ambas as câmeras, ou seja, um mesmo ponto no objeto que pode ser visto pela câmera da esquerda e pela câmera da direita. Tal conjunto será necessário para realizar o cálculo da profundidade, que irá gerar a nuvem de pontos.

4 Calculando a nuvem de pontos

Uma nuvem de pontos é um conjunto de pontos no espaço tridimensional que descreve a forma (superfície) de um objeto 3D. Para o cálculo desse conjunto, se faz importante o entendimento da *geometria epipolar* (ver figura 6).

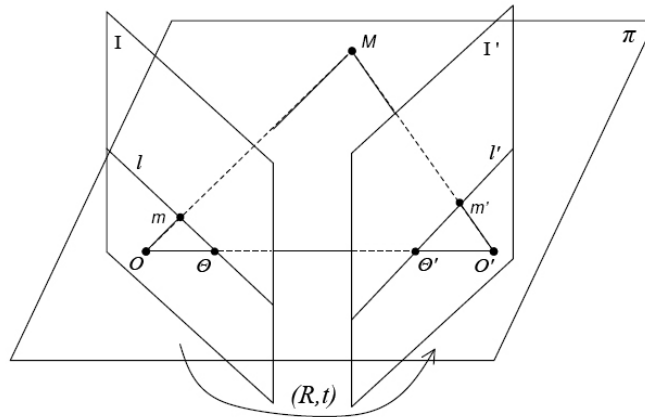


Figura 6: Geometria epipolar.

A geometria epipolar entre duas imagens de uma mesma cena capturadas pelas câmeras, é essencialmente a geometria da intersecção dos dois planos de imagem I e I' com o feixe de planos que tem como eixo a linha de base (linha que une os dois centros ópticos O e O'). Um ponto M do espaço tem projeções m e m' (pontos correspondentes respectivamente na câmera da esquerda e na câmera da direita), em que se deseja encontrar a relação entre eles. A determinação dessa correspondência é o primeiro passo para se derivar informações do espaço tridimensional a partir de duas imagens em estéreo. As intersecções entre os planos I e I' e a linha OO' são os chamados *epipolos* θ e θ' , representando a projeção do centro óptico da outra câmera no plano de imagem. O plano contendo a linha OO' e o ponto M é o *plano epipolar*. A intersecção do plano epipolar com os planos de imagem define as linhas l e l' , chamadas *linhas epipolares*. Dessa forma, todas as linhas epipolares de uma imagem passam pelo epipolo da mesma. Então, para um dado ponto m no plano I , seu correspondente no plano I' deve estar sobre a linha l' e vice versa. A transformação que mapeia os pontos de uma imagem para linhas da outra imagem, e vice versa, é representada por uma matriz F

chamada matriz fundamental. Na equação 1 temos Fm' descrevendo a linha epipolar onde o ponto correspondente m deve estar na outra imagem.

$$m^T F m' = 0 \quad (1)$$

Quando se utilizam câmeras calibradas para obtenção das imagens, ou a mesma câmera sem alteração de configuração em posições diferentes, pode-se utilizar coordenadas normalizadas e ignorar a matriz de parâmetros intrínsecos no modelo da câmera. Nesse caso, a matriz fundamental passa a ser denominada *matriz essencial*.

Porém, o fato de acharmos pontos nas interseções das faixas resolve-se o problema da correspondência de pontos, já que cada um desse ponto tem um código que o torna único na imagem. Assim, conhecendo os parâmetros das câmeras e tendo 2 pontos correspondentes, é possível obter a informação de profundidade de uma forma mais simples.

Portanto, ao final do tratamento dos conjuntos de imagens referentes a cada posição de rotação do objeto e das respectivas análises de correspondência, temos gerado uma nuvem de pontos para cada conjunto. A partir daí, deve-se agora tratar todos esses dados para que no final tenha-se apenas uma nuvem de pontos representando todo o modelo 3D do objeto fotografado.

5 ICP e Triangulação

Com as nuvens de pontos calculadas, as próximas etapas são alinhar essas nuvens e gerar a malha que irá formar o objeto 3D. Para o alinhamento foi utilizado o algoritmo *ICP (Iterative Closest Point)* [4], e para a malha foi feita uma triangulação. Então, vamos entender inicialmente como funciona o ICP.

O problema do alinhamento consiste em encontrar um movimento rígido (R, t) que minimiza o erro

$$\min \sum_i \| q(p_i) - (Rp_i + t) \|^2$$

onde dado 2 nuvens de pontos P e Q, $q(p_i)$ é o ponto de Q correspondente ao ponto p_i de P. Porém, a priori, não se conhece tal ponto em Q que corresponde a p_i .

O algoritmo ICP inicia-se com um “chute” inicial de R e t . A partir daí, $q(p_i)$ é escolhido como o ponto mais próximo em Q de $Rp_i + t$. Com isso, R e t são recalculados a fim de minimizar o erro descrito anteriormente. Esse processo é realizado iterativamente até que não haja mais alguma melhora no ajuste. Assim, o algoritmo é aplicado para cada par de nuvens até que só sobre apenas uma nuvem representando o objeto 3D.

Finalizando a reconstrução 3D, a nuvem de pontos é analisada e a malha do objeto é criada. Para realizar essa operação utiliza-se a técnica de triangularização, que significa encontrar uma forma de conectar os pontos formando vários triângulos, sem que haja algum cruzamento na malha. Logo, tais pontos serão os vértices do modelo 3D.

Referências

[1] GATTASS, M. *Visão Computacional e Realidade Aumentada*. Disponível em <http://www.tecgraf.puc-rio.br/~mgattass/ra/ra.html>.

[2] VIEIRA, P. S. Calibração de câmera. Disponível em <http://www.inf.puc-rio.br/~pvieira/ra.htm>.

[3] SA, A. M.; CARVALHO, P. C. P.; VELHO, L. (b,s)-BCSL: Structured Light Color Boundary Coding for 3D Photography. In: VMV 2002 - 7th International Fall Workshop on Vision, Modeling and Visualization, 2002, Erlangen, Germany. Vision, Modeling, and Visualization 2002 Proceedings. Amsterdam: IOS Press, 2002. p. 299-306.

[4] BESL, P. J.; MCKAY, N. D. A Method for Registration of 3-D Shapes. **IEEE Transactions on Pattern Analysis and Machine Intelligence**. v.14, n.12, feb1992.