SERVOVISÃO 3D BASEADA EM INTENSIDADE

Geraldo Silveira*

*CTI Renato Archer, Divisão de Robótica e Visão Computacional, Campinas/SP, Brasil

Email: Geraldo.Silveira@cti.gov.br

Abstract— This article considers the problem of pose-based visual servoing whose equilibrium state is defined via a reference image. Differently from most solutions, this work directly exploits the pixel intensities without any feature extraction or matching. Intensity-based schemes provide for higher accuracy and versatility. Another central idea of this work is the exploitation of the observability issue associated to monocular systems, which always occurs around the equilibrium. This overall framework allows for developing new 3D visual servoing methods with varying degrees of computational complexity and prior knowledge. Comparative results of three new techniques are presented to assess their closed-loop performances. As a side result, they refute the common belief that correct camera calibration and pose recovery are crucial to the accuracy of 3D visual servoing techniques.

Keywords— Appearance-based methods, template-based methods, position-based visual servoing, visual servo control, vision-based control, pose reconstruction, visual localization, structure from motion.

Resumo— Este artigo considera o problema de servovisão baseado em pose cujo estado de equilíbrio é definido por uma imagem de referência. Diferentemente da maioria das soluções, este trabalho explora diretamente a intensidade dos pixels sem qualquer extração de primitivas geométricas. Esquemas baseados em intensidade fornecem maior precisão e versatilidade. Outra ideia central do trabalho é a exploração do problema de observabilidade associado a sistemas monoculares, o qual sempre ocorre em torno do equilíbrio. Este arcabouço permite o desenvolvimento de novos métodos de servovisão 3D com diferentes graus de complexidade computacional e de conhecimento prévio. Resultados comparativos de três novas técnicas são apresentados para avaliar seus desempenhos em malha fechada. Como contribuição adicional, esses resultados refutam a crença comum de que a correta reconstrução da pose e calibração da câmera são cruciais para a precisão das técnicas de servovisão 3D.

Palavras-chave— Métodos baseados em aparência, controle servo-visual, controle baseado em visão, reconstrução de pose, localização visual, estrutura a partir do movimento.

1 Introdução

Servovisão (do inglês, *visual servoing*) se refere ao controle dos movimentos de um robô através da realimentação de imagens. Uma aplicação típica consiste em estabilizá-lo em uma pose definida por meio de uma imagem de referência, também chamada de imagem desejada.

As soluções clássicas para esse problema são em geral classificadas em dois grupos: as baseadas em imagem ou em pose, também conhecidas simplesmente por 2D ou 3D, respectivamente. A primeira é assim chamada porque o erro de controle é definido no espaço imagem. Sua ideia básica é a construção de um erro de controle (pelo menos localmente) difeomórfico à pose da câmera. Como o controle é realizado na imagem, o objeto provavelmente permanece no campo de visão durante a execução da tarefa. Por outro lado, a trajetória da câmera no espaço Cartesiano não é a ideal (uma reta). No caso da servovisão 3D, o erro de controle é expresso naquele espaço. Assim, informações visuais são utilizadas para reconstruir explicitamente a pose da câmera, um processo também chamado localização ou odometria visual. Suas principais vantagens e desvantagens são o oposto da anterior. Como o controle é definido no espaço Cartesiano, a trajetória da câmera é teoricamente ideal, mas o objeto pode sair do campo de visão. Acredita-se também que a correta calibração da câmera, bem como sua localização visual, são cruciais para a precisão de técnicas 3D (Chaumette e Hutchinson, 2006)[pág. 90]. Uma das contribuições deste artigo é mostrar que essa conjectura não é necessariamente válida. As principais motivações para desenvolver as técnicas de servovisão 3D são três. Em primeiro lugar, elas teoricamente induzem a uma trajetória ótima da câmera. Em segundo, existem vários trabalhos que combinam técnicas 2D e 3D, também chamadas de híbridas. Por fim, diversas estratégias para controlar robôs não holonômicos ou subatuados baseiam-se mesmo no espaço Cartesiano (Morin, 2004).

Independente do espaço do erro de controle, técnicas de estimação baseada em visão podem geralmente ser classificados em dois grupos: baseadas em características ou em intensidade. O primeiro requer a extração de primitivas geométricas (e.g., pontos, linhas, etc.), bem como a sua associação, nas imagens. Diferentemente, as técnicas baseadas em intensidade exploram diretamente esses valores, sem etapas intermediárias, para estimar os parâmetros necessários. Portanto, eles fazem uso de dados brutos e densos de imagem, o que permite a obtenção de elevados níveis de precisão e versatilidade. Outra vantagem refere-se à possibilidade de assegurar robustez a mudanças arbitrárias de iluminação, mesmo em imagens omnidirecionais (Silveira, 2013).

Este artigo considera o problema de estabilização de robôs de seis graus de liberdade (g.d.l.) via técnicas de servovisão 3D baseadas em intensidade, dada uma imagem de referência. Com efeito, as intensidades dos pixels das imagens corrente e de referência são diretamente exploradas para construir o erro de controle, sem qualquer extração de características. Outra ideia central desse trabalho refere-se à exploração da questão da ob-

servabilidade intrinsecamente associada aos sistemas monoculares. Para estes sistemas, a translação entre imagens deve ser suficientemente grande em relação à profundidade da cena de modo a obter estimativas de pose precisas. Note que tal questão *sempre* ocorre quando a estabilização é suficientemente próxima ao equilíbrio.

Como primeira contribuição deste trabalho, é mostrado que técnicas de servovisão 3D podem ser realizadas com precisão, mesmo que a pose do robô não seja estimada corretamente, bem como ainda se a câmera e o robô são apenas grosseiramente calibrados. Técnicas baseadas em intensidade desempenham um papel importante neste aspecto. Em verdade, neste esquema a tarefa é concluída com sucesso se e somente se a imagem atual coincide com a de referência. Em seguida, mostra-se que este arcabouço permite também o desenvolvimento de uma família de novos métodos de servovisão 3D, com diferentes graus de complexidade computacional e de conhecimento prévio. Resultados comparativos de três novas técnicas são apresentados para avaliar seus desempenhos em malha fechada. Resultados experimentais são obtidos utilizando um braço robótico de seis g.d.l. com uma câmera convencional em seu efetuador.

2 Fundamentação Teórica

Esta seção apresenta modelos e métodos essenciais ao desenvolvimento do trabalho. Permita que a norma Euclidiana, a estimativa e a versão transformada de uma variável $\boldsymbol{v}$ seja escrita como $\|\boldsymbol{v}\|$, $\widehat{\boldsymbol{v}}$ e $\boldsymbol{v}'$, respectivamente. Um asterisco superescrito, e.g., $\boldsymbol{v}^*$, é usado para indicar que $\boldsymbol{v}$ é definido com relação ao sistema de coordenadas de referência $\mathcal{F}^*$. Dado um vetor real $\boldsymbol{\omega} = [\omega_1, \omega_2, \omega_3]^\top$, a notação $[\boldsymbol{\omega}]_\times$ representa sua matriz antisimétrica associada, i.e.

$$[\boldsymbol{\omega}]_\times = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix} \quad \in \mathfrak{so}(3). \quad (1)$$

Dado um vetor $\boldsymbol{\nu} = [\boldsymbol{v}^\top, \boldsymbol{\omega}^\top]^\top \in \mathbb{R}^6$, as notações $\mathbf{A}(\boldsymbol{\nu})$ e $\mathrm{ziv}\big(\mathbf{A}(\boldsymbol{\nu})\big)$ denotam, respectivamente, sua matriz torção associada e seu operador inverso, i.e.

$$\mathbf{A}(\boldsymbol{\nu}) = \begin{bmatrix} [\boldsymbol{\omega}]_\times & \boldsymbol{v} \\ \mathbf{0} & 0 \end{bmatrix} \quad \in \mathfrak{se}(3), \quad (2)$$

e

$$\mathrm{ziv}\big(\mathbf{A}(\boldsymbol{\nu})\big) = \boldsymbol{\nu} \quad \in \mathbb{R}^6. \quad (3)$$

2.1 Geometria Euclidiana entre duas imagens

A relação geral entre pixels correspondentes $\mathbf{p} \leftrightarrow \mathbf{p}^*$ em duas imagens perspectivas calibradas é dada por (Faugeras et al., 2001)

$$\mathbf{p} \propto \mathbf{K}\,\mathbf{R}\,\mathbf{K}^{-1}\mathbf{p}^* + \mu^*\,\mathbf{K}\,\mathbf{t} \quad \in \mathbb{P}^2, \quad (4)$$

onde o símbolo "$\propto$" denota proporcionalidade, $\mathbf{K} \in \mathbb{R}^{3\times3}$ contem os parâmetros intrínsecos da câmera, $\mathbf{R} \in \mathbb{SO}(3)$ e $\mathbf{t} \in \mathbb{R}^3$ denotam, respectivamente, a rotação e a translação do sistema de referência $\mathcal{F}^*$ relativo ao corrente $\mathcal{F}$, e $\mu^* \in \mathbb{R}$ é inversamente proporcional à profundidade do ponto 3D projetado em $\mathbf{p}^*$. Vide Fig. 1.

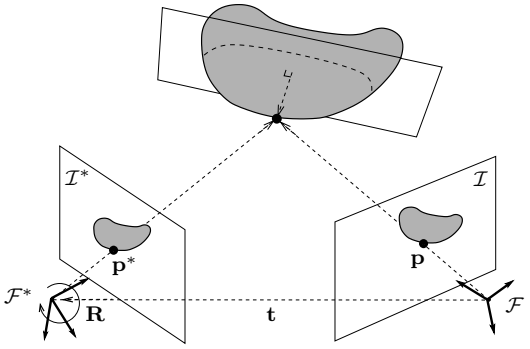

Figura 1: Geometria Euclidiana entre duas imagens.

Nota 2.1 (Observabilidade) *Independente do formato do objeto e do algoritmo de estimação, sistemas monoculares possuem problemas de observabilidade. Casos de interesse à servovisão são:*

- *deslocamentos puramente rotacionais entre a aquisição de duas imagens. Neste caso,* $\mathbf{t} = \mathbf{0}$*, o que sempre ocorre no equilíbrio;*
- *objetos no infinito. Neste caso, que é dual do anterior,* $\mu^* = 0$ *para todos os seus pontos.*

Ambos casos conduzem o segundo termo de (4) *a*

$$\mu^*\,\mathbf{K}\,\mathbf{t} = \mathbf{0}, \quad (5)$$

e todos os pixels em correspondência são relacionados somente por $\mathbf{p} \propto \mathbf{K}\,\mathbf{R}\,\mathbf{K}^{-1}\mathbf{p}^*$*. Em qualquer um desses casos, ambas a translação e a estrutura da cena não são observáveis e, portanto, não podem ser estimadas com precisão.*

A relação geral (4) pode ser escrita como

$$\mathbf{p} \propto [\,\mathbf{K}\;\;\mathbf{0}\,]\,\mathbf{T}\,[\,(\mathbf{K}^{-1}\mathbf{p}^*)^\top\,\mu^*]^\top, \quad (6)$$

com

$$\mathbf{T} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \quad \in \mathbb{SE}(3). \quad (7)$$

A matriz $\mathbf{T}$ contem o deslocamento da câmera, enquanto que o vetor $\boldsymbol{\mu}^* = [\mu_1^*, \mu_2^*, \ldots, \mu_n^*]^\top$ descreve a estrutura do objeto (descrito por n pixels).

2.2 Arcabouço básico de estimação

O arcabouço básico de estimação baseada em intensidade é o registro direto de imagens. Esse registro consiste em obter os parâmetros que melhor transformam a imagem corrente de tal forma que cada intensidade de pixel $\mathcal{I}(\mathbf{p})$ case o mais próximo possível da sua correspondente na imagem de referência $\mathcal{I}^*(\mathbf{p}^*)$. Dessa forma, um modelo de transformação $\mathcal{I}'(\cdot)$ é necessário. Por simplicidade, considere o modelo puramente geométrico

$$\mathcal{I}'(\mathbf{x}(\mathbf{z}), \mathbf{p}^*) = \mathcal{I}\big(\mathbf{w}(\mathbf{x}(\mathbf{z}), \mathbf{p}^*)\big) \quad \geq 0, \quad (8)$$

onde o operador $\mathbf{w}\colon \mathbb{SE}(3)\times\mathbb{R}^n\times\mathbb{P}^2 \to \mathbb{P}^2$ pode ser definido a partir de (6), com $\mathbf{x} = \{\mathbf{T}, \boldsymbol{\mu}^*\}$ e sua respectiva parametrização $\mathbf{z} = [\boldsymbol{\nu}^\top \boldsymbol{\sigma}^\top]^\top \in \mathbb{R}^m$, i.e., $\mathbf{x} = \mathbf{x}(\mathbf{z}) = \{\mathbf{T}(\boldsymbol{\nu}), \boldsymbol{\mu}^*(\boldsymbol{\sigma})\}$.

Um sistema de registro direto de imagens calibradas pode então ser formulado como o problema de otimização não linear (Silveira et al., 2008):

$$\min_{\mathbf{z}\in\mathbb{R}^m} \frac{1}{2}\sum_{i=1}^{n}\left[\mathcal{I}'\big(\mathbf{x}(\mathbf{z}), \mathbf{p}_i^*\big) - \mathcal{I}^*(\mathbf{p}_i^*)\right]^2, \quad (9)$$

o qual busca os parâmetros $\mathbf{z}$ para descrever $\mathbf{x}$ que minimizam as diferenças de intensidade

$$\mathbf{d}\big(\mathbf{x}(\mathbf{z})\big) = \begin{bmatrix} \mathcal{I}'\big(\mathbf{x}(\mathbf{z}), \mathbf{p}_1^*\big) - \mathcal{I}^*(\mathbf{p}_1^*) \\ \mathcal{I}'\big(\mathbf{x}(\mathbf{z}), \mathbf{p}_2^*\big) - \mathcal{I}^*(\mathbf{p}_2^*) \\ \vdots \\ \mathcal{I}'\big(\mathbf{x}(\mathbf{z}), \mathbf{p}_n^*\big) - \mathcal{I}^*(\mathbf{p}_n^*) \end{bmatrix} \in \mathbb{R}^n. \quad (10)$$

Outras funções custo podem ser consideradas em (9), por exemplo, uma função robusta (Huber, 1981), se aquele vetor contem medidas aberrantes, e.g., oclusões. No entanto, modificações neste aspecto não alteram o arcabouço básico.

O problema de otimização não linear (9) 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. 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 $\mathbf{z}_k \in \mathbb{R}^m$ nas variáveis de transformação é calculada na iteração k por

$$\mathbf{z}_k = -\alpha \mathbf{L}_{\mathbf{x}}^{+} \mathbf{d}(\widehat{\mathbf{x}}_k), \quad (11)$$

com $\alpha > 0$ e, para os métodos clássicos,

$$\mathbf{L}_{\mathbf{x}}^{+} = \widehat{\mathbf{H}}_{\mathbf{x}}^{-1} \mathbf{J}_{\mathbf{x}}^\top, \quad (12)$$

onde $\mathbf{J}_{\mathbf{x}} \in \mathbb{R}^{n\times m}$ denota a matriz Jacobiana de (10) com relação a $\mathbf{x}$ em $\mathbf{z}$, e $\widehat{\mathbf{H}}_{\mathbf{x}} \in \mathbb{R}^{m\times m}$ é uma matriz positiva definida que aproxima[1] adequadamente a matriz Hessiana da função custo. O incremento (11) atualiza a variável $\widehat{\mathbf{x}}_k$ via

$$\widehat{\mathbf{x}}_{k+1} = \mathbf{x}(\mathbf{z}_k) \circ \widehat{\mathbf{x}}_k, \quad (13)$$

o que é iterado até a convergência. O símbolo "$\circ$" denota o operador de composição associado ao grupo envolvido. Vide (Warner, 1987) para maiores detalhes. Na prática, a convergência pode ser obtida quando o deslocamento incremental $\mathbf{x}(\mathbf{z}_k)$ for suficientemente próximo do elemento identidade do grupo envolvido, i.e., quando $\|\mathbf{z}_k\| < \epsilon_e$, para algum valor $\epsilon_e > 0$ suficientemente pequeno.

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

Nota 2.2 (Método de otimização) *Embora cada método de otimização impacte de forma diferente no desempenho global da servovisão, as ideias gerais desse artigo são independentes da aproximação para calcular o incremento* (11).

3 Técnicas Propostas

Esta seção descreve novas técnicas de servovisão 3D a partir das intensidades dos pixels. De fato, o sistema de estimação é construído utilizando o vetor de diferenças de intensidade $\mathbf{d} \in \mathbb{R}^n$ (vide Seção 2.2), e sua estrutura é explorada para a construção do erro de pose. Estes novos métodos são organizados hierarquicamente em termos de custo computacional e de conhecimento prévio. Na sequência, considere uma câmera montada na extremidade de um robô de seis graus de liberdade observando um objeto imóvel, rígido e de forma desconhecida. Permita que os sinais de controle sejam as velocidades de translação e rotação da câmera, representadas por $\mathbf{v} \in \mathbb{R}^6$.

3.1 Método Pobre

Este primeiro método de servovisão 3D baseada em intensidade é o mais simples em termos de custo computacional. Seu baixo custo é devido tanto à sua estratégia de estimação quanto ao número de parâmetros estimados. Quanto à estratégia de estimação, a ideia é realizar apenas uma iteração de (11) dado que o cálculo da direção de descida pode ser custoso. Quanto ao número de parâmetros, a estratégia é aplicada apenas para obter o deslocamento da câmera $\mathbf{T}$. Os parâmetros relacionados com a estrutura $\boldsymbol{\mu}^*$ é fornecido pelo usuário. Este esquema é parcialmente motivado pela questão de observabilidade (vide Nota 2.1) que, em todo do equilíbrio, a estrutura do objeto não pode ser obtida com precisão.

Este método pode ser formalizado como se segue. 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 realizado apenas uma vez para cada imagem capturada:

$$\boldsymbol{\nu}_0 = -\alpha \mathbf{L}_{\mathbf{T}}^{+} \mathbf{d}(\widehat{\mathbf{x}}_0), \quad (14)$$

onde $\mathbf{L}_{\mathbf{T}} \in \mathbb{R}^{n\times 6}$, e n é o número de pixels.

Nota 3.1 (Integração) *As velocidades* (14) *já poderiam ser as entradas de controle, i.e.,* $\mathbf{v} = \boldsymbol{\nu}_0$. *Contudo, durante a tarefa (ou dado uma estimativa* $\widehat{\mathbf{T}}_0 \neq \mathbf{I}$*), a integração de movimento abaixo conduz a uma solução mais próxima da correta.*

O incremento (14) atualiza $\widehat{\mathbf{T}}_0$ via

$$\widehat{\mathbf{T}}_1 = \mathbf{T}(\boldsymbol{\nu}_0) \circ \widehat{\mathbf{T}}_0, \quad (15)$$

a qual pode ser utilizada como estimativa inicial na próxima imagem. Como informado, a variável $\widehat{\boldsymbol{\mu}}_0^*$ é fornecida pelo usuário e não é ajustada. Ela é utilizada apenas para o cálculo de (14).

No aspecto de controle, o erro no método pobre $\mathbf{e}_{\mathrm{MP}} \in \mathbb{R}^6$ pode ser definida como

$$\mathbf{e}_{\mathrm{MP}} = -\mathrm{ziv}\big(\log(\widehat{\mathbf{T}}_1)\big), \tag{16}$$

onde o operador $\mathrm{ziv}\colon \mathfrak{se}(3) \to \mathbb{R}^6$ é definido em (3), e a função $\log\colon \mathbb{SE}(3) \to \mathfrak{se}(3)$ representa o logaritmo matricial. Finalmente, a lei de controle respectiva pode ser definida como

$$\mathbf{v} = -\lambda\, \mathbf{e}_{\mathrm{MP}}, \tag{17}$$

com ganho de controle $\lambda > 0$. A convergência da tarefa pode ser estabelecida quando, e.g., $\|\mathbf{e}_{\mathrm{MP}}\| < \epsilon_c$, para $\epsilon_c > 0$ suficientemente pequeno.

3.2 Otimização Parcial

Nesta técnica servovisão 3D baseada em intensidade, a estratégia de estimação baseia-se na otimização de um subconjunto do variáveis. De fato, a ideia consiste em otimizar apenas os parâmetros relacionados com o deslocamento da câmera $\mathbf{T}$. A estrutura do objeto $\boldsymbol{\mu}^*$ é fornecido pelo usuário e não é ajustado. Este esquema também é parcialmente motivado pela questão da observabilidade (ver Nota 2.1). Em troca de ser computacionalmente mais caro do que o método pobre, seu domínio e taxa de convergência são melhorados. Esses são obtidos dado que ao menos quantidades subótimas são estimadas. Na verdade, se o modelo de objeto correto for fornecido, então a pose obtida também é a correta.

Este método pode ser descrito como se segue. Dado uma estimativa inicial $\widehat{\mathbf{x}}_0 = \{\widehat{\mathbf{T}}_0, \widehat{\boldsymbol{\mu}}_0^*\}$, a determinação do incremento $\boldsymbol{\nu}_0 \in \mathbb{R}^6$ é realizada para a imagem corrente na iteração k via

$$\boldsymbol{\nu}_k = -\alpha \mathbf{L}_{\mathbf{T}}^{+} \mathbf{d}(\widehat{\mathbf{x}}_k), \tag{18}$$

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

$$\widehat{\mathbf{T}}_{k+1} = \mathbf{T}(\boldsymbol{\nu}_k) \circ \widehat{\mathbf{T}}_k, \tag{19}$$

o qual é iterado até a convergência, e.g., $\|\boldsymbol{\nu}_k\| < \epsilon_e$, com $\epsilon_e > 0$. Uma vez mais, $\widehat{\boldsymbol{\mu}}_0^*$ é fornecido pelo usuário e não é ajustado.

No aspecto de controle, permita-nos definir o erro de controle respectivo $\mathbf{e}_{\mathrm{OP}} \in \mathbb{R}^6$ como

$$\mathbf{e}_{\mathrm{OP}} = -\mathrm{ziv}\big(\log(\widehat{\mathbf{T}})\big), \tag{20}$$

onde $\widehat{\mathbf{T}} \in \mathbb{SE}(3)$ é o deslocamento estimado, o qual pode ser usado como estimativa inicial na próxima imagem. Finalmente, a lei de controle respectiva pode ser definida como

$$\mathbf{v} = -\lambda\, \mathbf{e}_{\mathrm{OP}}, \tag{21}$$

com $\lambda > 0$. Convergência da tarefa é obtida quando, e.g., $\|\mathbf{e}_{\mathrm{OP}}\| < \epsilon_c$, com $\epsilon_c > 0$.

3.3 Otimização Completa

Esta técnica de servovisão 3D baseada na intensidade baseia-se em uma otimização simultânea de todos os parâmetros utilizados, incluindo, evidentemente, aqueles relacionados ao deslocamento da câmera $\mathbf{T}$ e da estrutura da cena $\boldsymbol{\mu}^*$. Esta técnica é mais custosa computacionalmente do que os métodos anteriores mas, por outro lado, tem o maior domínio e taxa de convergência entre eles. Com efeito, o sistema de controle utiliza estimativas ótimas em todas as circunstâncias. É, portanto, o método de escolha quando nenhum conhecimento prévio do sistema está disponível.

Este método pode ser formulado como se segue. Dado uma estimativa inicial $\widehat{\mathbf{x}}_0 = \{\widehat{\mathbf{T}}_0, \widehat{\boldsymbol{\mu}}_0^*\}$, a obtenção dos incrementos $\boldsymbol{\nu}_k \in \mathbb{R}^6$ e $\boldsymbol{\sigma}_k \in \mathbb{R}^{\dim(\boldsymbol{\sigma})}$ para a imagem corrente na iteração k escreve

$$[\boldsymbol{\nu}_k^\top \ \boldsymbol{\sigma}_k^\top]^\top = -\alpha [\mathbf{L}_{\mathbf{T}} \ \mathbf{L}_{\boldsymbol{\mu}}]^{+} \mathbf{d}(\widehat{\mathbf{x}}_k), \tag{22}$$

onde $\mathbf{L}_{\boldsymbol{\mu}} \in \mathbb{R}^{n \times \dim(\boldsymbol{\sigma})}$. Esses incrementos atualizam as variáveis via

$$\widehat{\mathbf{T}}_{k+1} = \mathbf{T}(\boldsymbol{\nu}_k) \circ \widehat{\mathbf{T}}_k, \tag{23}$$

$$\widehat{\boldsymbol{\mu}}_{k+1}^* = \boldsymbol{\mu}^*(\boldsymbol{\sigma}_k) \circ \widehat{\boldsymbol{\mu}}_k^*, \tag{24}$$

que são iterados até a convergência, e.g., $\| [\boldsymbol{\nu}_k^\top \ \boldsymbol{\sigma}_k^\top] \| < \epsilon_e$, com $\epsilon_e > 0$.

Nota 3.2 *Este trabalho considera objetos rígidos. Após a obtenção da estrutura correta $\boldsymbol{\mu}^*$ durante a tarefa, ou se a estimativa inicial é suficientemente próxima da solução, o esquema de estimação na OP deve ser usado em seu lugar. A estimação dos parâmetros da estrutura (e, portanto, de $\boldsymbol{\sigma}$) é praticamente inútil nesses casos.*

No aspecto de controle, o erro respectivo poderia ser definido de forma similar aos métodos precedentes. Entretanto, dado que o deslocamento correto é obtido, não existe razão para aplicar mapeamentos locais (o logaritmo matricial). Usando (7), o erro de controle pode ser definido a partir de $\widehat{\mathbf{T}} = \widehat{\mathbf{T}}(\widehat{\mathbf{R}}, \widehat{\mathbf{t}})$ como:

$$\mathbf{e}_{\mathrm{OC}} = -[\widehat{\mathbf{t}}^\top \ \widehat{\theta}\widehat{\mathbf{u}}^\top]^\top, \tag{25}$$

onde $\widehat{\theta}$ e $\widehat{\mathbf{u}}$ representam a parametrização ângulo-eixo de $\widehat{\mathbf{R}} \in \mathbb{SO}(3)$. Finalmente, a lei de controle neste arcabouço ótimo pode ser definido como:

$$\mathbf{v} = -\lambda\, \mathbf{e}_{\mathrm{OC}}, \tag{26}$$

com $\lambda > 0$. A convergência da tarefa servovisual é obtida quando, e.g., $\|\mathbf{e}_{\mathrm{OC}}\| < \epsilon_c$, com $\epsilon_c > 0$. A análise de estabilidade do sistema em malha fechada é imediata, e o equilíbrio $\mathbf{e}_{\mathrm{OC}} = \mathbf{0}$ é provado localmente assintoticamente estável, vide, e.g., (Chaumette e Hutchinson, 2006).

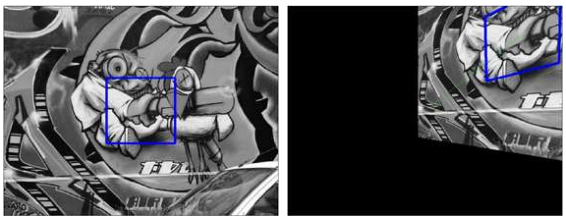

Figura 2: Servovisão 3D baseada em intensidade usando uma câmera não calibrada (erros de 10%), estimativas iniciais grosseiras da estrutura da cena (erros de 22°), e sob deslocamentos iniciais da câmera relativamente grandes (49%) com respeito às suas profundidades. (Esquerda) Imagem de referência. (Direita) Imagem inicial. Todos os pixels dentro do quadrado azul são explorados.

4 Resultados Experimentais

Esta seção apresenta resultados experimentais obtidos pelas três novas técnicas de servovisão 3D baseadas na intensidade propostos na Seção 3, ou seja, utilizando o método pobre (MP), via otimização parcial (OP) e pela otimização completa (OC). Em todos os casos, o objetivo de controle é estabilizar um robô de seis graus de liberdade dotado de uma câmera em sua extremidade, de tal forma que a imagem corrente do objeto coincida com a sua imagem capturada na pose de referência. Sem perda de generalidade, os resultados são obtidos usando um objeto planar por simplicidade. O arcabouço de estimação baseado em intensidade descrito em (Silveira et al., 2008) é aplicado para obter os parâmetros necessários nas leis de controle. Simplificações foram, obviamente, efetuadas naquele arcabouço de modo a satisfazer cada uma das estratégias propostas.

4.1 Dados sintéticos

Uma taxa de amostragem de 30ms foi utilizada em todos os experimentos com dados simulados. Todas as intensidades dentro de uma região de 200×200 pixels são exploradas. O critério de convergência para o controle e para a estimação utiliza $\epsilon_c = 10^{-5}$ e $\epsilon_e = 10^{-7}$ (quando aplicável), respectivamente. O ganho de controle utilizado é $\lambda = 1$. O deslocamento da câmera entre o frame de referência e o inicial é relativamente grande com relação às profundidades da cena. De fato, o objeto está a 1m da pose de referência, e foi imposta uma translação de $[0.28, 0.4, 0]$m (norma: 0.49m) e uma rotação de $-45°$ em torno do eixo $\vec{y}$, ambas relativas ao frame de referência. Assim, ele compreende uma translação de 49% das profundidades. A pose de referência corresponde a uma câmera observando um plano fronto-paralelo, i.e., com um vetor normal $\mathbf{n}^* = [0, 0, 1]^\top$. Não obstante, é fornecido para os algoritmos uma estimativa inicial daquele vetor de $\mathbf{n}^*\|\mathbf{n}^*\| = [0.4, -0.1, 1]^\top$. Isso corresponde a um erro inicial de aproximadamente 22.41°. Em adição, os parâmetros da câmera utilizados possuem um erro de 10%. Este cenário está mostrado na Fig. 2. A despeito de todas essas pertubações, a servovisão 3D baseada em intensidade converge com sucesso em todas as técnicas propostas com $\lambda = 1$, exceto para o método MP, mas este converge com $\lambda = 0.1$. A norma dos erros Cartesianos finais são menores que 0.1mm e 0.1°. O comportamento de cada técnica está mostrado na Fig. 3, e são descritos abaixo.

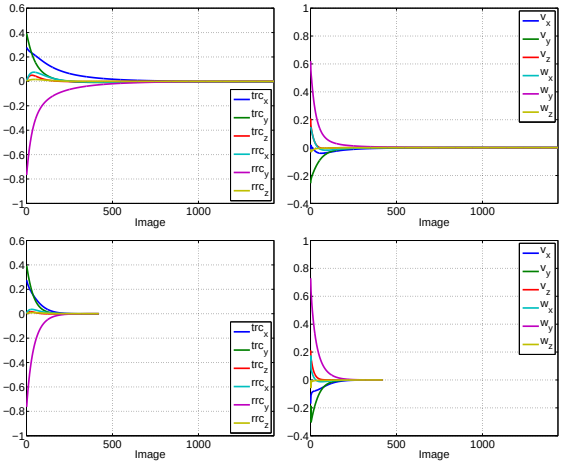

Figura 3: Evolução da câmera no espaço Cartesiano (esquerda) e os sinais de controle (direita) para os métodos OP e OC, usando $\lambda = 1$, sob um grande deslocamento inicial da câmera e vários tipos de perturbações (vide Fig. 2). O método MP não convergiu, mas converge se $\lambda = 0.1$ (não mostrado). Ambas as técnicas OP e OC convergem neste cenário. A última converge mais rápido, na imagem #421.

Método Pobre: Neste cenário de grandes perturbações (estimativas iniciais grosseiras, grandes deslocamentos iniciais), o método MP falha usando $\lambda = 1$, mas converge para $\lambda = 0.1$. No último caso são necessárias 14.407 imagens para realizar a tarefa de servovisão. Além disso, as entradas de controle não são completamente suaves no início da tarefa. Esses resultados sugerem que este método requer a aplicação de um ganho de controle variável de forma a melhorar seu desempenho. Em todo caso, seus requisitos computacionais são extremamente reduzidos dado que um único passo é aplicado na otimização.

Otimização Parcial: Para essa técnica, a tarefa converge com sucesso mesmo para o ganho de controle $\lambda = 1$. Neste caso, 1.439 imagens foram necessárias à estabilização. Ele também converge para $\lambda = 0.1$, e com sinais de controle suaves. Contudo, utilizando esse pequeno ganho de controle em tal cenário de grandes perturbações ele ainda necessita de diversas imagens, similarmente à técnica anterior. A complexidade computacional é maior que o método MP dado que mais iterações são efetuadas na otimização.

Otimização Completa: Utilizando o método OC, a estabilização é realizada com sucesso para ambos ganhos de controle $\lambda = 1$ e $\lambda = 0.1$, com um total de 421 e 7.334 imagens, respectivamente. Esses resultados mostram que não apenas o domínio de convergência mas também sua taxa são aumentadas em relação às técnicas anteriores. Isto ocorre a um custo computacional maior para estimar mais parâmetros e/ou para efetuar mais iterações no procedimento de otimização.

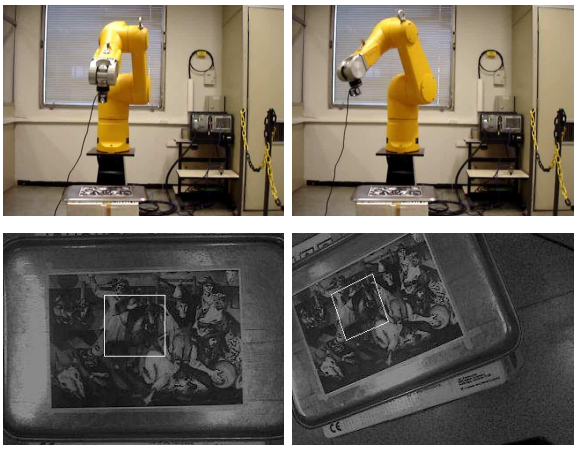

Figura 4: Servovisão 3D baseada em intensidade usando câmera e robô apenas grosseiramente calibrados, e uma estimativa inicial trivial da cena. (Acima) Configurações do robô com relação ao objeto. (Abaixo) Imagens capturadas naquelas poses relativas. Os algoritmos exploram todos os pixels dentro do quadrado branco.

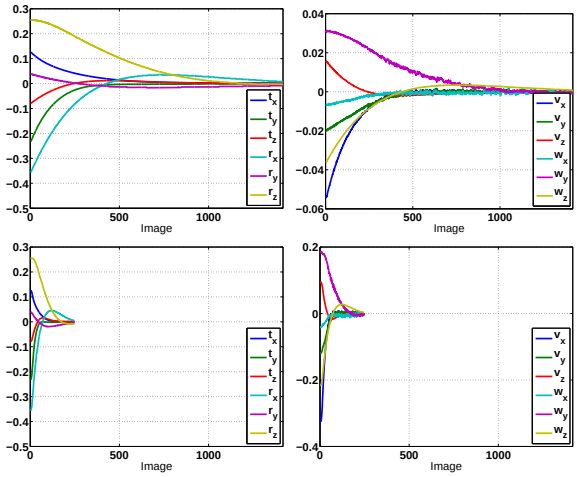

Figura 5: Evolução da câmera (esquerda) e dos sinais de controle (direita) para os métodos MP e OP no cenário da Fig. 4. O primeiro converge apenas para pequenos ganhos de controle. O segundo converge para um ganho maior, o que permitiu finalizar a tarefa na imagem #248.

4.2 Dados reais

Este segundo conjunto de experimentos utiliza um braço robótico de 6 g.d.l. O objeto é disposto a ≈0.7m da pose de referência da câmera. O deslocamento inicial do robô relativo à pose de referência é de $[0.13, -0.23, -0.08]^\top$m em translação (norma de 0.27m, i.e., de ≈38% das profundidades), e de $[-20.3, 2.17, 14.59]^\top$graus em rotação (norma de 25°). Com o intuito de demonstrar a robustez das técnicas, uma webcam grosseiramente calibrada é utilizada. De fato, as distâncias focais utilizadas são simplesmente 400 pixels, fator de obliquidade zero e ponto principal no meio da imagem, a qual possui 320 × 240 pixels. Esta câmera é posta no efetuador, e a calibração mão/olho é também muito grosseira. A condição inicial trivial na estrutura $\mathbf{n}^* = [0, 0, 1]^\top$ é fornecida. A taxa de quadro é de ≈30Hz, e o template de referência tem 70×70 pixels a fim de satisfazer as restrições de tempo real. A condição de parada utiliza $\epsilon_c = 10^{-3}$. Este cenário está mostrado na Fig. 4 e alguns resultados são mostrados na Fig. 5.

Método Pobre: Uma vez mais, o ganho de controle teve de ser reduzido a $\lambda = 0.1$ para que esta técnica convirja. Neste cenário, 1.417 imagens foram necessárias para estabilizar o sistema.

Otimização Parcial: Esta técnica de servovisão converge utilizando $\lambda = 0.1$, e também com um ganho maior de $\lambda = 0.6$. A taxa de convergência é assim aumentada, ao custo de realizar mais iterações por imagem na otimização.

Otimização Completa: O algoritmo de estimação não atualizou a estrutura da cena neste experimento (vide Nota 3.2). Portanto, a aplicação deste método não foi aqui necessária.

5 Conclusões

Este artigo investigou novas técnicas de servovisão 3D baseadas na intensidade dos pixels. Constatou-se que elas podem ser eficazes mesmo se os parâmetros estimados não são ótimos. Isto é especialmente válido em torno do equilíbrio, e se deve à exploração direta da intensidade e à questão de observabilidade associada aos sistemas monoculares. Esse fato permitiu o desenvolvimento de uma família de novos métodos, organizados hierarquicamente em termos de custo computacional e de conhecimento prévio. Trabalhos futuros podem ser dedicados à categorização dos subtipos de métodos no arcabouço proposto. Isso abre outros tópicos de pesquisa, tais como a definição de critérios para a seleção e/ou comutação entre eles. Com efeito, acredita-se que este artigo abre uma nova direção de pesquisa. Outro trabalho pode ser focado na análise da sua robustez aos erros de estimação. Os resultados apresentados são promissores dado que uma grande robustez foi observada nos experimentos. Esta questão tem sido completamente negligenciada em servovisão 3D, pois acreditava-se que parâmetros de controle corretos eram cruciais para a sua precisão final.

Referências

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

Faugeras, O., Luong, Q.-T. e Papadopoulo, T. (2001). *The geometry of multiple images*, The MIT Press.

Huber, P. J. (1981). *Robust Statistics*, John Wiley & Sons.

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

Morin, P. (2004). Stabilisation de systèmes non linéaires critiques et application à la commande de véhicules, *Habilit. à diriger des recherches*, Université de Nice.

Silveira, G. (2013). Direct 3-D tracking for central omnidirectional cameras under general lighting variations, *Journal of Control, Autom. Electr. Syst.* **24**: 129–138.

Silveira, G., Malis, E. e Rives, P. (2006). Visual servoing over unknown, unstructured, large-scale scenes, *Proc. of the IEEE ICRA*, USA, pp. 4142–4147.

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

Warner, F. W. (1987). *Foundations of differential manifolds and Lie groups*, Springer Verlag.