

Projeto de Iniciação Científica

Técnicas Atuais de Visão Estéreo Multi-Ocular
Aplicadas à Digitalização de Patrimônio
Comparando Abordagens Passivas e Ativas de Escaneamento 3D

Prof. Ricardo Fabbri, Ph.D.



Grupo de Visualização
Instituto Politécnico – IPRJ
Universidade do Estado do Rio de Janeiro

Nova Friburgo, 6 de Novembro de 2016

Orientador

Prof. Ricardo Fabbri, Ph.D.

Projetos Relacionados

FAPERJ Jovem Cientista do Nosso Estado E25/2014 204167
FAPERJ APQ1 01/09335-8
FAPERJ E28/2014/204747
UERJ Prociênciia 2014–2017

1 Contextualização

O presente plano de pesquisa se refere a atividades propostas a nível de iniciação científica, a serem realizadas no Laboratório de Visualização do Instituto Politécnico – IPRJ da UERJ, Departamento de Modelagem Computacional. Pretende-se trabalhar em conjunto com os grupos de pesquisa do IPRJ, os quais vêm adquirindo notoriedade através de um quadro de professores dedicados a tecnologias audiovisuais bastante diferenciado, notadamente os grupos de pesquisa dos Prof. Francisco Duarte Moura Neto, Prof. Edirlei Soares de Lima, Profa. Alessandra Paulino, Profa. Lis Custodio e Prof. Guilherme Domingues. Este projeto também contará com a colaboração com outros pesquisadores na realização de projetos e publicação de artigos, tanto no IPRJ/UERJ como na UFRJ, IMPA, ICMC (Instituto de Ciências Matemáticas e de Computação) e IFSC (Instituto de Física) da USP de São Carlos, bem como a Brown University (EUA). Também pretende-se colaborar de perto com coletivos de desenvolvedores de software livre como o LabMacambira.sourceforge.net para a tecnologia de software, dentre outros. Este projeto também se contextualiza na formação de um centro de tecnologias audiovisuais mais amplo, sendo instaurado no IPRJ/UERJ, denominado CENA. Diversos projetos têm sido elaborados e aprovados em conjunto pelo grupo de pesquisadores mencionados, de forma a consolidar o CENA.

2 Introdução e Justificativa

A reconstrução 3D de cenas gerais a partir de múltiplos pontos de vista usando-se câmeras convencionais, sem aquisição controlada, é um dos grandes objetivos de pesquisa em visão computacional, ambicioso até mesmo para os dias de hoje. Aplicações incluem a reconstrução de modelos 3D para uso em videogames [1], filmes [1], arqueologia, arquitetura, modelagem 3D urbana (p. ex., Google Streetview); técnicas de *match-moving* em cinematografia para fusão de conteúdo virtual e filmagem real [2], a organização de uma coleção de fotografias com relação a uma cena (p. ex., o sistema *Phototourism* [?]) e a funcionalidade *Look Around* do Google Panoramio e Street View), manipulação robótica, e a metrologia a partir de câmeras na indústria automobilística e metal-mecânica.

Os desafios estão ligados às escolhas de grande escala de representações adequadas e de técnicas que possam modelar simultaneamente com materiais drásticamente diferentes (p. ex., não-Lambertianos), modelos geométricos (p. ex., variedades curvilíneas gerais, descontinuidades, texturas, deformações, em escalas diferentes), tipos de regiões (com ou sem textura), condições de iluminação variadas, sombras, fortes diferenças de perspectivas, desbalanceamento devido a excesso de detalhes em partes menos importantes, número arbitrário de objetos e câmeras não-calibradas.

Mesmo que um sistema completo esteja fora do alcance da tecnologia atual, um progresso significa-

tivo tem sido atingido nos últimos anos. Por um lado, uma tecnologia operacional tem evoluído, mais recentemente para sistemas de grande escala [?, ?], a partir do desenvolvimento da detecção robusta de *features* [?], o *fitting* robusto e seleção de correspondências baseados em RANSAC, e o desenvolvimento de métodos de geometria projetiva para calibrar duas ou três imagens e progressivamente adicionar imagens e extrair estrutura 3D dessas *features* na forma de nuvens de pontos. Com o código fonte do sistema Bundler [?, ?] liberado por Noah Snavely, e sua subsequente incorporação ao sistema VisualSfM [3], é possível utilizar este sistema para a reconstrução de patrimônio.

No paradigma usando-se apenas imagens convencionais – denominado **reconstrução estéreo multiocular passiva** – a posição das câmeras são estimadas a partir apenas de imagens, usando pontos de interesse, em seguida uma nuvem de pontos é reconstruída 1.

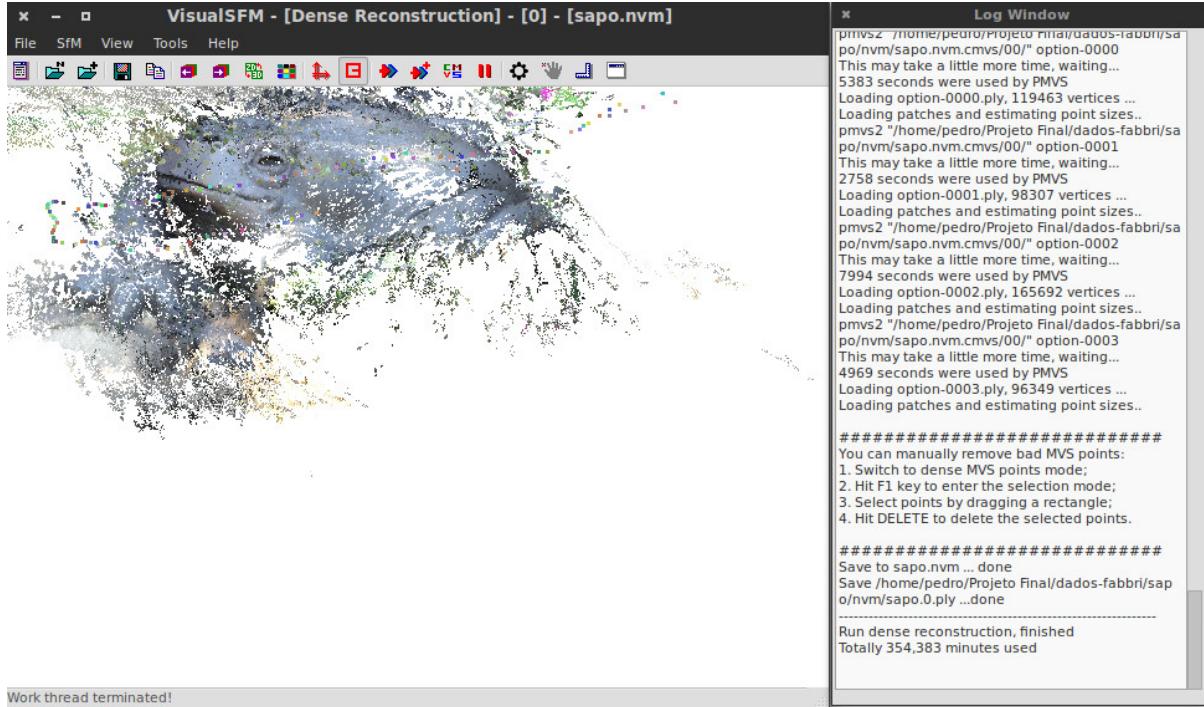


Figura 1: A reconstrução usando-se apenas imagens, sem controle de aquisição, como em um vídeo de um smartphone filmado em torno do objeto, fornece uma nuvem de pontos, que pode ser densificada [?, ?, 3, 4, 9], ou atribuída de curvas [5, 6, 7, 8], de forma a preservar a resolução em áreas de alto conteúdo informativo. Tais representações estão sendo atualmente unificadas na pesquisa da área. Este projeto propõe explorar os limites da reconstrução 3D usando-se apenas imagens, no contexto de preservação de patrimônio.

As câmeras podem então ser utilizadas para obter modelos mais detalhados de reconstrução, como algoritmos de densificação [4] e interpolação [?] da nuvem de pontos, bem como demais algoritmos densos de visão estéreo multi-perspectiva/multi-ocular, como os do grupo de Michel Goesele [9], também com código disponível. Tais algoritmos, no entanto, têm problemas, em particular a reconstrução suaviza partes bem-delineadas do objeto, e pode conter buracos em áreas homogêneas. Pode-se, portanto, utilizar

a reconstrução 3D de curvas do pesquisador proponente [5, 6, 7, 8] para auxiliar na reconstrução mais bem-delinada nesses casos problemáticos, bem como para ajudar no problema de escalabilidade quando a reconstrução 3D se torna muito grande.

Um segundo paradigma, denominado **reconstrução estéreo multiocular ativa**, tem se tornado viável devido à indústria de videogames, e consiste na utilização de sistemas que alteram o funcionamento de câmeras convencionais, típicamente usando-se projetores infra-vermelho, laser ou câmeras ToF (time of flight), como no caso dos dispositivos Kinect, figura 2.



Figura 2: Kinnects de primeira geração (a) consistindo de câmeras e projetores infra-vermelho (b) e de segunda geração, consistindo de tecnologia ToF (c). Ambos os kinnects são largamente utilizados para escaneamento em tempo real, formando a base de scanners manuais (d), porém nem sempre são úteis para preservação detalhada de patrimônio. Um dos objetivos deste projeto é explorar os limites desta tecnologia.

Diversos escanners a laser ou ToF que há alguns anos possuíam um custo elevado, hoje são acessíveis. Muitos deles são baseados nos sensores Kinect, e há software berto disponível na internet para realizar a calibragem externa e reconstrução 3D usando-se tais dispositivos, incluindo a fusão de mapas de profundidade para obtenção de reconstruções 3D com precisão razoável, mesmo que aquém dos sistemas de alto custo, figura 3. No entanto, comparando-se com câmeras convencionais, o hardware ainda é especializado e o custo, apesar de baixo, é uma ordem de magnitude maior que o de câmeras convencionais.

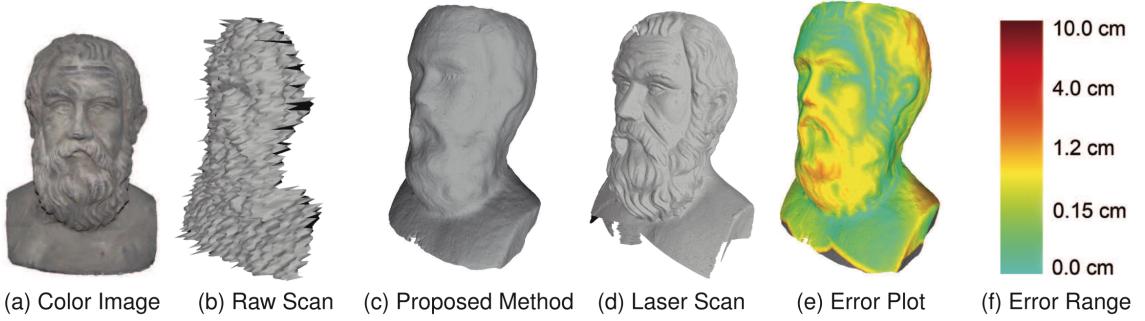


Figura 3: A reconstrução usando-se Kinect (de primeira ou segunda geração) usando software atual de super-resolução (c) fornece precisão similar a um sistema estéreo de média resolução, inferior um sistema a laser de alta qualidade (d) porém de baixo custo e muito mais versátil devido ao sistema de aquisição manual e a software amplamente utilizado[?, ?].

A preservação de patrimônio tem sido realizada tradicionalmente com scanners dedicados de alto custo, como no projeto David [4](#).

O projeto teve início em 1992 e tem como objetivo a utilização de scanners a laser de profundidade (*rangefinder scanners*), aliado com algoritmos que combinam diferentes profundidades e cores da imagem, para realizar uma digitalização da parte externa e da superfície de forma acurada da estátua de David (porém, esse método pode ser utilizado em diferentes objetos no mundo real, como partes de máquinas, artefatos culturais e na indústria de video games, por exemplo). Para as partes mais detalhadas, foi utilizado um scanner de menor escala que faz uma pequena triangulação com laser de profundidade.

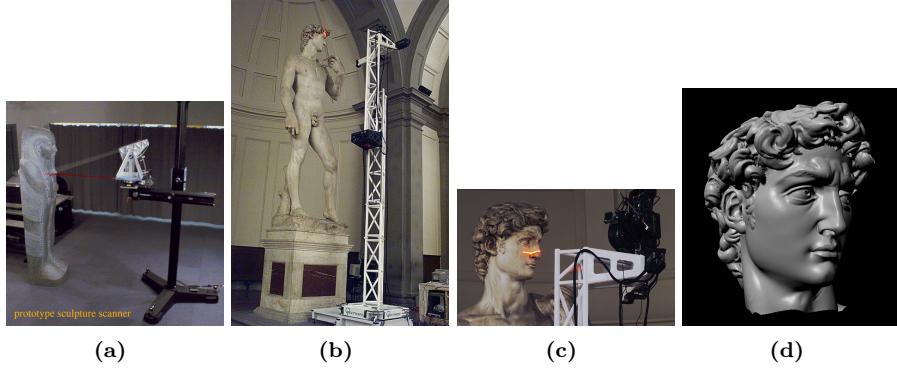


Figura 4: Protótipo do scanner a laser de triangulação. O objeto a ser escaneado é uma réplica em tamanho real de um sarcófago egípcio (a). O scanner foi reconfigurado para escanear objetos maiores, pois a escultura possui 517 centímetros (b), o da cabeça também sofreu uma reconfiguração, este scanner gira em 90 graus, que faz o laser rotacionar, da posição horizontal para a vertical e também roda em torno da cabeça como um todo (c). Para a reconstrução, o primeiro passo foi alinhar cerca de 100 scans em diversas posições, após isso, utilizado um alinhamento automático em pares dos scans, utilizando um algoritmo modificado de iterações de pontos próximos (ICP - *iterated-closest-points*). Após isso, faz-se um processo de relaxação global a fim de minimizar erros de alinhamento por toda a estátua. Depois de alinhados, usa-se o algoritmo de profundidade volumétrica de processamento de imagens (VRIP - *volumetric range image processing* - do Brian Curless) (d).

Seria de grande interesse explorar os dois paradigmas supracitados para avaliar as possibilidades disponíveis no estado da arte de reconstrução 3D para o escaneamento de baixo custo para a preservação de Patrimônio. O que se pode atingir com apenas uma filmagem de esculturas realizada por um smartphone, sem calibração prévia e *in situ*, ou seja, sem ambiente controlado? Como esta reconstrução se compara nos dias de hoje com a reconstrução realizada por um scanner padrão baseado em Kinect?

O Jardim do Nêgo, Nova Friburgo No caso de Nova Friburgo, há a necessidade redobrada de preservação do patrimônio, em especial devido às chuvas e deslizamentos inerentes à região. O Jardim do Nêgo consiste em grandes esculturas em encostas, cobertas por um tapete de vegetação, as quais desfrutam de grande reconhecimento regional e internacional [?], figura 5.



Figura 5: Algumas esculturas do Jardim do Nêgo

Idealizado e criado por Geraldo Simplicio (Nêgo), artista cearense que mora no local a mais de 30 anos, ganhou notoriedade por suas esculturas de barro, com traços singulares e técnicas únicas. Hoje, trabalha para reconstruir o Jardim da tragédia de 2011 na região serrana, onde algumas estruturas foram destruídas. Portanto, com o consentimento do Nêgo, surgiu a motivação desta pesquisa: além de explorar métodos de reconstrução, também tem o objetivo de eternizar um patrimônio que é reconhecido no mundo todo.

A preservação das esculturas do Jardim do Nego se torna um desafio à pesquisa em reconstrução 3D, pois apresentam curvas bem delineadas, que são representadas de maneira suavizada e empobrecida por métodos convencionais. Algumas esculturas apresentam pouca textura, quase sem nenhum padrão de textura/musgo. Seria de grande interesse availiar o potencial de técnicas atuais de reconstrução 3D geral sem controle de aquisição, as quais têm seu código fonte disponível na internet.

3 Objetivos

O presente projeto pretende fazer com que o aluno ganhe experiência com técnicas modernas de reconstrução 3D fotogramétrica, no contexto de uma aplicação bem-definida de preservação de patrimônio.

A entrada do sistema deverá ser um conjunto de vídeos realizados por câmeras de baixo custo, ou um conjunto de escaneamentos realizados por scanners à mão de baixo custo baseados em Kinect.

O objetivo concreto do aluno será explorar as tecnologias supracitadas para desenvolver um esquema de escaneamento usando software aberto, câmeras e scanners de baixo custo, representando o estado da arte em reconstrução 3D sem restrições de aquisição. Perguntas fundamentais a serem respondidas são: que nível de detalhe, facilidade e precisão se pode obter usando-se apenas imagens e software aberto? É possível utilizar scanners de baixo custo baseados em Kinect com melhorias significativas em termos de qualidade, conveniência ou tempo de processamento? Quais são as restrições desses sistemas? Seria útil na prática uma reconstrução de curvas para auxiliar na reconstrução de nuvem de pontos e de superfícies densas? Onde o estado da arte deve ser avançado de forma a permitir uma solução mais conveniente e completa para a preservação de patrimônio?

O principal objetivo em termos de pesquisa científica será comparar as diferentes abordagens do estado da arte disponíveis para reconstrução 3D e explicitar suas limitações práticas. O aluno deverá, com o entendimento das abordagens, desenvolver um esquema de aquisição de esculturas que permita ampliar os detalhes ou ajudar a resolver os problemas dos métodos. Com a experiência obtida, o aluno estará pronto para desenvolver pesquisa futura na área de reconstrução 3D, com conhecimento de causa para avaliar direções de pesquisa de efetivo e alto impacto na prática.

4 Metas

Elaboração de Publicações Como metas específica deste projeto, tem-se a elaboração de uma publicação em congresso ou revista contendo um estudo comparativo das diferentes técnicas modernas de reconstrução 3D para preservação de patrimônio.

Programação e Uso de Software Como meta deste projeto necessária à realização do estudo comparativo, tem-se o uso e a programação de software de reconstrução 3D, em especial VisualSFM [3], Visão estereópica densa de Michel Goesele [9], o sistema de reconstrução 3D do prof. proponente deste projeto [5, 6, 7, 8], Kinect Fusion [?], e técnicas por Luz Estruturada [?]. Por se tratarem de sistemas grandes, esforço considerável é necessário para aprender e modificar esses softwares.

5 Metodologia

A execução do projeto será dividida nas seguintes frentes, visando atingir as metas descritas na seção anterior, na ordem descrita na próxima seção:

- **Aquisição de Dados.** Esta etapa consiste na aquisição *in situ* de vídeo usando-se smartphone, bem como usando-se scanner de mão baseado em Kinect. Numa futura etapa, pretende-se realizar diversas re-aquisições de forma a melhorar a resolução das reconstruções. Na medida em que os dados forem crescendo, torna-se necessária estratégias de *big data* para o processamento, como armazenamento e processamento distribuído.
- **Escrita e Uso de Software Livre.** Uso e a programação de software de reconstrução 3D, em especial VisualSFM [3], Visao estereo densa de Michel Goesele [9], e o sistema de reconstrução 3D do prof. proponente deste projeto [5, 6, 7, 8], e Kinect Fusion [?]
- **Estudo de Tecnologias 3D.** Estudo junto ao orientador acerca do funcionamento dos sistemas de software sendo explorados.
- **Escalabilidade.** Na medida em que o número de imagens cresce, torna-se necessário estudar opções disponíveis para processamento distribuído/paralelo. Esta linha pode se tornar uma via de aprofundamento pelo aluno, se desejado, porém a princípio, num contexto de Iniciação Científica, trata-se de estudar soluções já existentes para uso no Cluster do IPRJ/UERJ.
- **Elaboração de Publicações.** Elaboração de uma publicação em congresso ou revista contendo um estudo comparativo das diferentes técnicas modernas de reconstrução 3D para preservação de patrimônio.

6 Cronograma das Atividades

A tabela abaixo mostra o cronograma para as principais frentes de atuação descritas na seção anterior. Para a execução deste projeto, o período foi dividido em etapas aproximadamente semestrais.

Etapas semestrais	1	2	3	4
Aquisição de Dados (etapas 1 e 2)	X		X	
Escrita e Uso de Software Livre	X	X	X	X
Estudo de Tecnologias 3D	X	X	X	
Escalabilidade			X	X
Elaboração de Publicações		X	X	X

7 Conclusão

Este projeto da área de visão computacional propõe a exploração e comparação das principais técnicas modernas de escaneamento 3D, no contexto bem-definido de preservação de patrimônio. Além da impor-

tância em obter experiência prática no estado da arte em de visão computacional 3D, o projeto atende a uma demanda real de preservação de patrimônio cultural e de outros objetos em áreas propensas a desastres naturais, como Nova Friburgo, onde tal preservanção toma uma importância imediata. As esculturas do Jardim do Nêgo, escolhidas para este projeto, fornecem uma geometria única e desafiadora, permitindo explorar os limites das técnicas atuais de reconstrução 3D. Pretende-se, assim, gerar uma publicação acerca da experiência prática com técnicas modernas para preservação de patrimônio cultural, além de gerar um arquivo digital das esculturas, bem como uma metodologia conveniente e de baixo custo para digitalização que possa ser utilizada em qualquer situação. Aplicações mais amplas incluem a monitoração contínua de encostas, reconstrução urbana e de interiores a partir de câmeras e scanners de baixo custo.

Referências

- [1] D. Ablan, *Digital photography for 3D imaging and animation*. Wiley, 2007. [1](#)
- [2] T. Dobbert, *Matchmoving: The Invisible Art of Camera Tracking*. Sybex, 2005. [1](#)
- [3] C. Wu, “VisualSFM – visual structure from motion system.” <http://ccwu.me/vsfm/>. [2](#), [6](#), [7](#)
- [4] Y. Furukawa and J. Ponce, “Accurate, dense, and robust multi-view stereopsis,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1–8, IEEE Computer Society, 2007. [2](#)
- [5] A. Usumezbas, R. Fabbri, and B. B. Kimia, “From multiview image curves to 3D drawings,” in *Proceedings of the European Conference on Computer Vision*, Lecture Notes in Computer Science, Springer, 2016. [2](#), [3](#), [6](#), [7](#)
- [6] R. Fabbri and B. B. Kimia, “Multiview differential geometry of curves,” *International Journal of Computer Vision*, vol. 117, pp. 1–23, 2016. [2](#), [3](#), [6](#), [7](#)
- [7] R. Fabbri and B. B. Kimia, “3D curve sketch: Flexible curve-based stereo reconstruction and calibration,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (San Francisco, California, USA), IEEE Computer Society Press, 2010. [2](#), [3](#), [6](#), [7](#)
- [8] R. Fabbri, P. J. Giblin, and B. B. Kimia, “Camera pose estimation using first-order curve differential geometry,” in *Proceedings of the IEEE European Conference in Computer Vision*, Lecture Notes in Computer Science, Springer, 2012. [2](#), [3](#), [6](#), [7](#)

[9] M. G. Simon Fuhrmann, Fabian Langguth, “Mve à a multi-view reconstruction environment,” 2014.

[2](#), [6](#), [7](#)

[10] R. Fabbri, L. Estrozi, and L. Costa, “On Voronoi diagrams and medial axes,” *Journal of Mathematical Imaging and Vision*, vol. 17, pp. 27–40, jul 2002.