

Plano de Trabalho de Doutorado

Machine learning por Difusão em Grafos para
Reconstrução 3D da Geometria Trinocular de Curvas

Juliana Santos Barcellos Chagas Ventura



Laboratório de Visualização
Laboratório de Tecnologia da Informação
Instituto Politécnico – IPRJ
Universidade do Estado do Rio de Janeiro

Nova Friburgo, 7 de Outubro de 2019

Projetos Relacionados

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

1 Contextualização

O presente plano de pesquisa se refere a atividades propostas para a aluna Juliana Ventura, a serem realizadas a nível de doutorado no Instituto Politécnico – IPRJ/UERJ, Departamento de Modelagem Computacional. Pretende-se trabalhar em conjunto com dois orientadores, os quais seriam prof. Ricardo Fabbri, do laboratório de visualização, e o prof. Francisco Duarte Moura Neto, do laboratório de tecnologia da informação (LTI), juntando-se os aspectos ativos recentemente das pesquisas de ambos. Pretende-se trabalhar em conjunto com outros grupos de pesquisa além desses dois laboratórios dentro do IPRJ, bem como outros pesquisadores no IMPA e em outras instituições em colaboração internacional na realização de projetos e publicação de artigos. Dentre as instituições internacionais envolvidas, tem-se: Brown University, University of Liverpool, Notre Dame University, Georgia Tech, Czech Technical University, e General Motors Research.

2 Introdução

Na presente década está-se presenciando um nível de investimentos sem precedentes em ecossistemas de tecnologia para modelagem 3D de fenômenos usando-se múltiplas imagens adquiridas por uma câmera em diferentes posições, múltiplas câmeras, ou vídeo [1]. Esse impacto cada vez maior da visão computacional 3D se deve em grande parte a tecnologias de *structure from motion* (SfM) desenvolvidas na década passada, a qual possibilitou novas ferramentas para resolver problemas chave de forma robusta e automática. No entanto, o vasto potencial das tecnologias 3D permanece ainda inexplorado, em parte devido à rentabilidade de aplicações mais simples e imediatas, e em parte devido a limitações da tecnologia evidentes na medida em que aplicativos práticos são desenvolvidos, como realidade aumentada [1]. Pode-se afirmar que tais limitações se devem em três problemas centrais: o uso de geometria projetiva clássica de pontos, sem a geometria diferencial necessária para modelar estruturas mais gerais; o uso sub-ótimo de múltiplas imagens; e a necessidade de modelos 3D mais estruturados.

O grupo de pesquisa do prof. Ricardo Fabbri tem tradição em um novo paradigma fotogramétrico para a reconstrução 3D de objetos rígidos e não-rígidos a partir de múltiplas imagens e vídeo [2, 3, 4, 5, 6, 7, 8]. A pesquisa do prof. Fabbri tem como objetivo suplantando as limitações supracitadas da tecnologia usual de reconstrução 3D, a qual é baseada em (e limitada por) pontos de interesse. Uma das formas para se atingir esse objetivo é empregando-se a informação disponível em curvas perceptuais em imagens, através de um modelo 3D com progressivos níveis de detalhes baseado no chamado *3D curve sketch* (esboço 3D de curvas) e o *3D curve drawing* (desenho 3D de curvas). Tendo obtido um esboço inicial dos objetos 3D através do casamento de curvas ao longo das múltiplas imagens, o modelo 3D pode ser refinado em um grafo de curvas, similar a um desenho 3D. Superfícies podem ser estendidas sobre esse desenho 3D para gerar uma reconstrução detalhada e robusta a iluminação. Além da proposta do paradigma baseado em curvas, prof. Fabbri e colaboradores foram os primeiros a introduzir a técnica CAD chamada Lofting à visão computacional em uma versão automática.

Este projeto objetiva unir a pesquisa em aprendizagem de máquina baseada em difusão em grafos – proveniente da colaboração entre os orientadores prof. Ricardo Fabbri e prof. Francisco Duarte Moura Neto – com a pesquisa específica em reconstrução 3D do prof. Fabbri, visando melhores algoritmos para inicializar modelos de câmera e de objetos curvilíneos com pouca textura. A ideia principal é capitalizar nos avanços recentes e de alto impacto realizados pelo prof. Fabbri durante sua estadia no ICERM / Brown University, ao final de 2018 e no início de 2019 como organizador do Research Cluster em “Visão Algébrica”, reunindo os melhores professores visitantes e seus alunos de doutorado. No início de 2019, o grupo organizado pelo prof. Fabbri conseguiu levar consigo o ex-aluno de mestrado do IPRJ David da Costa de Pinho, e sua tese de mestrado desenvolvida no IPRJ teve grandes desdobramentos. Este projeto também almeja capitalizar no sucesso e grande interesse demonstrado pela comunidade internacional nesta dissertação de mestrado do aluno do IPRJ. Apesar dos problemas algébricos locais terem sofrido grande avanço durante a estadia de David e Fabbri no ICERM/Brown University, ainda restam dois problemas chave: 1) como encontrar o melhor problema inicial para aumentar a velocidade da continuação homotópica para calcular configurações de três câmeras para novas imagens, e 2) como integrar a informação local para se utilizar a estrutura global de curvas.

Com o acabo de aprendizagem de máquina trabalhado entre os prof. Fabbri e prof. Moura Neto nos últimos anos, que envolve difusão em grafos e Teoria de Padrões, é possível no problema 1) treinar

um mapa entre as entradas de três pontos-tangentes em curvas e os 11 graus de liberdade de soluções iniciais para a escrita de um *solver* rápido por continuação homotópica, realizando aprendizagem com sistemas dinâmicos de soluções algébricas; e 2) realizar a estimativa inicial de um sistema de câmeras a partir de três imagens, explorando-se a estrutura global de curvas. O uso de três imagens fornece uma dimensionalidade e estrutura interessante: equações bem definidas, estrutura de curvas em duas e três dimensões, bem como um espaço de 11 dimensões descrevendo três câmeras. Apenas duas imagens não permitiria o uso de estruturas curvilíneas para a estimação de câmeras, e o estudo de mais de três imagens cai em problemas genéricos de aprendizagem e otimização e intratabilidade, com pouca robustez. Mais detalhes são fornecidos a seguir.

3 Reconstrução do Desenho 3D a partir de múltiplas perspectivas

Esta seção resume o trabalho anterior de Fabbri *et al.* sobre o paradigma do Desenho 3D para reconstrução de cenas (coleções de objetos) a partir de múltiplas imagens. O fundamento matemático está presente em [9], relacionando as propriedades diferenciais de uma curva 3D $\Gamma(S)$, onde Γ é um ponto 3D (possivelmente em uma curva) no sistema de coordenadas de câmera, e S is é uma parametrização da curva 2D correspondente, $\gamma(s)$, onde γ é um ponto em uma imagem e s é uma parametrização da curva 3D. Usando-se a relação $\Gamma(S) = \rho(s)\gamma(S)$, onde ρ é a profundidade do ponto denotado por Γ , a geometria diferencial da curva 3D $\{\mathbf{T}, \mathbf{N}, \mathbf{B}, K, \dot{K}, \tau, G, G', G''\}$ é relacionada à geometria diferencial da curva $\{\gamma, \mathbf{t}, \mathbf{n}, \kappa, \dot{\kappa}, g, g', g'', \rho, \rho', \rho'', \rho'''\}$, onde $\{\mathbf{T}, \mathbf{N}, \mathbf{B}\}$ é o triedro de Frenet da curva 3D, *i.e.*, a tangente, normal e binormal unitário, K e τ são as curvatura e torção and torsion da curva espacial, resp., κ é a curvatura da curva numa imagem, e G e g são as velocidades de parametrização das curvas 3D e 2D, resp., e onde “ ’ ” e “ · ” simbolizam a diferenciação relativa a um parâmetro espacial arbitrário (S ou s) ou parametrização por comprimento de arco, resp.

Note-se que $\dot{\kappa}$ e \dot{K} compõe modelos intrínsecos das curvas. Foi demonstrado pela primeira vez por Fabbri *et al.* [9] como essas quantias estão relacionadas. Por exemplo, o vetor tangente à curva espacial pode ser escrito como $\mathbf{T} = \rho'\gamma + \rho g\mathbf{t}$, e a tangente à curva projetada é $\mathbf{t} = \frac{\mathbf{T} - e_3^T \mathbf{T} \gamma}{\|\mathbf{T} - e_3^T \mathbf{T} \gamma\|}$, onde $e_3^T = (0, 0, 1)$.

Outro exemplo é que a razão das velocidades de parametrização é uma quantia intrínseca $\frac{g}{G} = \frac{\|\mathbf{T} - e_3^T \mathbf{T} \gamma\|}{e_3^T \mathbf{T}}$. Também foi demonstrada a reconstrução de uma curva espacial (*i.e.*, de torção não-nula) pela primeira vez na literatura.

Essas relações são muito impotantes para se relacionar três ou mais imagens de multiplas vistas ou para reconstruir-se uma cena: foi mostrado como a geometria diferencial em duas vistas (tangente, curvatura, e derivada da curvatura) reconstroem a geometria diferencial em 3D (tangente, curvatura e torção em 3D), Figure 1 Neste projeto, focaremos na continuidade da pesquisa de Fabbri no ICERM/Brown University iniciada por seu aluno de mestrado David de Pinho, a qual foca nas propriedades de primeira ordem (tangente) como um primeiro porém importantíssimo passo que já despertou grande interesse da comunidade científica e industrial (no ramo de realidade aumentada e reconstrução de fluidos).

A linha de trabalhos mais práticos de Fabbri *et al.* permite extrair um desenho 3D e superfícies acopladas, representando objetos vistos por imagens. As etapas desse sistema prático de reconstrução 3D são ilustradas na figura 2, sendo que nem todos os módulos ainda foram publicados. De maneira resumida, uma vez que as posições de câmeras tenham sido calculadas, usando-se o trabalho desenvolvido por Fabbri junto ao aluno David de Pinho, e junto ao célebre matemático Peter Giblin, cada segmento de curva em uma imagem busca porções correspondentes de curvas em alguma outra imagem para assim hipotetizar um fragmento de curva 3D, hipótese esta que é confirmada ou refutada reprojetando-se em outras imagens. Esse processo resulta em um conjunto inicial desorganizado de fragmentos 3D denominado “esboço 3D de curvas” [10, 9], Figure 2 (*3D curve sketch*). Neste estágio inicial de reconstrução, o esboço 3D de curvas sofre de erros de agrupamentos e fragmentações, além de redundância já que a informação ainda não foi integrada para diferentes pares de imagens, o que motivou Fabbri *et al.* [11] a explorar a conectividade topologica para produzir um grafo conexo de curvas longas, denominado “desenho 3D de curvas”, ver terceira linha na Figura 2.

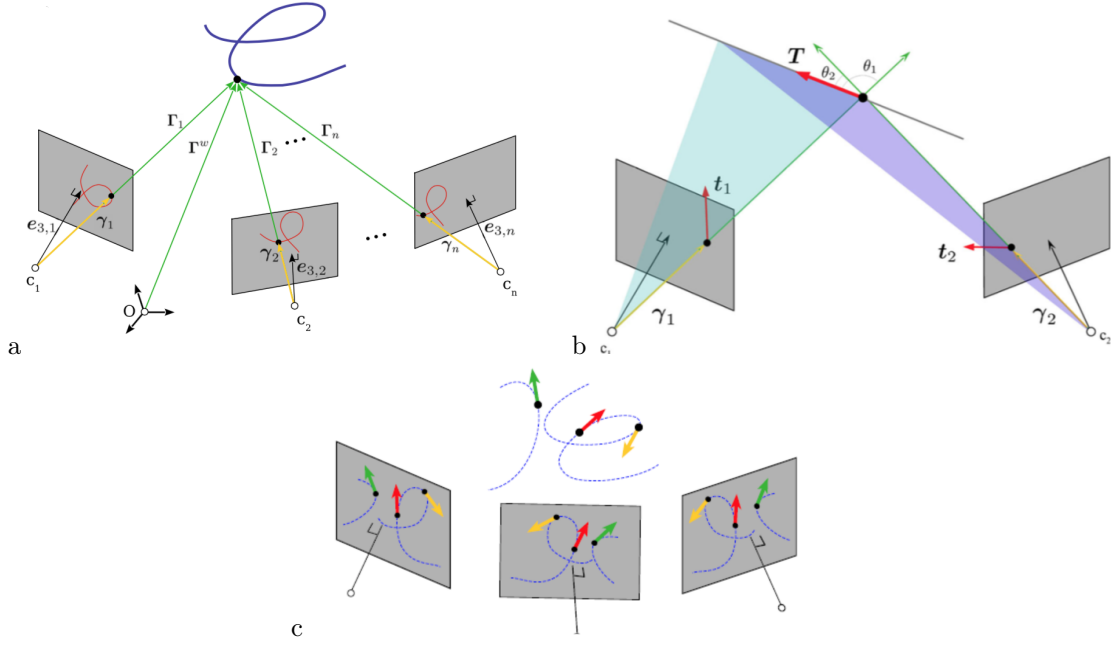


Figura 1: A geometria diferencial de uma curva 3D e suas projeções são relacionadas por equações, no caso um modelo local de primeira ordem (vetor tangente) em duas imagens reconstroem as tangentes no espaço tridimensional, e, através deste, as tangentes (modelos locais) em outras imagens.

4 Objetivos

Na Figura 2 ilustrando os módulos do sistema de reconstrução 3D que compõe a pesquisa do prof. Fabbri, pretende-se abordar as caixas em verde na primeira linha, em especial o módulo “Curve-Based Trifocal Pose Estimation”, usando-se ferramentas de aprendizagem de máquina baseada em difusão em grafos e teoria de padrões.

5 Aprendizagem de caminhos em *solvers* por continuação homotópica

Pretende-se explorar a estrutura de sistemas dinâmicos proveniente do rastreamento de soluções de equações não-lineares algébricas para produzir um solver rápido baseado em continuação homotópica. Durante a estadia do prof. Fabbri liderando um “research cluster” (grupo de pesquisa temporário, similar a um workshop de longa duração), foi produzido, pela primeira vez, um novo *solver* para o problema de se determinar a posição relativa de três câmeras a partir de três imagens de curvas, Figura 1. Foi determinado que o problema é sobredeterminado, e ignorando-se uma tangente, o problema se reduz a um polinômio de grau 320 [12]. Foi produzido um solver rápido e determinístico com técnicas estado da arte de geometria algébrica numérica no rastreamento de soluções por continuação homotópica, resolvendo 11 variáveis de três câmeras (rotação e translação) na ordem de 600ms. No entanto, diversas aplicações práticas requerem um tempo de execução em nanosegundos ou mesmo microsegundos. Um dos objetivos deste projeto é aprender a estrutura das equações e das trajetórias desse sistema dinâmico realizado pelo *solver*, de forma a melhor inicializar as 320 soluções o mais próximo possível do resultado final. No entanto, o comportamento dessas trajetórias é pouco conhecido dos pesquisadores em geometria algébrica. Pretende-se, para isso, aprender o comportamento das trajetórias e utilizar o conhecimento de equações diferenciais do prof. Francisco Moura Neto.

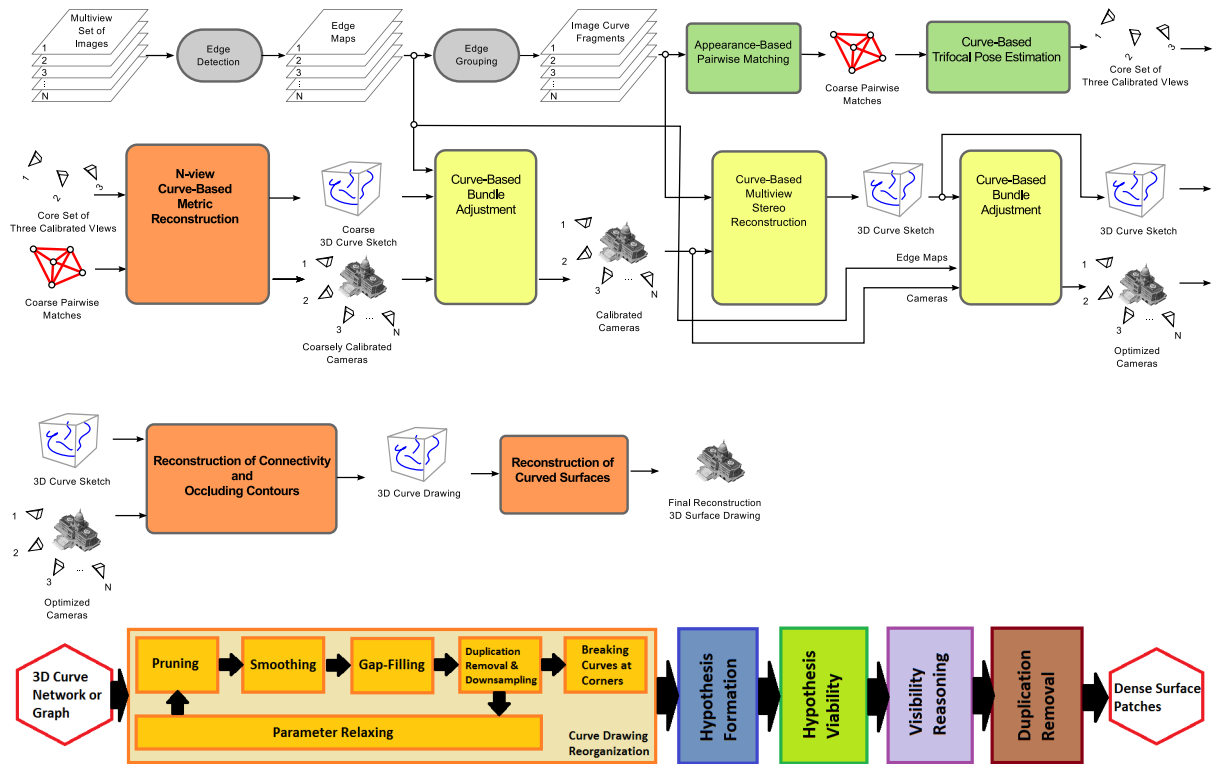


Figura 2: Uma ilustração de um sistema completo de visão 3D baseado em curvas e superfícies. Muitos desses módulos ainda estão em desenvolvimento.

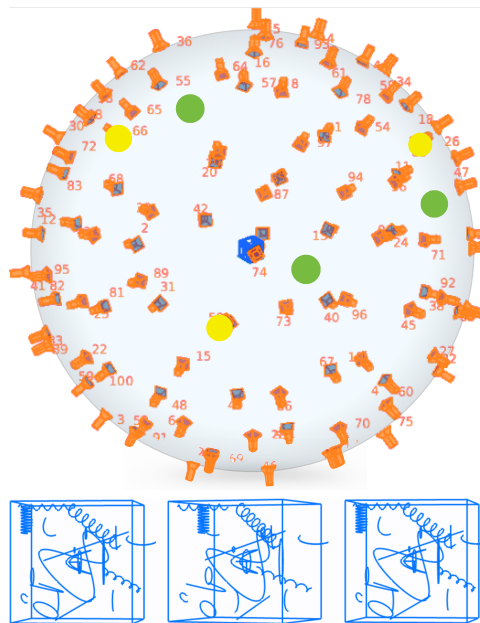


Figura 3: Dados de “treinamento” fornecendo uma amostra representativa de 100 configurações iniciais. Em cada imagem, o solver por continuação homotópica fornece trajetórias que podem ser aprendidas e otimizadas, para gerar um solver muito rápido.

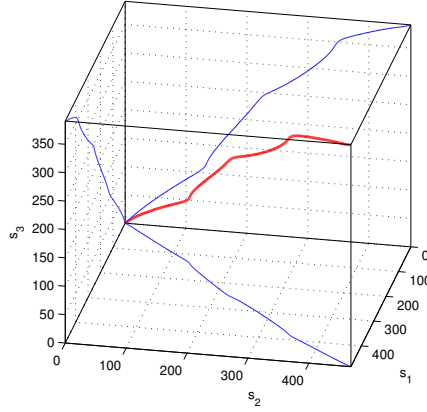


Figura 4: Curva de alinhamento entre três curvas em três imagens.

6 Geometria trifocal de curvas curvas: de local a global

O segundo grande objetivo deste trabalho é levar o *solver* local citado na seção anterior à construção de um *solver* global, que utilize não apenas triplas de pontos-tangentes nas curvas, como também integre solução na curva toda. Ao mesmo tempo, isto permite resolver o problema da correspondência entre curvas, através de um sistema de hipotetizar-e-testar.

Seja um conjunto de fragmentos de curvas em correspondência hipotética C_f , $f = 1, 2, \dots, F$ em F imagens. Seja $\alpha(\xi)$ a curva de alinhamento definida por

$$\alpha(\xi) = (s_1(\xi), s_2(\xi), \dots, s_F(\xi)), \quad \xi \in [0, 1],$$

tal que os pontos $\{C_1(s_1(\xi)), C_2(s_2(\xi)), \dots, C_F(s_F(\xi))\}$ estão em correspondência. Usa-se a notação $\alpha_{ij}(\xi) = (s_i(\xi), s_j(\xi))$ para simbolizar o alinhamento entre fragmentos i e j .

Dado o alinhamento entre quaisquer dois fragmentos α_{ij} , pode-se reconstruir a geometria epipolar (representando a disposição relativa entre as imagens), já que $\xi \in [0, 1]$ gera inúmeros pares em correspondência, p. ex., $\xi_1, \xi_2, \dots, \xi_7$ fornece sete pontos $(C_1(s_1(\xi_i)), C_2(s_2(\xi_i)))$, $i = 1, 2, \dots, 7$.

Para o caso de interesse de três imagens, $\alpha_{12}(\xi) = (s_1, s_2)$ e $\alpha_{13}(\xi) = (s_1, s_3)$ fornece, a princípio, curvas de alinhamento independentes (C_1, C_2) e (C_1, C_3) . No entanto, isto não é livre de restrições entre (C_2, C_3) . Na integração de informação local, mapas de difusão e teoria de padrões podem auxiliar; em particular a curva de alinhamento não é uma curva completamente livre, no início do sistema de reconstrução 3D, mas é parte do espaço de curvas monotônicas, cuja dimensão é restrita. Empregar esses tipos de algoritmos no caso de informação discreta é um dos pontos fortes da pesquisa com o prof. Mora NEto.

7 Justificativa: as vantagens das curvas.

Além de possibilitarem as inúmeras aplicações do sistema proposto citadas no início da Introdução, as curvas como *features* básicas para reconstrução 3D e calibração têm as seguintes vantagens. As curvas provenientes de descontinuidade de bordas nas imagens são mais densas e mais bem-estruturadas do que pontos isolados, e representam eficientemente a imagem ou cena 3D subjacente. Elas fornecem um bom compromisso entre uma representação altamente redundante na forma de arrays de pixels, e uma representação altamente esparsa das nuvens de pontos de interesse. Isso permite o cálculo rápido de reconstruções 3D (os *3D curve sketch*), que, apesar de não possuírem superfícies explícitas, são ainda reconhecíveis, estruturadas, e de eficiente manipulação e armazenagem. Essa eficiência representacional é evidenciada em uma tendência recente em computação gráfica de usar renderizações baseadas em curvas (*line-based rendering*) para aplicações de internet que exigem eficiência, e também por razões estéticas. A riqueza estrutural de curvas é ilustrada pelo fato de ser impossível registrar uma nova imagem não-calibrada com relação a uma nuvem de pontos 3D por si só, ao passo que essa tarefa se torna possível com curvas 3D dada a maior estruturação geométrica (p. ex., reprojetoando-se as curvas 3D na nova imagem

e alinhando-as às curvas detectadas na imagem). Ademais, é bem sabido que representações baseadas em bordas representam eficiente o conteúdo de uma imagem, o que motiva uma reconstrução de curvas 3D eficiente para armazenar a informação geométrica mais relevante de uma cena 3D. Curvas também têm maior invariância do que pontos de interesse a mudanças de iluminação, são mais estáveis sob uma maior separação entre as imagens, e são mais precisamente localizadas na direção normal. Ademais, estrutura de curvas de bordas em uma imagem é correlacionada com propriedades das superfícies subjacentes: curvas estacionárias (como as de reflectância) são condições de contorno para reconstrução de superfícies, e a variação de contornos de oclusão sob diferentes pontos de vista indica a curvatura da superfície

Impacto tecnológico. curve-based multiview reconstruct has broad impacts in applications working with scenes that are largely textureless, as frequently occurring in indoor environments, *e.g.*, the Hololens project at Microsoft [13].

8 Metas

Elaboração de Publicações Como metas específica deste projeto, tem-se a elaboração de publicações (idealmente 1 por ano) nas principais revistas e conferências da área. Dentre as revistas visadas, temos: International Journal of Computer Vision, IEEE Pattern Analysis and Machine Intelligence, IEEE Transactions in Image Processing, Pattern Recognition, ACM Computing Surveys, e Journal of Mathematical Imaging and Vision. Também pretende-se enviar trabalhos completos (arigos) às mais importantes conferências na área, a saber: IEEE CVPR (Computer Vision and Pattern Recognition), IEEE ECCV (European Conference in Computer Vision), IEEE ICCV (International Conference in Computer Vision), ACM SIGGRAPH, e o IEEE SIBIGRAPI.

9 Metodologia

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

- **Auto-calibração trifocal.** Será estudado o trabalho do prof. Fabbri e seu aluno David de Pinho, de forma compreender as soluções iniciais. Inicialmente serão adquiridas e visualizadas as trajetórias do solver por continuação homotópica. Esta etapa envolve a compreensão de um algoritmo a partir de uma teoria clara de geometria algébrica de diferencial, a ser desenvolvida em conjunto com o matemático Peter J. Giblin da Universidade de Liverpool.
- **Estudo de Geometria Algébrica Numérica.** Pretende-se estudar a solução de sistemas de equações algébricas de maneira aprofundada, para auxiliar no item anterior. Esta etapa também contará com reuniões com o matemático Peter J. Giblin, dentre outros.
- **Validação e Experimentos de Larga escala.** Pretende-se realizar experimentos na reconstrução em larga escala de objetos industriais partir de milhares de fotografias ou quadros de vídeo. A validação será realizada através de comparação com objetos conhecidos *a priori* e impressos em impressoras 3D. O foco será encontrar objetos e cenas para os quais as técnicas do *solver* estudado sejam imprescindíveis.
- **Desenvolvimento de Software livre para Visão Computacional.** Pretende-se continuar o desenvolvimento de software livre, como uma forma de contribuir para a sociedade, angariar colaborações, garantir um controle de qualidade, e disseminar a pesquisa.
- Elaboração de Publicações como descrito nas metas.

10 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 anuais.

Etapas anuais do doutorado	2	3	4
Auto-calibração Trifocal	X		
Estudo de Geometria Algébrica Numérica	X		
Estudo de Mapas de Difusão e Teoria de Padrões	X	X	
Experimentos de Larga Escala		X	X
Escrita de Software Livre	X	X	X
Elaboração de Publicações	X	X	X

11 Conclusão

Este projeto propõe o desenvolvimento da tecnologia computacional de base necessária para efetivar a praticidade de um novo paradigma baseado em curvas e superfícies para reconstrução 3D e calibração, complementando as presentes técnicas baseadas em pontos, as quais estão sendo usadas em larga escala. A tecnologia proposta extenderá a capacidade das técnicas atuais, permitindo o processamento de uma maior variedade de objetos de interesse industrial, ao mesmo tempo impondo menos restrições na aquisição.

Referências

- [1] ARKit Team, “Understanding ARKit tracking and detection.” Apple, WWDC, 2018. [1](#)
- [2] R. Fabbri and B. B. Kimia, “Multiview differential geometry of curves,” *International Journal of Computer Vision*, vol. 117, pp. 1–23, 2016. [1](#)
- [3] R. Fabbri, *Multiview Differential Geometry in Application to Computer Vision*. Ph.D. dissertation, Division Of Engineering, Brown University, Providence, RI, 02912, July 2010. [1](#)
- [4] 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. [1](#)
- [5] R. Fabbri and B. B. Kimia, “High-order differential geometry of curves for multiview reconstruction and matching,” in *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pp. 645–660, Springer, 2005. [1](#)
- [6] 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, CA, USA), IEEE Computer Society Press, 2010. [1](#)
- [7] A. Usumezbas, R. Fabbri, and B. B. Kimia, “The surfacing of multiview 3d drawings via lofting and occlusion reasoning,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017. [1](#)
- [8] A. Usumezbas, R. Fabbri, and B. B. Kimia, “From multiview image curves to 3D drawings,” in *Proceedings of the European Conference in Computer Vision*, 2016. [1](#)
- [9] R. Fabbri and B. B. Kimia, “Multiview differential geometry of curves,” *International Journal of Computer Vision*, vol. 120, no. 3, pp. 324–346, 2016. [2](#)
- [10] R. Fabbri, *Multiview differential geometry in application to computer vision*. PhD thesis, Brown University, 2011. [2](#)
- [11] A. Usumezbas, R. Fabbri, and B. B. Kimia, “From multiview image curves to 3D drawings,” in *European Conference on Computer Vision*, pp. 70–87, Springer, 2016. [2](#)
- [12] R. Fabbri, T. Duff, H. Fan, M. H. Regan, D. da Costa de Pinho, E. Tsigaridas, C. Wrangler, J. D. Hauenstein, B. B. Kimia, A. Leykin, and T. Pajdla, “Trifocal relative pose from lines at points and its efficient solution,” *CoRR*, vol. abs/1903.09755, 2019. [3](#)

- [13] I. Nurutdinova and A. Fitzgibbon, “Towards pointless structure from motion: 3d reconstruction and camera parameters from general 3d curves,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2363–2371, 2015. 6