Servovisão Métrica Baseada em Primitivas: Novas Técnicas 3D e Suas Conexões 2D

Geraldo Silveira

Divisão de Sistemas Ciberfísicos, CTI, Campinas/SP, Brasil
(e-mail: Geraldo.Silveira@cti.gov.br*)*

Abstract: This paper investigates various feature-based metric visual servoing techniques under the teach-by-showing framework. Metric techniques include all classical 3D and 2D control approaches because they require or estimate some metric information, e.g., scene depth. Unlike standard solutions, this work defines the 3D estimation problem as a 2D-to-2D feature nonlinear optimization task, and also exploits the observability issue associated to monocular systems. Such formulation allows the development of a family of new feature-based 3D visual servoing strategies with varying degrees of computational complexity and of prior knowledge. Three new 3D techniques are then proposed and, as a key contribution, strong connections are found between them and the classical 2D methods. These lead to a unifying theory that paves the way for relevant future research on visual servoing.

Resumo: Este artigo investiga várias técnicas de servovisão métrica baseada em primitivas onde o equilíbrio é definido via uma imagem de referência. Técnicas métricas incluem todas as abordagens clássicas de servovisão 3D e 2D dado que essas requerem ou estimam alguma informação métrica, e.g., profundidade da cena. Diferentemente das soluções usuais, este trabalho define o problema de estimação 3D como uma tarefa de otimização não linear de primitivas 2D para 2D, e também explora as degenerações intrínsecas dos sistemas monoculares. Tal formulação permite o desenvolvimento de uma família de novas estratégias de servovisão 3D baseada em primitivas, com diferentes graus de complexidade computacional e de conhecimento prévio do sistema. Três novas técnicas 3D são então propostas e, como contribuição-chave, fortes conexões são encontradas entre elas e os métodos 2D clássicos. Essas conexões fornecem uma teoria unificada de servovisão com importantes desdobramentos para futuras pesquisas na área.

Keywords: Robotics; 3D visual servoing; 2D visual servoing; vision-based robot control.
Palavras-chave: Robótica; servovisão 3D; servovisão 2D; controle de robôs baseado em visão.

1. INTRODUÇÃO

Servovisão se refere ao controle de robôs a partir da realimentação de imagens. Uma tarefa típica consiste em estabilizá-lo em uma pose de referência a qual é especificada via uma imagem de referência, também chamada imagem desejada. *Soluções gerais* para essa tarefa são aquelas capazes de controlar todos os seis graus de liberdade do robô independentemente das características do objeto (e.g., formato e tamanho), do deslocamento da câmera entre imagens e da sua pose relativa ao objeto. Com poucas exceções, e.g., (Silveira, 2014b; Silveira et al., 2020), as abordagens gerais comprovadamente estabilizantes exigem ou estimam alguma informação métrica (Chaumette and Hutchinson, 2006). Isso obviamente ocorre em todas as técnicas de servovisão 3D (i.e., as baseadas em pose) dado que elas utilizam a pose da câmera para definir seus erros de controle. É importante ressaltar que inclusive as técnicas 2D clássicas (i.e., as baseadas em imagem) são também métricas dado que a profundidade aproximada do objeto à câmera é necessária em suas leis de controle. Em verdade, a esmagadora maioria dos trabalhos sobre controle de robôs baseado em visão consiste em *servovisão métrica*.

⋆ Este trabalho foi em parte financiado pelo Projeto InSAC (FAPESP–2014/50851-0 e CNPq–465755/2014-3).

Independentemente se informações métricas são utilizadas ou não, imagens são exploradas para controle e estimação de duas maneiras básicas. A primeira maneira explora diretamente o valor da intensidade dos pixels sem etapas intermediárias. Assim, elas fazem uso de dados originais e densos de imagem, o que permite a obtenção de elevados níveis de precisão e versatilidade. Por outro lado, seus algoritmos de tempo real dependem de métodos locais de otimização não linear, pois os algoritmos globais são usualmente muito custosos computacionalmente. A segunda maneira básica de explorar dados visuais para aqueles objetivos é descrever as imagens através de primitivas geométricas, e.g., pontos e retas, as quais são extraídas e associadas entre imagens. Logo, essa maneira depende da existência e da detecção de primitivas particulares nas imagens, de um procedimento de associação entre elas que é altamente propenso a erros, e de ajustes especiais de parâmetros para a execução bem-sucedida dessas etapas intermediárias. Em contrapartida, esses algoritmos possuem um domínio de convergência relativamente maior e contam com uma literatura disponível abundante. De fato, a vasta maioria das técnicas existentes de servovisão são *baseadas em primitivas*, particularmente em primitivas 2D (i.e., projeções de primitivas 3D nas imagens).

Especificamente para o objetivo de estimação 3D (i.e., da pose da câmera e da estrutura da cena) monocular a partir da associação de primitivas 2D para 2D, a grande maioria das técnicas existentes se baseiam em alguma estratégia de filtragem ou em métodos numéricos para solução de sistemas de equações. A técnica de filtragem clássica para aquele objetivo específico consiste nas baseadas ou derivadas do filtro de Kalman, e.g., filtro de Kalman estendido. Em relação aos métodos numéricos, a solução clássica passa primeiramente pela determinação da matriz essencial, que relaciona geometricamente duas imagens perspectivas calibradas. O algoritmo proposto por Nistér (2003) para este fim tem sido amplamente utilizado na área de visão computacional, e mostra-se funcionar para objetos planares ou não. Enfatiza-se que primitivas 3D (sejam elas medidas ou obtidas através de triangulação) não são aqui consideradas. Portanto, estratégias de estimação 3D a partir da associação de primitivas 3D para 2D, ou de primitivas 3D para 3D, estão fora do escopo do presente trabalho.

Este artigo endereça a ampla maioria das abordagens existentes de controle de robôs baseado em visão. De fato, trata-se da servovisão métrica geral baseada em primitivas 2D para a estabilização de robôs holonômicos monoculares onde o equilíbrio é definido via uma imagem de referência. Este trabalho é inspirado de Silveira (2014a), cujas técnicas de servovisão 3D são exclusivamente baseadas em intensidade dos pixels. A primeira diferença aqui consiste, portanto, na formulação ótima da estimação 3D baseada em primitivas 2D para 2D, sem etapas intermediárias de filtragem ou de determinação da matriz essencial, e nem de triangulação e consequente projeção de primitivas. A partir desta formulação e das degenerações intrínsecas dos sistemas monoculares, três novas técnicas de servovisão 3D baseada em primitivas são então propostas. Como contribuição-chave, fortes conexões são encontradas entre essas novas técnicas de servovisão 3D e as 2D clássicas. Essas conexões fornecem uma teoria unificada de servovisão com importantes desdobramentos para futuras pesquisas na área. Por exemplo, novos controladores servo visuais de chaveamento suave poderão ser investigados. Do ponto de vista de estimação, vislumbra-se a extensão da abordagem 2D ótima unificada usando intensidade de pixels e primitivas 2D proposta em Nogueira et al. (2020) para o presente caso de estimação 3D.

2. FUNDAMENTAÇÃO TEÓRICA

Esta seção define a notação e recorda modelos essenciais utilizados no artigo. Sejam $\mathbf{I}$ e $\mathbf{0}$ respectivamente a matriz identidade e a matriz composta de zeros, ambas de dimensões apropriadas. A norma Euclidiana, uma estimativa e uma versão transformada da variável $\boldsymbol{v}$ são escritas $\|\boldsymbol{v}\|$, $\widehat{\boldsymbol{v}}$ e $\boldsymbol{v}'$, respectivamente. Um asterisco sobrescrito, e.g., $\boldsymbol{v}^*$, indica que $\boldsymbol{v}$ é expresso em relação ao sistema de coordenadas de referência $\mathcal{F}^*$. As notações $\mathbf{A}(\boldsymbol{\nu})$ e $\mathrm{ziv}\big(\mathbf{A}(\boldsymbol{\nu})\big)$ representam, respectivamente, a matriz torção associada ao vetor $\boldsymbol{\nu} = [\boldsymbol{v}^\top, \boldsymbol{\omega}^\top]^\top$ e o seu mapeamento inverso, i.e.,

$$\mathbf{A}(\boldsymbol{\nu}) = \begin{bmatrix} [\boldsymbol{\omega}]_\times & \boldsymbol{v} \\ \mathbf{0} & 0 \end{bmatrix} \in \mathfrak{se}(3) \tag{1}$$

e

$$\mathrm{ziv}\big(\mathbf{A}(\boldsymbol{\nu})\big) = \boldsymbol{\nu} \in \mathbb{R}^6. \tag{2}$$

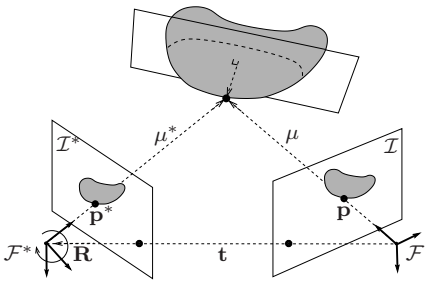

Figura 1. Geometria entre duas imagens de um objeto desconhecido.

2.1 Geometria Euclidiana entre Duas Imagens

Usando coordenadas homogêneas, a relação geral entre pixels correspondentes $\mathbf{p}_i \leftrightarrow \mathbf{p}_i^*$, $i = 1, 2, \ldots, p$, em duas imagens perspectivas calibradas é dada por

$$\mathbf{p}_i \propto [\ \mathbf{K}\ \ \mathbf{0}\ ]\ \mathbf{T}\ [\ (\mathbf{K}^{-1}\mathbf{p}_i^*)^\top\ \mu_i^*]^\top \in \mathbb{P}^2, \tag{3}$$

com

$$\mathbf{T} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \in \mathbb{SE}(3), \tag{4}$$

onde o símbolo "$\propto$" denota proporcionalidade, $\mathbf{K} \in \mathbb{R}^{3\times3}$ contem os parâmetros intrínsecos da câmera, $\mu_i^* \in \mathbb{R}$ é a inversa da profundidade (i.e., a proximidade) do ponto 3D projetado na imagem de referência $\mathcal{I}^*$ como $\mathbf{p}_i^*$, e $\mathbf{R} \in \mathbb{SO}(3)$ e $\mathbf{t} \in \mathbb{R}^3$ denotam, respectivamente, a rotação e a translação de $\mathcal{F}^*$ relativo ao sistema de coordenadas corrente $\mathcal{F}$ (vide Fig. 1).

Nota 1. *Independente das características do objeto observado (e.g., formato e tamanho) e do algoritmo de estimação, sistemas monoculares possuem problemas de observabilidade. Casos particulares de interesse à servovisão são:*

- *deslocamentos puramente rotacionais da câmera. Neste caso,* $\mathbf{t} = \mathbf{0}$*, o que sempre ocorre no equilíbrio; e*
- *objetos no infinito. Neste caso,* $\mu_i^* = 0$ *para todos os seus pontos.*

Ambos casos conduzem o segundo termo do desenvolvimento do lado direito de (3) *a*

$$\mu_i^*\,\mathbf{K}\,\mathbf{t} = \mathbf{0}, \tag{5}$$

$\forall \boldsymbol{\mu}^* = [\mu_1, \mu_2, \ldots, \mu_p]^\top$*. Em qualquer um deles, a translação da câmera e a estrutura da cena não são observáveis e, portanto, não podem ser estimadas com precisão.*

2.2 Sistema Robótico

Este trabalho considera um sistema robótico holonômico que evolui com velocidade instantânea $\boldsymbol{\nu} \in \mathbb{R}^6$ expressa em $\mathcal{F}$. Sua cinemática pode ser descrita por

$$\dot{\mathbf{T}} = -\mathbf{A}(\boldsymbol{\nu})\,\mathbf{T}, \tag{6}$$

onde $\mathbf{T} \in \mathbb{SE}(3)$ então descreve sua pose e pode ser localmente parametrizada através da algebra de Lie associada $\mathfrak{se}(3)$ (Varadarajan, 1974), i.e., $\mathbf{T} = \mathbf{T}(\boldsymbol{\nu})$. O mecanismo para passar informação de $\mathfrak{se}(3)$ para $\mathbb{SE}(3)$ é via o mapeamento exponencial:

$$\mathbf{T}(\boldsymbol{\nu}) = \exp\big(\mathbf{A}(\boldsymbol{\nu})\big), \tag{7}$$

cuja inversa em torno da origem de $\mathfrak{se}(3)$ e do elemento identidade de $\mathbb{SE}(3)$ é a função logaritmo:

$$\mathbf{A}(\boldsymbol{\nu}) = \log\big(\mathbf{T}(\boldsymbol{\nu})\big). \tag{8}$$

3. ARCABOUÇO DE ESTIMAÇÃO 3D ÓTIMA BASEADA EM PRIMITIVAS

A estimação 3D baseada em primitivas 2D pode ser efetuada de diferentes formas, e.g., métodos numéricos ou através de filtragem. Esta seção apresenta uma formulação ótima para esse problema, bem como suas soluções clássicas. Na sequência, considere que um conjunto suficientemente grande e não degenerado de primitivas 2D foi extraído e corretamente associado $\mathbf{s} \leftrightarrow \mathbf{s}^*$ (e.g., pontos e retas) entre as imagens corrente e a de referência.

3.1 Otimização Não Linear de Primitivas 2D para 2D

Este trabalho formula o problema de estimação 3D como uma tarefa de registro de primitivas 2D para 2D. Isso é principalmente motivado pela sua conexão com a servovisão, conforme será investigada mais profundamente nas seções posteriores. Essa tarefa de registro consiste em obter os parâmetros que melhor transformam as primitivas de referência $\mathbf{s}^* \in \mathbb{R}^n$ de tal forma que cada primitiva transformada $\mathbf{s}_j^{*\prime}$, $j = 1, 2, \ldots, \ell$, case o mais próximo possível da sua primitiva corrente associada $\mathbf{s}_j \in \mathbb{R}^n$. O conjunto de primitivas de referência transformadas pode ser escrito

$$\mathbf{s}^{*\prime}\big(\mathbf{x}(\boldsymbol{\tau})\big) = \begin{bmatrix} \mathbf{w}\big(\mathbf{x}(\boldsymbol{\tau}), \mathbf{s}_1^*\big) \\ \mathbf{w}\big(\mathbf{x}(\boldsymbol{\tau}), \mathbf{s}_2^*\big) \\ \vdots \\ \mathbf{w}\big(\mathbf{x}(\boldsymbol{\tau}), \mathbf{s}_\ell^*\big) \end{bmatrix} \in \mathbb{R}^n, \tag{9}$$

onde $\mathbf{w}(\cdot, \cdot)$ é uma função de transformação das primitivas. Se pontos 2D são utilizados, essa função pode ser definida a partir de (3) usando a variável $\mathbf{x} = \{\mathbf{T}, \boldsymbol{\mu}^*\}$ e sua respectiva parametrização $\boldsymbol{\tau} = [\boldsymbol{\nu}^\top \; \boldsymbol{\sigma}^\top]^\top \in \mathbb{R}^m$, i.e., $\mathbf{x} = \mathbf{x}(\boldsymbol{\tau}) = \{\mathbf{T}(\boldsymbol{\nu}), \boldsymbol{\mu}^*(\boldsymbol{\sigma})\}$. A parametrização da distribuição das proximidades $\boldsymbol{\mu}^* = \boldsymbol{\mu}^*(\boldsymbol{\sigma})$ é considerada a fim de permitir uma regularização da superfície. Uma propriedade importante dessa função é estabelecida abaixo.

Propriedade 1. *O conjunto de elementos $e = \{\mathbf{I}, \boldsymbol{\mu}^*\}$, $\forall \boldsymbol{\mu}^*$, define o mapeamento identidade para a função de transformação em* (9)*, i.e., $\forall \mathbf{s}_j^*$*

$$\mathbf{w}(e, \mathbf{s}_j^*) = \mathbf{s}_j^*. \tag{10}$$

Assim, um sistema de registro 3D de primitivas 2D para 2D pode ser formulado como o seguinte problema de otimização não linear:

$$\min_{\mathbf{x}(\boldsymbol{\tau})} \; \frac{1}{2} \left\| \mathbf{s} - \mathbf{s}^{*\prime}\big(\mathbf{x}(\boldsymbol{\tau})\big) \right\|^2, \tag{11}$$

o qual busca os parâmetros incrementais $\boldsymbol{\tau}$ para descrever a variável $\mathbf{x}$ que minimizam aquela norma.

Nota 2. *Outras funções custo podem ser consideradas em* (11)*, e.g., uma função robusta (Meer, 2004) pode ser aplicada para remover medidas aberrantes (e.g., oclusões parciais, falsas correspondências). No entanto, modificações neste aspecto não alteram esse arcabouço de estimação.*

3.2 Métodos Iterativos Clássicos

O problema de otimização não linear (11) pode ser eficientemente resolvido por métodos iterativos clássicos como, por exemplo, Gauss–Newton. Estes métodos baseiam-se em uma aproximação da função custo em série de Taylor. Assim, suas propriedades de convergência dependem fortemente de tal aproximação e das condições iniciais, também chamadas de estimativas iniciais. Estes métodos consistem nos seguintes passos (vide, e.g., (Luenberger, 1984) para maiores detalhes). Dada uma estimativa inicial $\widehat{\mathbf{x}}_0$ suficientemente próxima da solução, o incremento $\boldsymbol{\tau}_k \in \mathbb{R}^m$ nas variáveis de transformação é calculada na iteração k por

$$\boldsymbol{\tau}_k = -\alpha \, \mathbf{L}_{\boldsymbol{\tau}}^+ \left(\mathbf{s} - \mathbf{s}^{*\prime}(\widehat{\mathbf{x}}_k)\right), \tag{12}$$

com o tamanho do passo $\alpha > 0$ e, para os métodos clássicos,

$$\mathbf{L}_{\boldsymbol{\tau}}^+ = \widehat{\mathbf{H}}_{\boldsymbol{\tau}}^{-1} \mathbf{J}_{\boldsymbol{\tau}}^\top, \tag{13}$$

onde $\mathbf{J}_{\boldsymbol{\tau}} \in \mathbb{R}^{n \times m}$ denota a matriz Jacobianade (11) com relação a $\mathbf{x}$ em $\boldsymbol{\tau}$, e $\widehat{\mathbf{H}}_{\boldsymbol{\tau}} \in \mathbb{R}^{m \times m}$ é uma matriz positiva definida que aproxima [1] adequadamente a matriz Hessiana da função custo. O incremento (12) atualiza a variável $\widehat{\mathbf{x}}_k$ via

$$\widehat{\mathbf{x}}_{k+1} = \mathbf{x}(\boldsymbol{\tau}_k) \circ \widehat{\mathbf{x}}_k, \tag{14}$$

e o processo é iterado até a convergência. O símbolo "$\circ$" denota o operador de composição associado ao grupo envolvido. Vide (Varadarajan, 1974) para maiores detalhes. Na prática, a condição de convergência pode ser estabelecida quando o deslocamento incremental $\mathbf{x}(\boldsymbol{\tau}_k)$ for suficientemente próximo do elemento identidade do grupo envolvido, i.e., quando $\|\boldsymbol{\tau}_k\| < \epsilon_e$ para algum valor $\epsilon_e > 0$ suficientemente pequeno, ou quando um número máximo de iterações for efetuado.

A performance da estimação depende da escolha do método de otimização e das parametrizações das variáveis. No entanto, as ideias gerais deste artigo são independentes dessas definições, conforme mostrado abaixo.

4. NOVAS TÉCNICAS DE SERVOVISÃO 3D BASEADA EM PRIMITIVAS

Esta seção propõe novas técnicas de servovisão 3D em uma formulação unificada e baseada em primitivas. Elas são desenvolvidas a partir da exploração do arcabouço de estimação descrita na Seção 3 e dos problemas de observabilidade ressaltados na Nota 1. Na sequência, considere uma câmera montada na extremidade de um robô holonômico de seis graus de liberdade observando um objeto rígido e imóvel. Permita que os sinais de controle sejam as velocidades de translação e rotação da câmera, representadas por $\mathbf{v} \in \mathbb{R}^6$.

4.1 Servovisão 3D Ótima Baseada em Primitivas

Essa técnica de servovisão baseia-se na otimização simultânea da estrutura da cena $\boldsymbol{\mu}^*$ e da pose da câmera $\mathbf{T}$, os quais são então realimentados ao sistema de controle. Este procedimento de estimação pode ser computacionalmente custoso mas, por outro lado, o sistema de controle utiliza estimativas ótimas em todas as circunstâncias. Este método pode ser formulado como se segue.

[1] Exemplos dessa aproximação são: 1) para o método da maior descida, ela é simplesmente $\widehat{\mathbf{H}}_{\boldsymbol{\tau}} = \mathbf{I}$; 2) para o método Gauss–Newton, ela é dada por $\widehat{\mathbf{H}}_{\boldsymbol{\tau}} = \mathbf{J}_{\boldsymbol{\tau}}^\top \mathbf{J}_{\boldsymbol{\tau}}$, onde $[\cdot]^+$ então corresponde à pseudoinversa; 3) para o método Levenberg–Marquardt, ela é $\widehat{\mathbf{H}}_{\boldsymbol{\tau}} = \mathbf{J}_{\boldsymbol{\tau}}^\top \mathbf{J}_{\boldsymbol{\tau}} + \gamma \mathbf{D}$, onde $\gamma > 0$ e $\mathbf{D} \in \mathbb{R}^{m \times m}$ é uma matriz diagonal, e.g., $\mathbf{D} = \mathbf{I}$ ou $\mathbf{D} = \operatorname{diag}\big(\mathbf{J}_{\boldsymbol{\tau}}^\top \mathbf{J}_{\boldsymbol{\tau}}\big)$.

Estimação. Dado uma estimativa inicial $\widehat{\mathbf{x}}_0 = \{\widehat{\mathbf{T}}_0, \widehat{\boldsymbol{\mu}}_0^*\}$, os incrementos $\boldsymbol{\nu}_k \in \mathbb{R}^6$ e $\boldsymbol{\sigma}_k \in \mathbb{R}^{\dim(\boldsymbol{\sigma})}$ são computados na iteração k para a imagem corrente via

$$[\boldsymbol{\nu}_k^\top \ \boldsymbol{\sigma}_k^\top]^\top = -\alpha\,[\mathbf{L}_{\boldsymbol{\nu}} \ \mathbf{L}_{\boldsymbol{\sigma}}]^+ \big(\mathbf{s} - \mathbf{s}^{*\prime}(\widehat{\mathbf{x}}_k)\big), \tag{15}$$

onde $\mathbf{L}_{\boldsymbol{\nu}} \in \mathbb{R}^{n\times 6}$ e $\mathbf{L}_{\boldsymbol{\sigma}} \in \mathbb{R}^{n\times\dim(\boldsymbol{\sigma})}$. Os subscritos $\boldsymbol{\nu}$ e $\boldsymbol{\sigma}$ sob $\mathbf{L}$ são empregados apenas para especificar suas respectivas estimativas, não para denotar suas dependências. De fato, ambos $\mathbf{L}_{\boldsymbol{\nu}}$ e $\mathbf{L}_{\boldsymbol{\sigma}}$ dependem intrinsicamente de todas as variáveis envolvidas. Os incrementos (15) atualizam as variáveis via

$$\widehat{\mathbf{T}}_{k+1} = \mathbf{T}(\boldsymbol{\nu}_k) \circ \widehat{\mathbf{T}}_k, \tag{16}$$

$$\widehat{\boldsymbol{\mu}}_{k+1}^* = \boldsymbol{\mu}^*(\boldsymbol{\sigma}_k) \circ \widehat{\boldsymbol{\mu}}_k^*, \tag{17}$$

e o processo é iterado até a convergência, e.g., $\big\| [\boldsymbol{\nu}_k^\top \ \boldsymbol{\sigma}_k^\top] \big\| < \epsilon_e$. As estimativas ótimas $\mathbf{T}$ e $\boldsymbol{\mu}^*$ podem ser usadas como as iniciais na próxima imagem a ser processada.

Nota 3. *Este trabalho considera objetos rígidos. Se os parâmetros corretos da estrutura são obtidos, ou se a servovisão estiver suficientemente próxima do equilíbrio, a estimação desses parâmetros não faz sentido (vide Nota 1).*

Controle. No aspecto de controle, defina seu erro como

$$\mathbf{e}_{\mathrm{O3D}} = -\mathrm{ziv}\big(\log(\mathbf{T})\big). \tag{18}$$

Adicionalmente, defina a lei de controle desta técnica como

$$\mathbf{v} = -\lambda\,\mathbf{e}_{\mathrm{O3D}}, \tag{19}$$

com ganho de controle $\lambda > 0$. A convergência da tarefa servovisual pode ser estabelecida quando, e.g., $\|\mathbf{e}_{\mathrm{O3D}}\| < \epsilon_c$ para algum valor suficientemente pequeno $\epsilon_c > 0$. As propriedades de estabilidade e convergência desta técnica de servovisão são descritas abaixo.

Lema 1. *Considere o modelo de robô* (6)*. Seja* $\lambda > 0$*. Então, a lei de controle* (19) *assegura estabilidade exponencial local do equilíbrio* $\mathbf{e}_{O3D} = \mathbf{0}$*.*

De fato, note que esta técnica impõe ao Sistema (6) as velocidades (19) que são exatamente na mesma direção de $\boldsymbol{\nu} \in \mathbb{R}^6$ em (7). Ademais, observe que isso ocorre apenas em torno do equilíbrio $\mathbf{e}_{\mathrm{O3D}} = \mathbf{0}$ dado que a função logaritmo é definida apenas localmente. Uma propriedade interessante dessa lei de controle é o decaimento exponencial (embora local) de todas as componentes de velocidade em uma formulação concisa, i.e., não somente das rotacionais como nas abordagens 3D clássicas (vide, e.g., (Chaumette and Hutchinson, 2006, Eq. (18))). Essas propriedades estão ilustradas na Seção 6. Adicionalmente, ela fornece uma abordagem unificada para todas as técnicas propostas neste trabalho conforme mostrado abaixo.

4.2 Servovisão 3D Eficiente Baseada em Primitivas

Nesta técnica de servovisão métrica, a estratégia de estimação baseia-se na otimização de um subconjunto de variáveis. De fato, a ideia consiste em otimizar apenas os parâmetros relativos a pose da câmera $\mathbf{T}$. A estrutura do objeto $\boldsymbol{\mu}^*$ é fornecida pelo usuário e não é ajustada. Este esquema também é motivado pela questão da observabilidade (vide Nota 1) onde $\mu_i^*\,\mathbf{K}\,\mathbf{t} \longrightarrow \mathbf{0}, \forall i \in \{1, 2, \ldots, p\}$, se o robô estiver em torno do equilíbrio ou se o objeto estiver distante da câmera. Essa técnica é mais computacionalmente eficiente do que a anterior, mas sua performance depende das condições iniciais e/ou desses casos particulares. Este método pode ser descrito como se segue.

Estimação. Dado uma estimativa inicial $\widehat{\mathbf{x}}_0 = \{\widehat{\mathbf{T}}_0, \widehat{\boldsymbol{\mu}}_0^*\}$, a determinação do incremento $\boldsymbol{\nu}_k \in \mathbb{R}^6$ é realizada na iteração k para a imagem corrente via

$$\boldsymbol{\nu}_k = -\alpha\,\mathbf{L}_{\boldsymbol{\nu}}^+ \big(\mathbf{s} - \mathbf{s}^{*\prime}(\widehat{\mathbf{x}}_k)\big). \tag{20}$$

Este incremento atualiza a estimativa $\widehat{\mathbf{T}}_k$ via

$$\widehat{\mathbf{T}}_{k+1} = \mathbf{T}(\boldsymbol{\nu}_k) \circ \widehat{\mathbf{T}}_k, \tag{21}$$

e o processo é iterado até a convergência, e.g., $\|\boldsymbol{\nu}_k\| < \epsilon_e$. O vetor $\widehat{\boldsymbol{\mu}}_0^*$ não é ajustado, mas é necessário em (20).

Controle. Defina o respectivo erro de controle como

$$\mathbf{e}_{\mathrm{E3D}} = -\mathrm{ziv}\big(\log(\widehat{\mathbf{T}})\big), \tag{22}$$

onde $\widehat{\mathbf{T}} \in \mathbb{SE}(3)$ corresponde à estimativa obtida, a qual pode ser usada como inicial na próxima imagem a ser processada. Adicionalmente, a respectiva lei de controle pode ser definida como

$$\mathbf{v} = -\lambda\,\mathbf{e}_{\mathrm{E3D}}. \tag{23}$$

A convergência da servovisão é alcançada quando, e.g., $\|\mathbf{e}_{\mathrm{E3D}}\| < \epsilon_c$ para algum pequeno valor $\epsilon_c > 0$.

Corolário 2. *Assuma que a estrutura da cena fornecida pelo usuário* $\widehat{\boldsymbol{\mu}}_0^*$ *está correta ou que a Condição* (5) *é válida. Então, o Lema 1 também é válido para o equilíbrio* $\mathbf{e}_{E3D} = \mathbf{0}$ *usando a lei de controle* (23)*.*

De fato, note que se uma das premissas acima é satisfeita, então $\widehat{\mathbf{T}} = \mathbf{T}$ e (23) torna-se igual a (19). Se nenhum dos casos é estritamente verificado, então as propriedades de convergência desta estratégia de controle são obviamente em função de seus reais desvios.

4.3 Servovisão 3D Pobre Baseada em Primitivas

Esta seção apresenta a técnica 3D mais simples em termos de complexidade computacional. Seu baixo custo é devido à sua estratégia de estimação: apenas uma única iteração de (20) é executada e, como na técnica anterior, apenas em relação aos parâmetros associados com a pose da câmera. Os parâmetros relativos a estrutura são fornecidos pelo usuário e não são ajustados. Este método pode ser formalizado como se segue.

Estimação. Dado uma estimativa inicial $\widehat{\mathbf{x}}_0 = \{\widehat{\mathbf{T}}_0, \widehat{\boldsymbol{\mu}}_0^*\}$, a estimação do incremento $\boldsymbol{\nu}_0 \in \mathbb{R}^6$ é realizada apenas em uma única iteração para cada imagem capturada:

$$\boldsymbol{\nu}_0 = -\alpha\,\mathbf{L}_{\boldsymbol{\nu}}^+ \big(\mathbf{s} - \mathbf{s}^{*\prime}(\widehat{\mathbf{x}}_0)\big). \tag{24}$$

A atualização de $\widehat{\mathbf{T}}_0$ é também efetuada uma única vez para cada imagem capturada:

$$\widehat{\mathbf{T}}_1 = \mathbf{T}(\boldsymbol{\nu}_0) \circ \widehat{\mathbf{T}}_0, \tag{25}$$

a qual pode ser usada como estimativa inicial para a próxima imagem a ser processada. Enfatiza-se que a variável $\widehat{\boldsymbol{\mu}}_0^*$ não é ajustada, porém é necessária em (24).

Controle. Defina o erro de controle para o método pobre como

$$\mathbf{e}_{\mathrm{P3D}} = -\mathrm{ziv}\big(\log(\widehat{\mathbf{T}}_1)\big), \tag{26}$$

e a respectiva lei de controle como

$$\mathbf{v} = -\lambda\,\mathbf{e}_{\mathrm{P3D}}. \tag{27}$$

A convergência desta servovisão pode ser estabelecida quando, e.g., $\|\mathbf{e}_{\mathrm{P3D}}\| < \epsilon_c$ para algum valor suficientemente pequeno $\epsilon_c > 0$.

Nota 4. *A lei de controle* (27) *utiliza a estimativa* $\widehat{\mathbf{T}}_1$ *obtida a partir de um único passo de descida, e de uma dada estimativa inicial da estrutura* $\widehat{\boldsymbol{\mu}}_0^*$. *Enquanto que esta técnica é computacionalmente barata, suas propriedades de convergência (em particular, seu comportamento transitório) dependem fortemente das condições iniciais e/ou do conhecimento prévio do sistema.*

5. CONEXÕES COM AS TÉCNICAS DE SERVOVISÃO 2D CLÁSSICAS

Esta seção propõe e discute as fortes conexões entre as técnicas de servovisão 3D propostas e as 2D clássicas. Em particular, são mostradas as similaridades dessas últimas com a técnica 3D Pobre proposta na Seção 4.3.

5.1 Uma Nova Formulação da Servovisão 2D Clássica

As técnicas clássicas de servovisão 2D podem ser formuladas usando o mesmo procedimento das Seção 4. Para isso, um importante resultado deste artigo é apresentado.

Teorema 3. *Considere a técnica de servovisão 3D Pobre proposta na Seção 4.3. Seja* $\widehat{\mathbf{x}}_0 = \{\mathbf{I}, \widehat{\boldsymbol{\mu}}_0^*\}$ *a estimativa inicial para cada imagem a ser processada. Então, sua lei de controle* (27) *se escreve como nas técnicas 2D clássicas:*

$$\mathbf{v} = -\beta\, \mathbf{L}_{\boldsymbol{\nu}}^{+}\, (\mathbf{s} - \mathbf{s}^*), \tag{28}$$

com $\beta = \lambda\alpha > 0$.

Prova. A demonstração segue o mesmo procedimento usado para desenvolver a técnica de servovisão 3D Pobre.

Estimação. Dado a estimativa inicial particular $\widehat{\mathbf{x}}_0 = \{\mathbf{I}, \widehat{\boldsymbol{\mu}}_0^*\}$ e a Propriedade 1, tem-se que $\widehat{\mathbf{x}}_0 = e$. Então, o cálculo do incremento em uma única iteração para cada imagem capturada (24) se escreve

$$\boldsymbol{\nu}_e = -\alpha\, \mathbf{L}_{\boldsymbol{\nu}}^{+}\, \big(\mathbf{s} - \mathbf{s}^{*\prime}(e)\big), \tag{29}$$

$$= -\alpha\, \mathbf{L}_{\boldsymbol{\nu}}^{+}\, (\mathbf{s} - \mathbf{s}^*), \tag{30}$$

dado (10). Usando (30) e aquela estimativa particular, a atualização (25) se escreve

$$\widehat{\mathbf{T}}_e = \mathbf{T}(\boldsymbol{\nu}_e) \circ \mathbf{I}, \tag{31}$$

$$= \exp\big(\mathbf{A}(\boldsymbol{\nu}_e)\big), \tag{32}$$

dado (7). Obviamente, esta estimativa não será usada como inicial para a próxima imagem a ser processada. Conforme declarado, o elemento identidade continuará a ser aplicado para cada uma delas.

Controle. Usando (32), o erro de controle (26) torna-se

$$\mathbf{e}_{\mathrm{C2D}} = -\mathrm{ziv}\big(\log(\widehat{\mathbf{T}}_e)\big), \tag{33}$$

$$= -\boldsymbol{\nu}_e, \tag{34}$$

o qual é obtido através das aplicações inversas (8) e (2). O erro de controle (34) define um difeomorfismo entre $\mathbb{R}^6$ e $\mathbb{SE}(3)$ em torno da identidade. Finalmente, a lei de controle respectiva é dada por

$$\mathbf{v} = -\lambda\, \mathbf{e}_{\mathrm{C2D}} \tag{35}$$

$$= \lambda\, \boldsymbol{\nu}_e. \tag{36}$$

A prova é concluída com a substituição de (30) em (36). A lei de controle (28) corresponde às das técnicas de servovisão 2D clássicas. Vide, e.g., (Chaumette and Hutchinson, 2006; Malis, 2004), onde as provas de difeomorfismo local e de estabilidade assintótica local também podem ser encontradas. ■

5.2 Discussão

A partir da formulação acima, as técnicas clássicas de servovisão 2D podem ser vistas como a 3D Pobre proposta na Seção 4.3 com o elemento identidade como estimativa inicial da pose da câmera para cada imagem capturada. De fato, elas compartilham do mesmo arcabouço, que pode ser dividido em estimação e controle. A parte de estimação executa um procedimento de descida em um único passo no registro 3D de primitivas 2D para 2D, otimizando apenas os parâmetros associados à pose da câmera. (Para ambas as classes de técnicas, a estrutura do objeto pode, opcionalmente, ser estimada separadamente usando, por exemplo, observadores de estado. No entanto, isso também não altera o arcabouço compartilhado.) Por sua vez, a parte de controle aplica o resultado usando exatamente as mesmas fórmulas.

A formulação da Seção 5.1 também permite outra analogia entre a servovisão 2D e os métodos de otimização não linear já estabelecidos. À luz de (12), o escalar $\beta > 0$ em (28) não é um ganho de controle estrito no sentido de que não determina apenas a resposta dinâmica do sistema. De fato, ele também abrange o tamanho do passo que o robô deve se mover ao longo da direção de descida. Dada sua grande dependência da função custo, a direção calculada de descida raramente irá induzir a um movimento geodésico do robô. De forma a evitar seu movimento ineficiente (sem falar na resposta dinâmica), esse escalar deve ser pequeno, principalmente quando o sistema está longe do equilíbrio. Em contrapartida, esse escalar é um ganho de controle estrito para todas as técnicas propostas de servovisão 3D.

Outras implicações dessa nova interpretação da servovisão 2D clássica podem ser listadas. Do ponto de vista teórico, ele fornece uma estrutura que unifica as abordagens de servovisão 3D e 2D. Do ponto de vista prático, esta teoria unificadora abre caminho para o desenvolvimento, por exemplo, de novos controladores servo visuais de chaveamento suave. A maioria dos existentes controladores chaveados baseados em visão, e.g., (Chesi et al., 2004; Gans and Hutchinson, 2003), fazem suposições ou induzem vibrações no sistema devido às suas transições (quase-)abruptas entre as leis de controle.

6. EXEMPLOS ILUSTRATIVOS

Esta seção apresenta resultados típicos obtidos pelas técnicas de servovisão 3D propostas e a 2D clássica. Em todos os casos, o objetivo de controle é estabilizar o sistema tal que a descrição da imagem corrente (i.e., as primitivas correntes) convirja para aquela da imagem de referência (i.e., as primitivas de referência).

6.1 Cenário

O sistema robótico modela um manipulador clássico de seis graus de liberdade com uma câmera perspectiva montada em seu efetuador. As distâncias focais da câmera considerada são de 500 pixels, sem obliquidade, ponto principal como o centro da imagem e uma taxa de aquisição de 33Hz. Sem perda de generalidade, o cenário clássico é implementado. A câmera observa um objeto planar, que é descrito

por quatro pontos (primitivas consideradas) dispostos a 1m de distância da pose de referência. A translação inicial da câmera entre a pose corrente e a de referência é de $[0.73, -0.1, -0.19]^\top$m (norma: 0.76m), e a respectiva rotação initial é de $[-4.5, -45, 45]^\top$ na parametrização ângulo-eixo (norma: 63.8°). Esses valores correspondem a grandes deslocamentos iniciais. A Fig. 2 ilustra esse cenário.

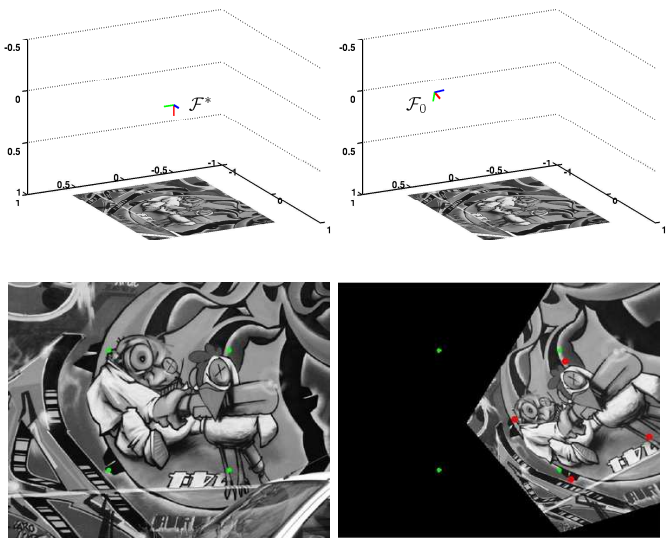

Figura 2. Cenário da tarefa de posicionamento. (Acima) Poses de referência e iniciais da câmera em relação ao objeto, respectivamente, e (abaixo) suas imagens correspondentes. Os quatro pontos verdes (resp. vermelhos) representam as primitivas de referência (resp. correntes) consideradas pela servovisão.

A parte das estimações é executada com $\alpha = 1$, o símbolo $(\cdot)^+$ representa a pseudoinversa, e as parametrizações e respectivos Jacobianos estão descritos em Silveira et al. (2008) (obviamente, até a transformação geométrica, visto que as intensidades dos pixels não são aqui consideradas). Quando aplicável, a estimativa inicial da pose é $[0.91, 0.02, -0.08]^\top$m (norma: 0.91m) para a translação e $[-5.4, -54, 54]^\top$ (norma: 76,56°) para a rotação na parametrização ângulo-eixo. Esses valores correspondem a erros na estimativa inicial de cerca de 20% na translação e na rotação. A estimativa inicial para o vetor normal do objeto é dada como $[0.26, -0.17, 0.95]^\top$, o que corresponde a um erro de estimativa de cerca de 20° nos parâmetros da estrutura. Finalmente, o ganho de controle é definido como $\lambda = 1$ (portanto, $\beta = 1$), e os critérios de convergência para a servovisão e a estimação (quando aplicável) usa $\epsilon_c = 10^{-5}$ e $\epsilon_e = 10^{-7}$, respectivamente.

6.2 Comparações

Os resultados obtidos para as técnicas propostas 3D Ótima (**O3D**), 3D Eficiente (**E3D**), 3D Pobre (**P3D**), bem como para a técnica 2D Clássica (**C2D**) são todos descritos e comparados na sequência.

Técnica proposta **O3D**. A Fig. 3 mostra a evolução dos sinais de controle e do deslocamento da câmera em direção à convergência. Note o interessante decaimento exponencial de todos os componentes da velocidade, conforme descrito na teoria. Além disso, ele é obtido desde o início da tarefa, apesar do grande deslocamento inicial e erros nas estimativas iniciais. Como resultado, o posicionamento é efetuado rapidamente, levando apenas 310 imagens.

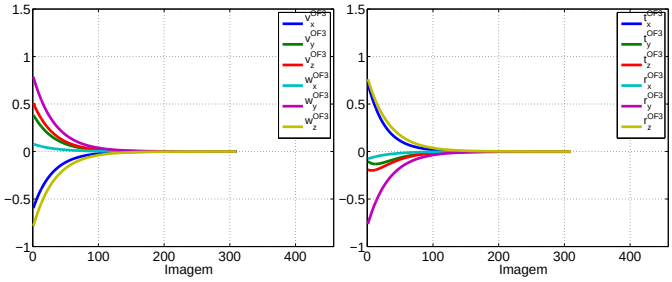

Figura 3. Sinais de controle (note o interessante decaimento exponencial) e o deslocamento da câmera no espaço Cartesiano, respectivamente, usando **O3D**.

Técnica proposta **E3D**. Nesta técnica, a estrutura inicial (e incorreta) do objeto não é otimizada. É, portanto, mais eficiente computacionalmente do que a estratégia anterior. Por outro lado, isso afeta adversamente as estimativas de pose e, como consequência, reduz a taxa de convergência. Conforme mostrado na Fig. 4, tal fato é mais pronunciado no início da tarefa, onde os sinais de controle foram ligeiramente acoplados. Em todo caso, o posicionamento é executado com sucesso após 444 imagens. Embora leve mais tempo para convergir do que **O3D**, este é um resultado interessante, pois também demonstra robustez a erros nas estimativas da estrutura do objeto.

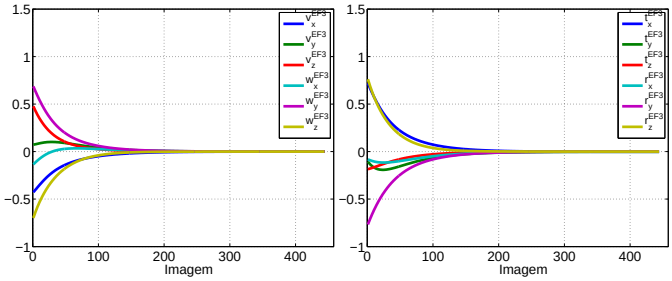

Figura 4. Sinais de controle (note os sutis acoplamentos) e o deslocamento da câmera no espaço Cartesiano, respectivamente, usando **E3D**.

Técnica proposta **P3D**. Como na **E3D**, esta técnica também utiliza as estimativas de profundidade iniciais (e incorretas) em todos os momentos. Sua principal diferença em relação à estratégia anterior consiste em estimar a pose da câmera a partir de um único passo de descida. Portanto, é altamente improvável que essa estimativa seja correta, especialmente para grandes deslocamentos de câmera. Além disso, tal estimativa pode mudar rapidamente de imagem para imagem, dependendo também da função de custo. A Fig. 5 ilustra este fenômeno, onde alguns picos nos sinais de controle podem ser observados no início da tarefa. Possíveis soluções para evitar esse comportamento indesejado consistem em reduzir o ganho de controle e/ou o tamanho do passo, mas ambas impactariam na taxa de convergência. Outra solução consiste em aplicar tal técnica somente quando o sistema estiver mais próximo do equilíbrio (ver Nota 4), possivelmente dentro de uma estratégia

de chaveamento. Em todo caso, o posicionamento ainda é concluído com sucesso também após 444 imagens, e com a menor complexidade computacional por imagem entre todas as técnicas 3D acima mencionadas.

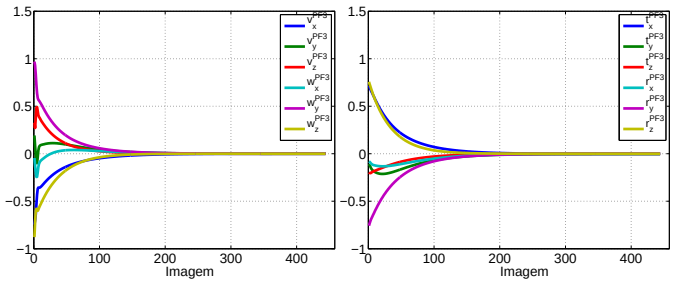

Figura 5. Sinais de controle (note os picos) e o deslocamento da câmera no espaço Cartesiano, respectivamente, usando **P3D**.

Técnica clássica **C2D**. Conforme discutido, a principal diferença entre esta técnica e a **P3D** proposta consiste na estimativa inicial da pose da câmera para cada imagem capturada. Conforme mostrado na Fig. 6, o fato de fornecer o elemento de identidade evita completamente os picos nos sinais de controle da Fig. 5. Por outro lado, isso ocorre às custas de um forte acoplamento entre eles, especialmente para grandes deslocamentos de câmera. Como consequência imediata, seu erro de controle é regulado apenas assintoticamente, levando cerca de 50% mais tempo do que **O3D**. Em todo caso, o posicionamento é realizado com sucesso após 458 imagens e, portanto, próxima à taxa de convergência da **P3D**. Finalmente, a Fig. 7 compara o decaimento da norma dos erros de controle para todas as técnicas de servovisão ora descritas.

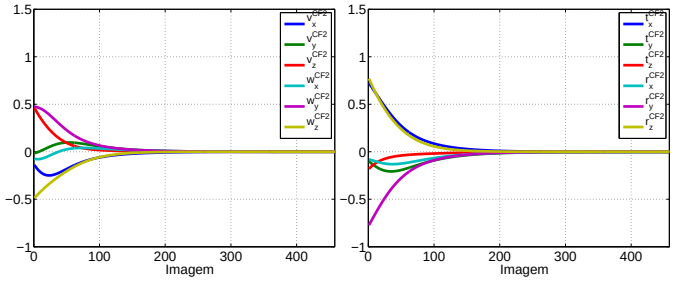

Figura 6. Sinais de controle (note os fortes acoplamentos) e o deslocamento da câmera no espaço Cartesiano, respectivamente, usando **C2D**.

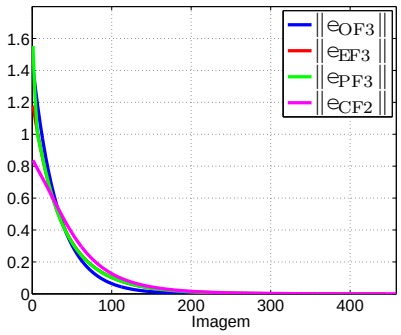

Figura 7. Evolução da norma dos erros de controle para todas as técnicas apresentadas no cenário considerado.

7. CONCLUSÕES

Este artigo aborda a servovisão métrica geral baseada em primitivas para a estabilização de robôs holonômicos monoculares onde o equilíbrio é definido via uma imagem de referência. O ponto de partida deste trabalho foi a formulação do problema de estimação 3D como um registro de primitivas 2D para 2D, sem etapas intermediárias de filtragem ou de determinação da matriz essencial, e nem de triangulação e consequente projeção de primitivas. A partir dela e das degenerações intrínsecas de sistemas monoculares, três novas técnicas de servovisão 3D baseadas em primitivas e de diferentes características foram aqui propostas. Ademais, o desenvolvimento dessas técnicas permitiu a descoberta de fortes conexões entre elas e os métodos 2D clássicos. Essa última contribuição é considerada chave, pois fornece uma estrutura que unifica as abordagens de servovisão 3D e 2D. Essa teoria unificada terá, certamente, importantes desdobramentos em futuras pesquisas na área. Como exemplo, vislumbra-se o desenvolvimento de novas estratégias de controle servo visual de chaveamento suave entre elas.

REFERÊNCIAS

Chaumette, F. and Hutchinson, S. (2006). Visual servo control part I: Basic approaches. *IEEE Robotics and Automation Magazine*, 82–90.

Chesi, G., Hashimoto, K., Prattichizzo, D., and Vicino, A. (2004). Keeping features in the field of view in eye-in-hand visual servoing: A switching approach. *IEEE Transactions on Robotics*, 20(5), 908–914.

Gans, N.R. and Hutchinson, S.A. (2003). An asymptotically stable switched system visual controller for eye in hand robots. In *Proc. IEEE/RSJ IROS*, 735–742.

Luenberger, D.G. (1984). *Linear and Nonlinear Programming*. Addison-Wesley.

Malis, E. (2004). Visual servoing invariant to changes in camera intrinsics parameters. *IEEE Transactions on Robotics an Automation*, 20(1), 72–81.

Meer, P. (2004). *Emerging Topics in Computer Vision*, chapter Robust techniques for computer vision, 107–190. Prentice Hall.

Nistér, D. (2003). An efficient solution to the five-point relative pose problem. In *Proc. IEEE CVPR*.

Nogueira, L., Paiva, E., and Silveira, G. (2020). Towards a unified approach to homography estimation using image features and pixel intensities. In *Proc. ICAS*, 110–115.

Silveira, G., Malis, E., and Rives, P. (2008). An efficient direct approach to visual SLAM. *IEEE Transactions on Robotics*, 24(5), 969–979.

Silveira, G., Mirisola, L., and Morin, P. (2020). Decoupled intensity-based nonmetric visual servo control. *IEEE Transactions on Control Systems Technology*, 28(2), 566–573.

Silveira, G. (2014a). On intensity-based 3D visual servoing. *Robotics and Autonomous Systems*, 62(11), 1636–1645.

Silveira, G. (2014b). On intensity-based nonmetric visual servoing. *IEEE Transactions on Robotics*, 30(4), 1019–1026.

Varadarajan, V. (1974). *Lie groups, Lie algebras, and their representations*. Prentice-Hall.