

SERVOVISÃO NÃO-MÉTRICA BASEADA EM INTENSIDADE: NOVAS TÉCNICAS E SUAS APLICAÇÕES

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 vision-based robot stabilization where the equilibrium state is defined via a reference image. Differently from most solutions, this work focuses on control techniques that directly use pixel intensities and only nonmetric information. Intensity-based techniques provide for higher accuracy, whereas not requiring metric information increases their versatility. In this context, this paper exploits the epipolar geometry and its intrinsic degeneracies, which always occur when that stabilization is sufficiently close to the equilibrium. This remarkable fact allows for developing new vision-based control strategies with varying degrees of computational complexity and of prior knowledge, all in a unified framework. Three new methods are proposed, and their performances are experimentally assessed using a six-dof robotic arm.

Keywords— Visual servo control, vision-based robot control, visual servoing, robot vision, computer vision, robotics.

Resumo— Este artigo considera o problema da estabilização de robôs por realimentação visual onde o estado de equilíbrio é definido via uma imagem de referência. Diferentemente da maioria das soluções, este trabalho foca em técnicas de controle que utilizam apenas informações não-métricas e exploram diretamente a intensidade dos pixels, sem etapas intermediárias. Técnicas baseadas em intensidade fornecem maior precisão, enquanto que dispensar informações métricas aumenta a versatilidade. Neste contexto, o presente trabalho explora a geometria epipolar e suas degenerações intrínsecas, as quais sempre ocorrem quando aquela estabilização está suficientemente próxima do equilíbrio. Este importante fato permite o desenvolvimento de novas estratégias de servovisão com diferentes graus de complexidade computacional e de conhecimento prévio. Três novos métodos são propostos, e seus desempenhos são experimentalmente avaliados utilizando um braço robótico de seis graus de liberdade.

Palavras-chave— controle servo visual, controle de robôs baseado em visão, servovisão, visão robótica, visão computacional, robótica.

1 Introdução

Controle servo visual, ou servovisão (do inglês, *visual servoing*), se refere ao controle de robôs por realimentação de imagens. Uma aplicação típica dessa abordagem consiste em estabilizá-los em torno de uma pose especificada via uma imagem de referência, também chamada de desejada.

A maioria das técnicas de servovisão requer informações métricas do objeto/cena para que a lei de controle seja comprovadamente estabilizante. Isto é válido inclusive para os métodos clássicos de servovisão baseada em imagem (Chaumette e Hutchinson, 2006), onde a profundidade (pelo menos aproximada) do objeto à câmera é necessária. Em relação às técnicas *não-métricas*, apesar de serem mais versáteis e robustas (Thaler e Goodale, 2010), existem apenas poucos trabalhos nesta classe. Uma possível razão para essa escassez é a dificuldade em definir-se um erro de controle difeomórfico à pose da câmera (pelo menos localmente). Um exemplo antigo dessas técnicas foi a navegação realizada por um robô de superfície (Beardsley et al., 1995). Neste caso, apenas três graus de liberdade (g.d.l.) foram controlados. Recentemente, técnicas não-métricas e gerais foram propostas (Silveira e Malis, 2012; Silveira et al., 2013). Elas são ditas *gerais* no sentido em que todos os seis g.d.l. de um robô são estabilizados, e em relação a objetos planares ou não, em movimentos rotacionais e translacionais.

Independentemente se informações métricas são utilizadas ou não, técnicas de estimação por visão são geralmente classificadas em dois grupos (Irani e Anandan, 1999): as baseadas em características visuais ou na intensidade dos pixels, também chamadas de *diretas*. O primeiro grupo requer a extração de um conjunto suficientemente grande de primitivas geométricas (e.g., pontos, retas, etc.), bem como sua correta associação, nas imagens. Diferentemente, as técnicas baseadas em intensidade exploram diretamente esses valores, sem etapas intermediárias, a fim de estimar os parâmetros necessários. Portanto, elas fazem uso de dados originais e densos de imagem, o que permite a obtenção de elevados níveis de precisão e versatilidade. Outra vantagem desse grupo se refere à possibilidade de assegurar robustez a mudanças arbitrárias de iluminação, mesmo em imagens coloridas e omnidirecionais (Silveira e Malis, 2009; Silveira, 2014).

Este trabalho investiga soluções computacionalmente eficientes para o problema da servovisão geral, não-métrica e direta. Para isso, são exploradas a geometria epipolar projetiva e suas degenerações, as quais *sempre* ocorrem quando a estabilização está próxima do equilíbrio. Este importante fato permite o desenvolvimento, neste trabalho, de três novas técnicas (duas subótimas e uma ótima) com diferentes graus de complexidade computacional e de conhecimento prévio. Diversos experimentos foram conduzidos para avaliá-las, utili-

zando objetos planares e não-planares, simulando e empregando um braço robótico industrial de seis g.d.l. As técnicas subótimas possuem aplicação quando o poder computacional disponível é baixo comparativamente à quantidade de dados visuais, ou se a translação entre a pose corrente e a desejada é pequena em relação às profundidades da cena. A técnica ótima é o método de escolha se nenhum conhecimento prévio estiver disponível.

2 Fundamentação Teórica

Esta seção define a notação e recorda modelos essenciais. Permita que $\|\mathbf{v}\|$ e $\hat{\mathbf{v}}$ denotem, resp., a norma Euclidiana e uma estimativa da variável \mathbf{v} . Um asterisco, e.g., \mathbf{v}^* , indica que \mathbf{v} é definida em relação ao sistema de coordenadas de referência \mathcal{F}^* . As notações $[\mathbf{w}]_{\times}$ e $\text{vex}([\mathbf{w}]_{\times})$ representam, resp., a matriz antissimétrica associada a $\mathbf{w} = [w_1, w_2, w_3]^T$ e o seu mapeamento inverso:

$$[\mathbf{w}]_{\times} = \begin{bmatrix} 0 & -w_3 & w_2 \\ w_3 & 0 & -w_1 \\ -w_2 & w_1 & 0 \end{bmatrix}, \text{vex}([\mathbf{w}]_{\times}) = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}.$$

Considere também o operador

$$P_a(\mathbf{W}) = \mathbf{W} - \mathbf{W}^T, \quad (1)$$

o qual é duas vezes o operador de projeção da matriz quadrada \mathbf{W} em sua parte antissimétrica.

2.1 Geometria epipolar projetiva

A relação geral entre pixels correspondentes $\mathbf{p} \leftrightarrow \mathbf{p}^*$ em duas imagens perspectivas é descrita em geometria projetiva por

$$\mathbf{p} \propto \mathbf{G}\mathbf{p}^* + \rho^*\mathbf{e} \in \mathbb{P}^2, \quad (2)$$

onde o símbolo “ \propto ” denota proporcionalidade, $\mathbf{G} \in \mathbb{SL}(3)$ é a homografia induzida por um plano dominante, $\mathbf{e} \in \mathbb{P}^2$ denota o epipolo, e $\rho^* \in \mathbb{R}$ é a paralaxe do ponto 3D cuja projeção na imagem de referência \mathcal{I}^* é \mathbf{p}^* , relativamente àquele plano. Esta paralaxe é proporcional à *distância* deste ponto 3D ao plano e é inversamente proporcional à *profundidade* deste ponto (vide Fig. 1). O epipolo é proporcional à *translação* entre o sistema de coordenadas corrente \mathcal{F} e o de referência \mathcal{F}^* .

A restrição epipolar básica que existe entre duas imagens de uma mesma cena pode ser obtida multiplicando pela esquerda ambos lados de (2) por $\mathbf{p}^T[\mathbf{e}]_{\times}$, i.e., $0 = \mathbf{p}^T \mathbf{F} \mathbf{p}^*$, onde $\mathbf{F} \propto [\mathbf{e}]_{\times} \mathbf{G}$ é usualmente chamada de matriz Fundamental. Embora esta matriz é em geral única para um par de imagens, existem casos degenerados onde pelo menos duas matrizes Fundamentais distintas (i.e., linearmente independentes) satisfazem aquela restrição. Casos degenerados de especial importância para a servovisão são descritos a seguir.

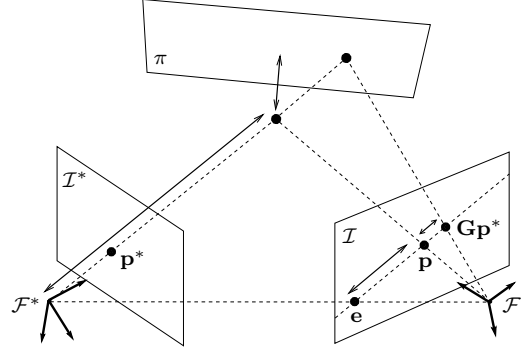


Figura 1: Geometria projetiva entre duas imagens.

Nota 2.1 (Casos degenerados) *Importantes degenerações da geometria epipolar ocorrem para:*

- *movimentos puramente rotacionais, independentemente do formato do objeto. Neste caso, a translação é zero e, portanto, $\mathbf{e} = \mathbf{0}$, o que sempre ocorre no equilíbrio;*
- *objetos no infinito, independentemente do formato do objeto. Neste caso, $\rho^* = 0$ para todos os seus pontos;*
- *objetos planares. Também neste caso, $\rho^* = 0$ para todos os seus pontos;*

Todos os casos acima verificam

$$\rho^*\mathbf{e} = \mathbf{0}, \quad (3)$$

independentemente do ponto do objeto, i.e., $\forall \rho_i^*$, $i = 1, 2, \dots, n$, onde n é o número de pixels que o descreve. A relação geral (2) engloba todos esses casos, e pode ser reescrita como

$$\mathbf{p} \propto [\mathbf{I} \ \mathbf{0}] \mathbf{Q} [\mathbf{p}^{*T} \ \rho^*]^T, \quad (4)$$

com

$$\mathbf{Q} = \begin{bmatrix} \mathbf{G} & \mathbf{e} \\ \mathbf{0} & 1 \end{bmatrix} \in \mathbb{SA}(3), \quad (5)$$

onde o grupo de Lie $\mathbb{SA}(3)$ é homeomórfico ao $\mathbb{SL}(3) \times \mathbb{R}^3$. Assim, todos os parâmetros globais estão em \mathbf{Q} , i.e., o deslocamento da câmera e a base projetiva. A estrutura do objeto relativa a essa base e ao \mathcal{F}^* está descrita em $\boldsymbol{\rho}^* = [\rho_1^*, \rho_2^*, \dots, \rho_n^*]^T$, que representa parâmetros locais.

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 com 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

$$\mathcal{I}'(\mathbf{x}(\mathbf{z}), \mathbf{p}^*) = \mathcal{I}(\mathbf{w}(\mathbf{x}(\mathbf{z}), \mathbf{p}^*)) \geq 0, \quad (6)$$

onde a função de transformação geométrica (do inglês *warping*) $\mathbf{w}: \mathbb{SA}(3) \times \mathbb{R}^n \times \mathbb{P}^2 \rightarrow \mathbb{P}^2$ pode ser definido a partir de (4), com $\mathbf{x} = \{\mathbf{Q}, \boldsymbol{\rho}^*\}$ e sua respectiva parametrização $\mathbf{z} = [\mathbf{v}^\top \boldsymbol{\gamma}^\top]^\top \in \mathbb{R}^m$, i.e., $\mathbf{x} = \mathbf{x}(\mathbf{z}) = \{\mathbf{Q}(\mathbf{v}), \boldsymbol{\rho}^*(\boldsymbol{\gamma})\}$.

Um sistema típico de registro direto de imagens pode ser formulado como o seguinte problema de otimização não-linear:

$$\min_{\mathbf{z} \in \mathbb{R}^m} \frac{1}{2} \sum_{i=1}^n [\mathcal{I}'(\mathbf{x}(\mathbf{z}), \mathbf{p}_i^*) - \mathcal{I}^*(\mathbf{p}_i^*)]^2, \quad (7)$$

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

$$\mathbf{d}(\mathbf{x}(\mathbf{z})) = \begin{bmatrix} \mathcal{I}'(\mathbf{x}(\mathbf{z}), \mathbf{p}_1^*) - \mathcal{I}^*(\mathbf{p}_1^*) \\ \mathcal{I}'(\mathbf{x}(\mathbf{z}), \mathbf{p}_2^*) - \mathcal{I}^*(\mathbf{p}_2^*) \\ \vdots \\ \mathcal{I}'(\mathbf{x}(\mathbf{z}), \mathbf{p}_n^*) - \mathcal{I}^*(\mathbf{p}_n^*) \end{bmatrix} \in \mathbb{R}^n. \quad (8)$$

Outras funções custo podem ser consideradas em (7). No entanto, modificações neste aspecto não alteram o arcabouço básico.

O problema de otimização não-linear (7) pode ser resolvido por métodos iterativos clássicos como, por exemplo, via 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 $\hat{\mathbf{x}}_0$ suficientemente próxima da solução, o incremento $\mathbf{z}_k \in \mathbb{R}^m$ nas variáveis de transformação é calculada na iteração k por

$$\mathbf{z}_k = -\alpha \mathbf{L}_x^+ \mathbf{d}(\hat{\mathbf{x}}_k), \quad (9)$$

com $\alpha > 0$ ($\alpha \approx 1$ em torno de um mínimo local) e, para os métodos clássicos,

$$\mathbf{L}_x^+ = \hat{\mathbf{H}}_x^{-1} \mathbf{J}_x^\top, \quad (10)$$

onde $\mathbf{J}_x \in \mathbb{R}^{n \times m}$ denota a matriz Jacobiana de (8) com relação a \mathbf{x} em \mathbf{z} , e $\hat{\mathbf{H}}_x \in \mathbb{R}^{m \times m}$ é uma matriz positiva definida que aproxima¹ adequadamente a matriz Hessiana da função custo. O incremento (9) atualiza a variável $\hat{\mathbf{x}}_k$ via

$$\hat{\mathbf{x}}_{k+1} = \mathbf{x}(\mathbf{z}_k) \circ \hat{\mathbf{x}}_k, \quad (11)$$

e o processo é iterado até a convergência. O símbolo “ \circ ” denota o operador de composição associado ao grupo envolvido. Na prática, a convergência pode ser obtida quando $\mathbf{x}(\mathbf{z}_k)$ for suficientemente próximo do elemento identidade do grupo

¹Exemplos dessa aproximação são: 1) para o método do gradiente, ela é simplesmente $\hat{\mathbf{H}}_x = \mathbf{I}$; 2) para o método de Gauss-Newton, ela é dada por $\hat{\mathbf{H}}_x = \mathbf{J}_x^\top \mathbf{J}_x$; 3) para o método de Levenberg-Marquardt, ela é $\hat{\mathbf{H}}_x = \mathbf{J}_x^\top \mathbf{J}_x + \sigma \mathbf{D}$, onde $\sigma > 0$ e $\mathbf{D} \in \mathbb{R}^{m \times m}$ é uma matriz diagonal, e.g., $\mathbf{D} = \mathbf{I}$ ou $\mathbf{D} = \text{diag}(\mathbf{J}_x^\top \mathbf{J}_x)$.

envolvido, i.e., quando $\|\mathbf{z}_k\| < \epsilon_e$, para algum valor $\epsilon_e > 0$ suficientemente pequeno, ou quando o número máximo de iterações for atingido.

Nota 2.2 *Embora cada método de otimização impacte de forma diferente no desempenho global da servovisão, as ideias gerais desse artigo não dependem do escolhido, como mostrado a seguir.*

3 Técnicas Propostas

Esta seção apresenta novas técnicas não-métricas de servovisão que exploram diretamente a intensidade dos pixels. De fato, o sistema de estimação é baseado nas diferenças de intensidade $\mathbf{d} \in \mathbb{R}^n$, o qual é construído a partir de parâmetros não-métricos. Em particular, esta seção mostra como beneficiar-se das degenerações da geometria epipolar para tornar essas técnicas mais computacionalmente eficientes. Para isso, considere um robô holonômico de seis g.d.l. com uma câmera convencional montada em seu efetador, observando um objeto imóvel, rígido e de forma desconhecida. Sejam as entradas de controle $\boldsymbol{\nu} \in \mathbb{R}^3$ e $\boldsymbol{\omega} \in \mathbb{R}^3$, respectivamente, as velocidades translacionais e rotacionais da câmera, cujos parâmetros intrínsecos estão agrupados em $\mathbf{K} \in \mathbb{R}^{3 \times 3}$.

3.1 Otimização Completa

Esta técnica de servovisão não-métrica baseada em intensidade efetua uma otimização simultânea de todas as variáveis geométricas envolvidas. O sistema de controle utiliza, portanto, estimativas ótimas. É, por conseguinte, o método de escolha se nenhum conhecimento prévio estiver disponível. Por outro lado, este procedimento de estimação pode ser computacionalmente intensivo.

Este método pode ser brevemente formulado como se segue. Dado uma estimativa inicial $\hat{\mathbf{x}}_0 = \{\hat{\mathbf{Q}}_0, \hat{\boldsymbol{\rho}}_0^*\}$, os incrementos $\mathbf{v}_k \in \mathbb{R}^{8+3}$ e $\boldsymbol{\gamma}_k \in \mathbb{R}^{\dim(\boldsymbol{\gamma})}$ são obtidos na iteração k para a imagem corrente via

$$[\mathbf{v}_k^\top \boldsymbol{\gamma}_k^\top]^\top = -\alpha [\mathbf{L}_Q \mathbf{L}_\rho]^\top \mathbf{d}(\hat{\mathbf{x}}_k), \quad (12)$$

onde $\mathbf{L}_Q \in \mathbb{R}^{n \times 11}$, $\mathbf{L}_\rho \in \mathbb{R}^{n \times \dim(\boldsymbol{\gamma})}$, e n é o número de pixels considerados na otimização. Esses incrementos atualizam as variáveis via

$$\hat{\mathbf{Q}}_{k+1} = \mathbf{Q}(\mathbf{v}_k) \circ \hat{\mathbf{Q}}_k, \quad (13)$$

$$\hat{\boldsymbol{\rho}}_{k+1}^* = \boldsymbol{\rho}^*(\boldsymbol{\gamma}_k) \circ \hat{\boldsymbol{\rho}}_k^*, \quad (14)$$

e o processo é iterado até sua convergência, e.g., $\|[\mathbf{v}_k^\top \boldsymbol{\gamma}_k^\top]^\top\| < \epsilon_e$.

Em relação aos aspectos de controle, defina seus erros não-métricos translacionais $\boldsymbol{\varepsilon}_\nu^F \in \mathbb{R}^3$ e rotacionais $\boldsymbol{\varepsilon}_\omega^F \in \mathbb{R}^3$ como

$$\begin{bmatrix} \boldsymbol{\varepsilon}_\nu^F \\ \boldsymbol{\varepsilon}_\omega^F \end{bmatrix} = \begin{bmatrix} [\mathbf{K}^{-1} \mathbf{0}] (\hat{\mathbf{Q}} - \mathbf{I}) [\mathbf{p}^{*\top} \hat{\boldsymbol{\rho}}^*]^\top \\ \text{vex}(\mathbf{P}_a([\mathbf{K}^{-1} \mathbf{0}] \hat{\mathbf{Q}} [\mathbf{K}^\top \mathbf{0}]^\top)) \end{bmatrix}, \quad (15)$$

onde $\hat{\mathbf{Q}} \in \mathbb{SA}(3)$ é a matriz ótima e $\hat{\rho}^* \in \mathbb{R}$ é o paralaxe ótimo do ponto de controle escolhido. Ambas estimativas podem ser utilizadas como inicialização na próxima imagem. Finalmente, a lei de controle pode ser definida como

$$\begin{bmatrix} \nu \\ \omega \end{bmatrix} = \lambda \begin{bmatrix} \varepsilon_\nu^F \\ \varepsilon_\omega^F \end{bmatrix}, \quad (16)$$

com ganho de controle $\lambda > 0$. A condição de parada da servovisão pode ser, e.g., $\|\varepsilon_\nu^{F\top} \varepsilon_\omega^{F\top}\| < \epsilon_c$ para um pequeno $\epsilon_c > 0$. Assim, esta técnica corresponde a uma versão reduzida de (Silveira e Malis, 2012). O equilíbrio $[\varepsilon_\nu^{F\top} \varepsilon_\omega^{F\top}] = \mathbf{0}$ do sistema de controle em malha fechada é, portanto, localmente exponencialmente estável.

3.2 Otimização Parcial

Nesta técnica de servovisão não-métrica baseada em intensidade, a estratégia de estimação otimiza apenas um subconjunto de variáveis. De fato, a ideia consiste em otimizar apenas os parâmetros relacionados com o deslocamento da câmera, i.e., \mathbf{Q} . Os parâmetros relacionados com a estrutura do objeto, i.e., ρ^* , é fornecido pelo usuário e não é ajustado. Essa ideia é inspirada nos casos degenerados (vide Nota 2.1) onde, em torno do equilíbrio ou se o objeto for planar, temos $\rho_i^* \mathbf{e} \rightarrow \mathbf{0}$, $\forall i \in \{1, 2, \dots, n\}$. Essa técnica exige, portanto, menos esforços computacionais do que a precedente, mas depende dessas condições de operação.

Este método pode ser descrito como se segue. Dado uma estimativa inicial $\hat{\mathbf{x}}_0 = \{\hat{\mathbf{Q}}_0, \hat{\rho}_0^*\}$, o incremento $\mathbf{v}_k \in \mathbb{R}^{8+3}$ é calculado na iteração k para a imagem corrente via

$$\mathbf{v}_k = -\alpha \mathbf{L}_Q^+ \mathbf{d}(\hat{\mathbf{x}}_k). \quad (17)$$

Este incremento atualiza $\hat{\mathbf{Q}}_k$ via

$$\hat{\mathbf{Q}}_{k+1} = \mathbf{Q}(\mathbf{v}_k) \circ \hat{\mathbf{Q}}_k, \quad (18)$$

e o processo é iterado até a convergência, e.g., $\|\mathbf{v}_k\| < \epsilon_c$. A variável $\hat{\rho}_0^*$ não é otimizada, mas é necessária para a determinação de (17).

Para os aspectos de controle, seus erros não-métricos translacionais $\varepsilon_\nu^Q \in \mathbb{R}^3$ e rotacionais $\varepsilon_\omega^Q \in \mathbb{R}^3$ são definidos como

$$\begin{bmatrix} \varepsilon_\nu^Q \\ \varepsilon_\omega^Q \end{bmatrix} = \begin{bmatrix} [\mathbf{K}^{-1} \mathbf{0}] (\hat{\mathbf{Q}} - \mathbf{I}) [\mathbf{p}^{*\top} \hat{\rho}_0^*]^\top \\ \text{vex}(\mathbf{P}_a([\mathbf{K}^{-1} \mathbf{0}] \hat{\mathbf{Q}} [\mathbf{K}^\top \mathbf{0}]^\top)) \end{bmatrix}, \quad (19)$$

onde $\hat{\mathbf{Q}} \in \mathbb{SA}(3)$ é a estimativa obtida, a qual pode ser usada como inicialização na próxima imagem. A lei de controle respectiva pode ser definida como

$$\begin{bmatrix} \nu \\ \omega \end{bmatrix} = \lambda \begin{bmatrix} \varepsilon_\nu^Q \\ \varepsilon_\omega^Q \end{bmatrix}, \quad (20)$$

com $\lambda > 0$. A convergência da tarefa é obtida quando, e.g., $\|\varepsilon_\nu^{Q\top} \varepsilon_\omega^{Q\top}\| < \epsilon_c$. A técnica de

servovisão em (Benhimane e Malis, 2006) corresponde a um caso particular desta. De fato, ela considera apenas objetos planares ou rotações puras da câmera, i.e., $\rho_i^* \mathbf{e} = \mathbf{0}$, $\forall i$, o que fornece $\mathbf{x} = \mathbf{G}$ (4). Neste caso, $\mathbf{v}_k \in \mathbb{R}^8$ e $\mathbf{L}_Q = \mathbf{L}_G \in \mathbb{R}^{n \times 8}$ em (17), e o equilíbrio $[\varepsilon_\nu^{Q\top} \varepsilon_\omega^{Q\top}] = \mathbf{0}$ do sistema de controle em malha fechada é provado localmente assintoticamente estável.

3.3 Método Pobre

Esta seção apresenta a técnica não-métrica baseada em intensidade computacionalmente mais simples dentre as acima descritas. Seu baixo custo é devido tanto à sua estratégia de estimação, quanto ao número de parâmetros estimados. Quanto à estimação, a ideia é realizar apenas uma iteração de (9) 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, como no método da Otimização Parcial. Os parâmetros associados à estrutura são fornecidos pelo usuário.

Este método pode ser formalizado como se segue. Dado uma estimativa inicial $\hat{\mathbf{x}}_0 = \{\hat{\mathbf{Q}}_0, \hat{\rho}_0^*\}$, o incremento $\mathbf{v}_0 \in \mathbb{R}^{8+3}$ é computado apenas uma única vez para cada imagem capturada:

$$\mathbf{v}_0 = -\alpha \mathbf{L}_Q^+ \mathbf{d}(\hat{\mathbf{x}}_0). \quad (21)$$

A atualização de $\hat{\mathbf{Q}}_0$ é também efetuada apenas uma única vez para cada imagem:

$$\hat{\mathbf{Q}}_1 = \mathbf{Q}(\mathbf{v}_0) \circ \hat{\mathbf{Q}}_0, \quad (22)$$

a qual pode ser utilizada como estimativa inicial na próxima imagem. Assim como no método precedente, a variável $\hat{\rho}_0^*$ não é ajustada.

O erro de controle no método pobre pode ser definido como

$$\begin{bmatrix} \varepsilon_\nu^1 \\ \varepsilon_\omega^1 \end{bmatrix} = \begin{bmatrix} [\mathbf{K}^{-1} \mathbf{0}] (\hat{\mathbf{Q}}_1 - \mathbf{I}) [\mathbf{p}^{*\top} \hat{\rho}_0^*]^\top \\ \text{vex}(\mathbf{P}_a([\mathbf{K}^{-1} \mathbf{0}] \hat{\mathbf{Q}}_1 [\mathbf{K}^\top \mathbf{0}]^\top)) \end{bmatrix}. \quad (23)$$

Finalmente, a lei de controle respectiva pode ser definida simplesmente como

$$\begin{bmatrix} \nu \\ \omega \end{bmatrix} = \lambda \begin{bmatrix} \varepsilon_\nu^1 \\ \varepsilon_\omega^1 \end{bmatrix}, \quad (24)$$

com $\lambda > 0$. A convergência da tarefa pode ser estabelecida quando, e.g., $\|\varepsilon_\nu^{1\top} \varepsilon_\omega^{1\top}\| < \epsilon_c$. Observe que se por um lado ela é a mais barata computacionalmente, por outro lado sua taxa e domínio de convergência são os menores possíveis, dada sua estratégia de estimação. Como consequência, seu sucesso depende de conhecimentos prévios sobre o sistema, e.g., sobre a qualidade de $\hat{\mathbf{x}}_0$.

4 Resultados Experimentais

Esta seção apresenta resultados experimentais obtidos pelas três novas técnicas de servovisão propostas, i.e., o Método Pobre, via Otimização Parcial e pela Otimização Completa. Essas técnicas

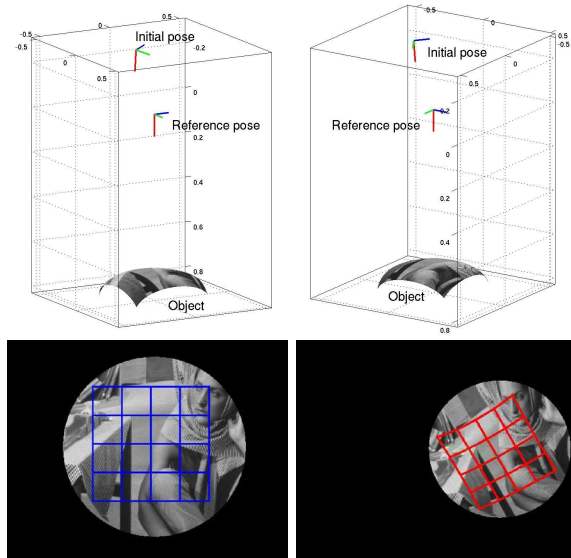


Figura 2: Experimento com dados sintéticos. (Acima) Poses da câmera em relação ao objeto, sob diferentes pontos de vista. Apenas a parte explorada do alvo é mostrada. (Abaixo) Imagens de referência e inicial, respectivamente.

não utilizam informações métricas, nem extraem ou associam características visuais nas imagens. As intensidades dos pixels são diretamente exploradas para estimar os parâmetros de controle necessários. O objetivo de controle é posicionar o robô de forma que a imagem corrente coincida com a imagem capturada na pose de referência. Os experimentos foram realizados tanto com dados sintéticos, quanto com um braço robótico de seis g.d.l. dotado de uma câmera em sua extremidade.

4.1 Dados sintéticos

Nesse primeiro conjunto de experimentos, o objeto considerado é esférico. O *template* (i.e., a região de interesse da imagem) de referência possui 200×200 pixels, cujas intensidades são todas exploradas. As distâncias focais são definidas como 500 pixels, fator de obliquidade zero, e o ponto principal como o centro da imagem. O período de amostragem é de 30 ms e o ganho de controle é $\lambda = 1$. 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 deslocamento da câmera entre a pose de referência e a inicial (norma de 0,33 m) é relativamente grande comparativamente às profundidades da cena (mediana de 0,72 m para a referência). Este cenário está mostrado na Fig. 2, enquanto que os resultados obtidos estão na Fig. 3 e descritos a seguir.

Método Pobre (MP): Neste cenário de deslocamento inicial relativamente grande ($\approx 50\%$ das profundidades), o MP falha para quaisquer ganhos de controle $\lambda > 0,1$. Com $\lambda = 0,06$, a servovição converge, porém isso ocorre apenas após 6 028 imagens. Para um deslocamento inicial de $\approx 5\%$ das profundidades, o MP funciona mesmo para $\lambda = 1$, embora as entradas de controle não são

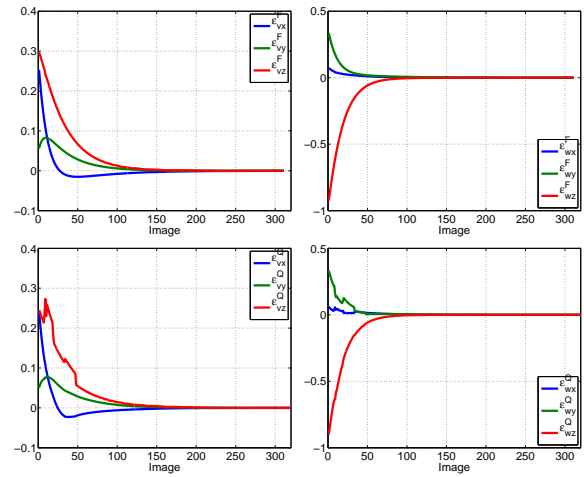


Figura 3: Evolução dos erros de controle para o experimento da Fig. 2 com $\lambda = 1$. (Acima) Erros translacionais e rotacionais para o método OC. (Abaixo) Erros translacionais e rotacionais para o método OP. O MP falhou, mas converge se $\lambda = 0,06$ (não mostrado).

completamente suaves. Esses resultados sugerem que o MP requer a aplicação de um ganho de controle variável. Em todo caso, seus requisitos computacionais são extremamente reduzidos dado que uma única iteração é realizada na otimização.

Otimização Parcial (OP): Para essa técnica e no cenário de deslocamento inicial relativamente grande, a tarefa converge com sucesso mesmo para o ganho de controle $\lambda = 1$, embora as entradas de controle ainda não são suaves. Neste caso, 320 imagens foram necessárias à estabilização. Para $\lambda = 0,06$, as entradas de controle são suaves e, para esse pequeno ganho, o método OP convergiu após 5 215 imagens. A complexidade computacional é maior que o MP dado que mais iterações por imagem são efetuadas durante o procedimento de estimação.

Otimização Completa (OC): Para este método, a estabilização é realizada com sucesso para ambos ganhos de controle $\lambda = 1$ e $\lambda = 0,06$, após um total de 311 e 4 883 imagens, respectivamente. Além de aumentar a taxa de convergência, as entradas de controle são suaves em ambos os casos. Por outro lado, o método OC exige maior poder computacional para estimar mais parâmetros e realizar mais iterações por imagem.

4.2 Dados reais

Este conjunto de experimentos considera um braço robótico observando um objeto planar, disposto a $\approx 0,3$ m da pose de referência. A norma do deslocamento inicial imposto é de $\approx 150\%$ dessa distância. É, portanto, uma perturbação significativa. Além disso, para demonstrar a robustez das técnicas, uma *webcam* apenas grosseiramente calibrada é considerada. 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

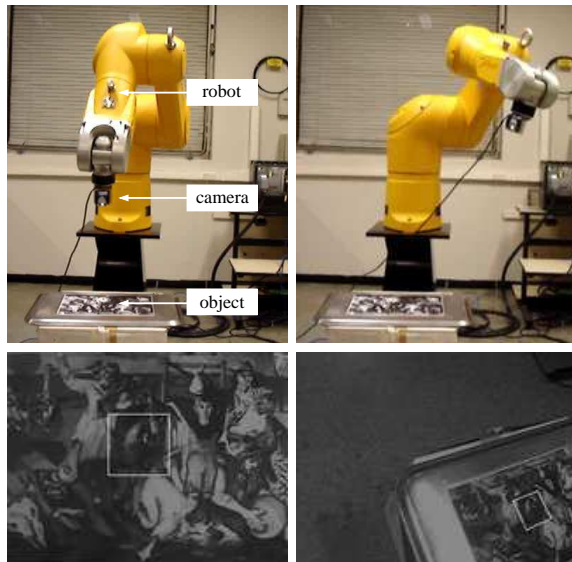


Figura 4: Experimento com dados reais, utilizando um braço robótico com câmera apenas grosseiramente calibrada. (Acima) Poses de referência e inicial do robô em relação ao objeto. (Abaixo) Imagens respectivas.

está montada no efetuador do robô, e a calibração mão/olho é também apenas aproximada. A taxa de quadro é de ≈ 30 Hz, e o *template* de referência tem 70×70 pixels a fim de satisfazer as restrições de tempo real. A condição de parada da servovisão utiliza $\epsilon_c = 10^{-4}$. Este cenário está mostrado na Fig. 4. A despeito de todas essas perturbações, a servovisão é realizada com precisão, cuja norma do erro em rotação é de $0,2^\circ$ e de $0,2$ mm em translação. Alguns resultados são mostrados na Fig. 5, e são brevemente descritos abaixo.

Método Pobre (MP): Neste cenário, o MP apenas converge para um pequeno ganho de controle de $\lambda = 0,06$, o que exige 3 990 imagens.

Otimização Parcial (OP): O método OP converge para uma ganho maior: $\lambda = 0,3$. Neste caso,² apenas 879 imagens foram necessárias para estabilizar o sistema.

Otimização Completa (OC): Dado que o objeto utilizado é planar, então $\rho^* = 0$ para todos os seus pontos. Portanto,³ o método OC é, neste caso, equivalente ao OP.

5 Conclusões

Este artigo investigou novas técnicas de servovisão não-métrica baseada em intensidade. Constatou-se que elas podem ser precisas e eficazes mesmo se os parâmetros de controle não são ótimos ou corretos. Isto é especialmente válido em torno do equilíbrio, independente da forma do objeto, graças às degenerações da geometria epipolar. Este importante fato permitiu o desenvolvimento de um arcabouço unificado, e três novas técnicas foram pro-

²Este resultado está disponível em vídeo no site <http://goo.gl/vbB6S6>

³Outro resultado está disponível em vídeo no site <http://goo.gl/1Q34Za>

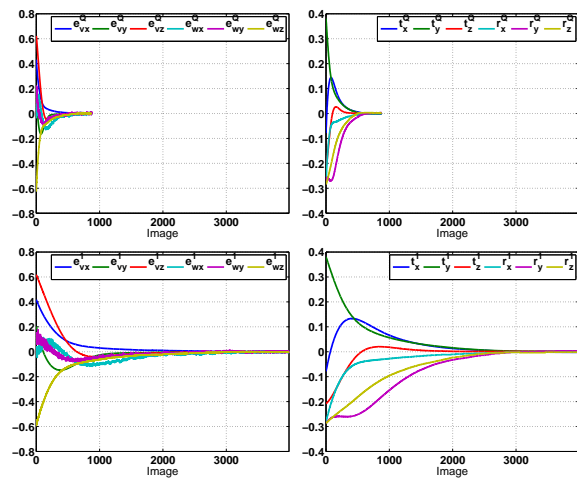


Figura 5: Evolução dos erros de controle e da câmera no espaço 3D para o experimento da Fig. 4. (Acima) Sinais para o método OP, o qual converge na imagem 879 com $\lambda = 0,3$. (Abaixo) Sinais para o MP, o qual apenas converge com $\lambda = 0,06$, o que exige 3 990 imagens.

postas. As técnicas subótimas propostas possuem aplicação quando o poder computacional disponível é baixo quando comparado à quantidade de dados visuais, ou se a translação da câmera é pequena em relação às profundidades da cena. Trabalhos futuros podem ser dedicados à categorização de todos os subtipos de métodos no arcabouço proposto, e.g., deslocamentos particulares, procedimentos de otimização, etc. Isto abre outros tópicos de investigação, tais como a definição de critérios para a seleção e comutação on-line entre os métodos. Acredita-se que este trabalho abre, de fato, uma nova frente de pesquisas em servovisão.

Referências

- Beardsley, P. A., Reid, I. D., Zisserman, A. e Murray, D. W. (1995). Active visual navigation using non-metric structure, *Proc. IEEE ICCV*, pp. 58–64.
- Benhimane, S. e Malis, E. (2006). Homography-based 2D visual servoing, *Proc. IEEE ICRA*, EUA.
- Chaumette, F. e Hutchinson, S. (2006). Visual servo control part I: Basic approaches, *IEEE Robotics & Automation Magazine* pp. 82–90.
- Irani, M. e Anandan, P. (1999). All about direct methods, *Workshop on Vision Algorithms: Theory and practice*.
- Luenberger, D. G. (1984). *Linear and Nonlinear Programming*, Addison-Wesley.
- Silveira, G. (2014). Photogeometric direct visual tracking for central omnidirectional cameras, *Journal of Mathematical Imaging and Vision* **48**(1): 72–82.
- Silveira, G. e Malis, E. (2009). Visual servoing from robust direct color image registration, *Proc. IEEE/RSJ IROS*, USA, pp. 5450–5455.
- Silveira, G. e Malis, E. (2012). Direct Visual Servoing: Vision-based estimation and control using only non-metric information, *IEEE Transactions on Robotics* **28**(4): 974–980.
- Silveira, G., Mirisola, L. e Morin, P. (2013). Decoupled direct visual servoing, *Proc. IEEE/RSJ IROS*, Japan, pp. 71–76.
- Thaler, L. e Goodale, M. A. (2010). Beyond distance and direction: The brain represents target locations non-metrically, *Journal of Vision* **10**(3): 1–27.