

**UNIVERSITY OF DORTMUND**

---

REIHE COMPUTATIONAL INTELLIGENCE

---

COLLABORATIVE RESEARCH CENTER 531

---

Design and Management of Complex Technical Processes  
and Systems by means of Computational Intelligence Methods

---

Weighted moments of SIFT-features for decoupled  
visual servoing in 6DOF

Thomas Nierobisch, Johannes Krettek,  
Umar Kahn and Frank Hoffmann

No. CI-226/07

Technical Report

ISSN 1433-3325

February 2007

Secretary of the SFB 531 · University of Dortmund · Dept. of Computer Science/LS 2  
44221 Dortmund · Germany

---

This work is a product of the Collaborative Research Center 531, "Computational Intelligence," at the University of Dortmund and was printed with financial support of the Deutsche Forschungsgemeinschaft.



# Weighted moments of SIFT-features for decoupled visual servoing in 6DOF

T. Nierobisch, J. Krettek, U. Khan and F. Hoffmann  
Chair for Control and Systems Engineering, University of Dortmund, Germany  
thomas.nierobisch@uni-dortmund.de

**Abstract** – *Mobile manipulation in service robotic applications requires the alignment of the end-effector with recognized objects of unknown pose. Image based visual servoing provides a means of model-free manipulation of objects solely relying on 2D image information. This contribution proposes a visual servoing scheme that utilizes the pixel coordinates, scale and orientation of the Scale Invariant Feature Transform (SIFT). The control is based on dynamic weighted moments aggregated from a mutable set of redundant SIFT feature correspondences between the current and the reference view. The key idea of our approach is to weight the point features in a way that eliminates or at least minimizes the undesired couplings to establish a one-to-one correspondence between feature and camera motion. The scheme achieves a complete decoupling in 4DOF, additionally the visual features in 6DOF are largely decoupled except for some minor residual coupling between the horizontal translation and the two rotational degrees of freedom. The decoupling results in a visual control in which the convergence in image and task space along a particular degree of motion is not effected by the remaining feature errors. Several simulations in virtual reality and experiments on a robotic arm equipped with a monocular eye-in-hand camera demonstrate that the approach is robust, efficient and reliable for 4DOF as well as 6DOF positioning.*

**Keywords:** image based visual servoing, SIFT features, robotic manipulation, decoupled visual features

## 1 Introduction

This paper advocates SIFT features [1] for 6DOF visual servoing of a robotic manipulator with an eye-in-hand camera configuration. The increasing availability of inexpensive cameras and powerful computers opens a novel avenue for integrating image processing systems as a sensor for real-time control of robotic manipulators. Image and position based visual servoing grows in visibility due to its importance for robotic manipulation and grasping [2].

Our point of departure is the conventional visual servoing paradigm developed for an eye-in-hand vision guided manipulation task originally introduced in [3]. The visual point features are defined directly in the 2D image plane, therefore a geometric object model or an explicit reconstruction

of the object pose becomes obsolete. The motion of a feature with image coordinates  $f = (u, v)^T$  is related to the camera motion via the image Jacobian or sensitivity matrix  $J_v$  according to  $\dot{f} = J_v(r)\dot{r}$ . Image based visual servoing builds upon this relationship by an error proportional control law in which the error between reference features  $\hat{f}$  and current features  $f$  is compensated by a camera motion

$$\dot{r} = -K \cdot J_v^+(r)(\hat{f} - f), \quad (1)$$

in which  $J_v^+$  denotes the pseudo-inverse of the image Jacobian and  $K$  is a gain matrix. The computation of the analytical image Jacobian requires knowledge about the depth of the scene and the intrinsic camera parameters. Image based visual servoing with point features suffers from the handicap that exclusive control of features in the image might result in an inferior or infeasible camera motion. The underlying problem is caused by the coupling between translational and rotational degrees of freedom and is particularly imminent in the presence of substantial errors in orientation. In order to obtain a better motion control, the approach proposed in [4] decouples the translational and rotational degrees of freedom by using the homography or the epipolar constraint held between the reference and the current image. In [5] the authors pursue a similar approach called 2-1/2 D visual servoing that includes a partial pose estimate to compensate the camera rotation. Both approaches require the online estimation of the homography or fundamental matrix in the servo loop, with at least four or eight feature point correspondences. The scheme is fairly sensitive to image noise for poses close to the reference view. The approach in [6] relies on image moments rather than point features to overcome the above mentioned shortcomings of visual servoing schemes. Our approach takes on the idea to design visual moments that are primarily sensitive to a single degree of motion and appear invariant to the remaining translations or rotations. Our scheme exploits certain invariance properties of geometric moments to decouple visual features across different degrees of motion. SIFT features for visual servoing applications were first introduced by [7], in which the authors explicitly reconstruct the object pose based on the epipolar geometry. A novel image-based controller that augments conventional point features by the additional attributes scale and keypoint

orientation of SIFT features is presented in [8]. The visual features scale and keypoint orientation turn out to be widely independent of translation and rotation along the other axes which qualifies them to control the distance to the object and the rotation around the optical axis. Six generic visual moments are defined based on the pixel coordinates, scale and orientation of multiple SIFT features. These moments are dynamically adapted to the current geometric distribution of available SIFT features. The trade-off between computational complexity and positioning accuracy is controlled by the number of scales at which SIFT-features are extracted. This contribution improves the moment based controller from [8] by completely decoupling the sensitivity matrix for 4DOF based on a geometric constraint. Introducing additional constraints, leads to a sensitivity matrix for visual servoing in 6DOF with only four off-diagonal elements, that exhibits a robust and efficient visual control of the robot pose.

The paper is organized as follows: Section 2 provides a brief description of the characteristics of SIFT-features, followed by the definition of the set of visual features for visual servoing in 6DOF. Section 3 introduces the concept of a virtual centroid computed from weighted point features that eliminates the undesired couplings between the visual features and the degrees of freedom in 4DOF. In Section 4 the concept of weighted moments is extended to further decouple the sensitivity matrix for visual servoing in 6DOF. Section 5 compares the coupled and decoupled control scheme in 4DOF and 6DOF and analyzes their convergence properties in robotic experiments. The paper concludes with a summary and outlook on future work in section 6.

## 2 Moments of SIFT-Features

Scale invariant feature transformations (SIFT) introduced by Lowe [1] occur frequently in textured objects and remain identifiable and distinguishable despite changes in scale, orientation, illumination and affine transformation. Each feature contains a compact representation of the surrounding image region, the so called keypoint descriptor, which uniqueness distinguishes it from other features. Since these features are visible and robustly related across different views they are particularly suited for model free image based visual servoing. Once the image of the object is acquired from the demonstrated reference pose, the SIFT features are extracted automatically. For a detailed description of the automatic feature selection the interested reader is referred to [8]. A single SIFT-feature  $F_i$  contains four attributes, namely the pixel coordinates  $u_i$  and  $v_i$ , the canonical orientation of the keypoint  $\phi_i$  and its scale  $\sigma_i$ . In the following, the desired appearance of SIFT-features in the reference position is denoted by  $\hat{F}_i = (\hat{u}_i, \hat{v}_i, \hat{\phi}_i, \hat{\sigma}_i)$  and the current SIFT-features are denoted by  $F_i = (u_i, v_i, \phi_i, \sigma_i)$ .  $n$  denotes the number of SIFT features correspondences established between the current and reference view. Scale and keypoint orientation are ideal to control the distance to the object and the rotation around the optical axis as they prove

widely insensitive to translation and rotation along the other axes [8]. A camera rotation around its optical axis by  $\gamma$  induces an inverse rotation of equal magnitude of the keypoint orientations  $\phi_i$ . The averaged keypoint rotation  $f_\gamma$  regulates the camera rotation, whereas the translation along the camera axis is governed by visual feature  $f_z$  defined as the average scale.

$$f_\gamma = \frac{\sum_{i=1}^n \phi_i}{n}, \quad f_z = \frac{\sum_{i=1}^n \sigma_i}{n} \quad (2)$$

The point features  $u_i, v_i$  are first of all aligned with the camera orientation in the reference view according to the observed feature  $f_\gamma$ . The correction along  $\gamma$  is therefore defined as  $\Delta f_\gamma = \hat{f}_\gamma - f_\gamma$ . The visual features are rotated by  $\Delta f_\gamma$  such that the current SIFT feature locations  $u_i$  and  $v_i$  are aligned with the camera orientation in the reference view. The new SIFT feature locations  $u'_i$  and  $v'_i$  are determined as follows,

$$\begin{bmatrix} u'_i \\ v'_i \end{bmatrix} = \begin{bmatrix} \cos(\Delta f_\gamma) & -\sin(\Delta f_\gamma) \\ \sin(\Delta f_\gamma) & \cos(\Delta f_\gamma) \end{bmatrix} \cdot \begin{bmatrix} u_i \\ v_i \end{bmatrix}. \quad (3)$$

In the following the corrected pixel coordinates  $u'_i$  and  $v'_i$  are used for the computation of the remaining visual moments and for better comprehensibility are denoted as  $u_i$  and  $v_i$ . The moment based controller in [8] operates with the geometric centroid of point features for regulating translations along the  $x$ - and  $y$ -axis. However, the geometric centroid suffers from a sensitivity to motions along the remaining degrees of freedom. The visual features  $f_x$  and  $f_y$  capturing translations along the  $x$ - and  $y$ -axis, are expressed as the weighted aggregation of the matched SIFT feature locations

$$f_x = \sum_{i=1}^n w_i u_i, \quad f_y = \sum_{i=1}^n w_i v_i \quad (4)$$

By proper selection of the weights  $w_i$  attributed to individual point features  $(u_i, v_i)$  it is possible to decouple  $f_x, f_y$  from the remaining degrees of freedom. In order to complete the 6DOF visual control two additional visual features corresponding to the rotations about the  $x$ - and  $y$ - axis  $f_\alpha$  and  $f_\beta$  are defined. The visual features  $f_\alpha$  and  $f_\beta$  capture the perspective distortions of lines connecting pairs of SIFT features caused by rotations.

$$f_\alpha = \sum_{i=1}^n \sum_{j=i+1}^n \frac{(-\hat{v}_i - \hat{v}_j) \cdot \|\vec{p}_j - \vec{p}_i\|}{\sum_{k=1}^n \sum_{l=k+1}^n \|\vec{p}_k - \vec{p}_l\|} \quad (5)$$

$$f_\beta = \sum_{i=1}^n \sum_{j=i+1}^n \frac{(-\hat{u}_i - \hat{u}_j) \cdot \|\vec{p}_j - \vec{p}_i\|}{\sum_{k=1}^n \sum_{l=k+1}^n \|\vec{p}_k - \vec{p}_l\|} \quad (6)$$

The term  $\|\vec{p}_j - \vec{p}_i\|$  denotes the length of the line connecting the two pixels, which weighted by the factor  $(-\hat{v}_i - \hat{v}_j)$ . Its sign indicates whether the line is above or below the u-scan-line through the cameras principal point. The absolute magnitude of the weight increases with the vertical distance

from the image center. The visual feature  $f_\beta$  represents the equivalent effect of dilations and compressions of lines caused by rotations along the y-axis. In the following the image Jacobian is derived in order to analyze the remaining couplings. The centroid feature  $[f_x, f_y]^T$  behaves similar to a virtual point feature and the Jacobian is obtained by averaging the individual point feature Jacobians stated in [2]

$$\mathbf{J}_{f_x, f_y} = \frac{\sum_{i=1}^n \begin{bmatrix} \frac{\lambda}{z} & 0 & \frac{-u_i}{z} & \frac{-u_i v_i}{z} & \frac{\lambda^2 + u_i^2}{\lambda} \\ 0 & \frac{\lambda}{z} & \frac{-v_i}{z} & \frac{-\lambda^2 - v_i^2}{\lambda} & \frac{u_i \cdot v_i}{z} \end{bmatrix}}{n} \quad (7)$$

in which  $\lambda$  denotes the focal length and  $z$  denotes the distance between the camera and the feature plane. The main difference with respect to a point feature is the simplifying assumption that all SIFT-features share the same depth  $z$ . This assumption is reasonable as long as the depth of the scene is small compared to the distance to the camera. The image Jacobian for the visual feature  $f_\alpha$  is given by

$$\mathbf{J}_{f_\alpha} = \frac{\sum_{j=i+1}^n \frac{p_{ij} \cdot \hat{v}_{ij} \cdot \lambda}{\sqrt{8} \cdot z^2} \begin{bmatrix} 0 & 0 & -\frac{2 \cdot \lambda}{z} & -v_{ij} & +u_{ij} \end{bmatrix}}{\sum_{k=1}^n \sum_{l=i+1}^n \|\vec{p}_k - \vec{p}_l\|} \quad (8)$$

in which  $v_{ij} = (v_i - v_j)/2$ ,  $u_{ij} = (u_i - u_j)/2$ ,  $\hat{v}_{ij} = (\hat{v}_i + \hat{v}_j)/2$ ,  $\hat{u}_{ij} = (\hat{u}_i + \hat{u}_j)/2$  and the length  $p_{ij} = \|p_i - p_j\|$ . In the Jacobian for the analogous visual feature  $f_\beta$  the  $u$  and  $v$  components are interchanged. The resulting full 6DOF Jacobian matrix exhibits the following block structure.

$$\begin{bmatrix} \dot{f}_x \\ \dot{f}_y \\ \dot{f}_z \\ \dot{f}_\alpha \\ \dot{f}_\beta \\ \dot{f}_\gamma \end{bmatrix} = \begin{bmatrix} K_x & 0 & \tilde{K}_{xz} & \tilde{K}_{x\alpha} & \tilde{K}_{x\beta} & 0 \\ 0 & K_y & \tilde{K}_{yz} & \tilde{K}_{y\alpha} & \tilde{K}_{y\beta} & 0 \\ 0 & 0 & K_z & 0 & 0 & 0 \\ 0 & 0 & 0 & K_\alpha & \tilde{K}_{\alpha\beta} & 0 \\ 0 & 0 & 0 & \tilde{K}_{\beta\alpha} & K_\beta & 0 \\ 0 & 0 & 0 & 0 & 0 & K_\gamma \end{bmatrix} \cdot \begin{bmatrix} T_x \\ T_y \\ T_z \\ \omega_\alpha \\ \omega_\beta \\ \omega_\gamma \end{bmatrix} \quad (9)$$

The objective of the remainder of this paper is to define the dynamic weights  $w_i$  in equation 4 in such a way that the residual couplings in the off-diagonal elements  $\tilde{K}_{xz}$ ,  $\tilde{K}_{x\alpha}$ ,  $\tilde{K}_{x\beta}$ ,  $\tilde{K}_{yz}$ ,  $\tilde{K}_{y\alpha}$ ,  $\tilde{K}_{y\beta}$  disappear.

### 3 Decoupled visual servoing in 4DOF

With this objective in mind the features  $f_x$  and  $f_y$  are supposed to only depend on  $T_x$  and  $T_y$ , respectively. The following analysis for the proper dynamic weight focuses on the element  $\tilde{K}_{xz}$  related to the  $u$ -component. Notice, that for the decoupling of the visual feature  $f_y$  the same methodology is applied only using the  $v_i$  instead of the  $u_i$ . The undesired off-diagonal elements of the sensitivity matrix  $\tilde{K}_{xz}$  is eliminated if the dynamic weights  $w_i$  satisfy the constraint

$$\tilde{K}_{xz} = J_{\{f_x, z\}} = \sum_{i=1}^n w_i \frac{-u_i}{z} = 0 \quad (10)$$

For an arbitrary set of point features  $(u_i, v_i)$ , this constraint is violated for the geometric centroid calculation with equal

weights  $w_i = 1/n$ . In order to maintain the similarity with the conventional centroid we seek a minimal variation  $\Delta w_i$  of the original weights  $w_i = 1/n + \Delta w_i$  that satisfies the constraint. This optimal variation is obtained by minimizing the following cost function in conjunction with a Lagrange multiplier  $\lambda_1$ .

$$F = 1/2 \sum_{i=1}^n (w_i - \frac{1}{n})^2 + \lambda_1 \sum_{i=1}^n w_i u_i \quad (11)$$

In order to solve the optimization problem the partial derivative of 10 with respect to  $w_i$  and the Lagrange multiplier  $\lambda_1$ , not to be confused with the focal length  $\lambda$ , is computed as

$$\begin{aligned} \frac{\partial F}{\partial w_i} &= \left( w_i - \frac{1}{n} \right) + \lambda_1 \cdot u_i = 0 \\ \frac{\partial F}{\partial \lambda_1} &= \sum_{i=1}^n w_i u_i = 0 \end{aligned}$$

which in turn yields the least squares solution

$$w_i = \frac{1}{n} - \frac{u_i \bar{u}}{\sum_{i=1}^n u_i^2}, \quad \bar{u} = \sum_{i=1}^n \frac{u_i}{n} \quad (12)$$

Intuitively, the weight of SIFT features which pixels possesses the opposite sign as the geometric centroid  $\bar{u} = \sum_i u_i/n$  is increased, whereas those with the same sign are down-weighted. Notice, that by definition the weighted centroid is always located at the origin of the current image thus  $f_x = f_y = 0$ . However, the reference features  $\hat{f}_x = \sum_i w_i \hat{u}_i$  and  $\hat{f}_y = \sum_i w_i \hat{v}_i$  are no longer constant, but depend indirectly on the current image via the dynamic weights  $w_i$  and are therefore implicitly susceptible to motions along multiple degrees of freedom. In order to verify the decoupling of the weighted visual features  $f_x$  and  $f_y$  from a motion  $T_z$ , it is necessary to show that the weights for an identical set of feature points remain indeed independent of the distance  $z$ . The following considerations assume that the image plane is oriented parallel to the feature plane. The perspective projection of a world point on the image plane is given by  $u_i = x_i \cdot \frac{f}{z}$ , whereas the same point displaced by  $\Delta z$  is projected to  $u_{i, \Delta z} = x_i \cdot \frac{f}{z + \Delta z}$ . Assuming that the weights  $w_i$  fulfill the constraint  $\sum_{i=1}^n u_i \cdot w_i$ , the weighted sum of the feature points  $u'_i$  at distance  $z + \Delta z$  is thus given by:

$$\sum_{i=1}^n u_{i, \Delta z} \cdot w_i = \frac{z}{z + \Delta z} \cdot \sum_{i=1}^n u_i \cdot w_i = 0. \quad (13)$$

Due to fact that  $\frac{z}{z + \Delta z}$  is a proportional factor common to all features, the optimal weights for the first set of feature points also transform the virtual centroid of the second set of feature points to zero. Nonetheless, the weights are effectively down-scaled by the factor  $\frac{z}{z + \Delta z}$ . In order to render

the weights themselves and not only the centroid independent of the distance  $z$ , the weights are normalized according to

$$w_{i,norm} = w_i \cdot \frac{\sum_{k=1}^n |w_k|}{n} \quad (14)$$

Under the assumption, that the weighted sum of the features points is initially zero for the reference view, the weights  $w_{i,norm}$  transform the virtual centroid of every other view to the image center independent of the distance  $z$ . Figure 1 depicts two projections of the same features, in which the corresponding view points differ by a displacement  $\Delta z$ . The weights  $w_i$  are optimized according to the first set of feature points, but are also applied to weight the second set. Both virtual centroids are located in the image center even though the weighted feature points for the two sets differ. The example demonstrates that neither the weights nor the visual features  $f_x$  and  $f_y$  change with a camera motion along  $z$ . Notice, that in the weighted scheme the role of current and reference features are reversed. Typically, the reference features are constant and the current features change with the motion of the camera. However, in the weighted scheme the current centroid ( $f_x = 0, f_y = 0$ ) is constant per definition and always coincides with the principal point due to equation 10. Instead the reference features features  $\hat{f}_x, \hat{f}_y$  change over time as the weights  $w_i$  change with the current view, even though the point features  $\hat{u}_i, \hat{v}_i$  themselves are constant.

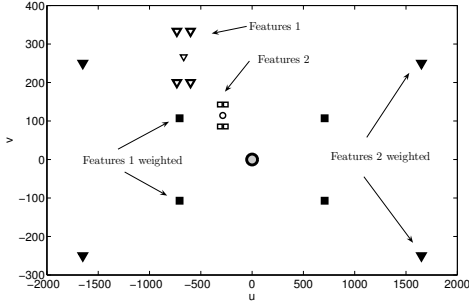


Figure 1: Virtual centroid for two sets of feature points at different distances between camera and feature plane

Figure 2 depicts the image projections of the conventional and weighted centroid u-components view as a function of the lateral task space error  $\Delta x$ . The two horizontal lines correspond to the constant conventional reference centroid and the constant weighted current centroid u-components. In this example, the reference image is deliberately chosen such that the majority of features is located in the right half plane. Therefore, the conventional centroid is shifted along the u-component by 120 units from the principal point. The current conventional centroid depends linearly on the lateral error and intersects the reference centroid for a zero lateral error  $\Delta x = 0$ . However, the offset and slope depend on the longitudinal distance be-

tween camera and feature plane. Current and reference features interchange their role for the weighted scheme, in that the former remains constant and the later changes with the lateral error in a nonlinear fashion. In this case the dependency of the centroid on the lateral error remains the same independent of the longitudinal pose error in  $z$ . Again, the reference and current centroid intersect for zero lateral error. Even though the slopes for the weighted and conventional centroid exhibit opposite sign, the actual image error  $\hat{f}_x - f_x$  has the same sign for both schemes. The figure also reveals the slight asymmetry of the weighted reference centroid for positive and negative lateral errors. For large positive lateral errors the image error even increases slightly with decreasing lateral error. This asymmetry is caused by the inhomogeneous distribution of point features  $u_i$ , which in turn effects the adaptation of the weight factors through equation 12. It should be noted, that the asymmetry and slope inversion do not effect the stability or convergence of the visual control and only occur if the point feature distribution in the reference image is significantly skewed.

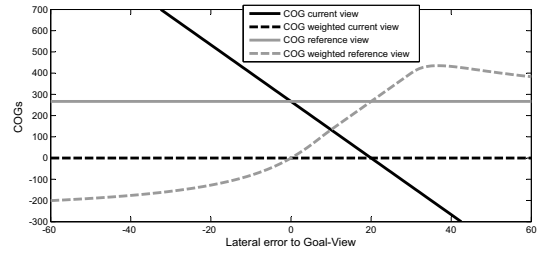


Figure 2: Centroid u-component as a function of the lateral error for the conventional and weighted centroid in the current and reference view

## 4 Decoupled visual servoing in 6DOF

Section 3 discussed a fully decoupled sensitivity matrix for visual servoing in 4DOF. This section is aimed at extending the decoupling to visual servoing in 6DOF. This requires that the features  $f_x$  and  $f_y$  do not only become independent on the motion  $T_z$  but also on the rotary motions  $\omega_\alpha$  and  $\omega_\beta$ . Again, the constraints emerge from an analysis of the Jacobian in equation 7. The optimization problem for decoupling the visual feature  $f_x$  from the motions  $T_z$  and  $\omega_\alpha$  according to the Jacobian in equation 5 is stated as follows:

$$F = \frac{1}{2} \sum_{i=1}^n \left( w_i - \frac{1}{n} \right)^2 + \lambda_1 \cdot \sum_{i=1}^n w_i \cdot u_i \quad (15)$$

$$+ \lambda_2 \cdot \sum_{i=1}^n w_i \cdot u_i \cdot v_i$$

Theoretically, it would be possible to cancel the component  $\sum_{i=1}^n w_i \cdot (\lambda^2 + u_i^2)$  of the Jacobian related to the motion  $\omega_\beta$  as well. If this additional constraint is included,

the minimization problem in equation 15 is algebraically still solvable. Unfortunately, the inclusion of this constraint substantially reduces the sensitivity of the feature  $f_x$  to its associated motion  $T_x$ , resulting in a deterioration of the visual control. Therefore, the weight optimization problem only includes constraints for the cancellation of the elements  $\tilde{K}_{x\alpha}$  and  $\tilde{K}_{y\beta}$  in the sensitivity matrix in equation 9. The cost function  $F$  is partially differentiated with respect to  $w_i$  and the Lagrangian multipliers  $\lambda_i$ .

$$\begin{aligned}\frac{\partial F}{\partial w_i} &= w_i - \frac{1}{n} + \lambda_1 \cdot u_i + \lambda_2 \cdot u_i \cdot v_i \\ \frac{\partial F}{\partial \lambda_1} &= \sum_{i=1}^n w_i \cdot u_i \quad \frac{\partial F}{\partial \lambda_2} = \sum_{i=1}^n w_i \cdot u_i \cdot v_i\end{aligned}$$

Changing to vector notation by substituting  $k = (\frac{1}{n}, \dots, \frac{1}{n})$ ,  $u = (u_1, \dots, u_n)$  and  $p = (u_1 \cdot v_1, \dots, (u_n \cdot v_n))$  a set of linear equations is obtained:  $w + \lambda_1 u + \lambda_2 p = k$ , with the constraints  $w \cdot u = 0$ ,  $w \cdot p = 0$ . In order to determine the Lagrangian multipliers, the weights  $w$  are eliminated by multiplying with the transpose of  $u$  and  $p$ . This results in a system of linear equations, which can be solved in closed form.

$$\begin{bmatrix} u^T u & u^T p \\ p^T u & p^T p \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} u^T k \\ p^T k \end{bmatrix} \quad (16)$$

The corresponding weights of the point features are determined as follows  $w = k - \lambda_1 \cdot u - \lambda_2 \cdot p$  and subsequently normalized according to equation 14. However, some residual coupling remains through the non-zero off-diagonal elements ( $\tilde{K}_{x\beta}$ ,  $\tilde{K}_{y\alpha}$ ,  $\tilde{K}_{\alpha\beta}$ ,  $\tilde{K}_{\beta\alpha}$ ) in the Jacobian. These terms capture the effect of a rotation  $\alpha$  along the  $x$ -axis on the motion of the  $v$ -components of point features, and vice versa a rotation  $\beta$  along the  $y$ -axis on the  $u$ -component.

## 5 Experimental results

The proposed scheme is evaluated for a free moving camera in virtual reality and in experiments on a 5DOF KATANA manipulator with an eye-in-hand configuration. This section reports results for the 4DOF case on the manipulator and the 6DOF case in virtual reality as the KATANA robot is restricted to 5DOF. For the evaluation of the decoupled visual features for 4DOF visual servoing, the manipulator is dislocated from the reference pose by an initial displacement  $\Delta x = 20\text{mm}$ ,  $\Delta z = 40\text{mm}$  and  $\Delta \gamma = 25^\circ$ . Substantially larger displacements are not feasible in the experiments due to the restricted workspace of the KATANA manipulator and the eye-in-hand constraint of keeping the object in view of the camera. Figure 3 compares the evolution of the task space error and the image error for the original coupled controller (left) and the decoupled controller (right). In case of the coupled controller the  $z$ -error introduces a dynamic shift of the current feature  $f_x$ . Although the image space error itself is compensated after 12 iterations the corresponding residual task space error in  $x$  remains until complete cancellation of the  $\Delta z$  after about 25

iterations. The decoupled controller eliminates the impact of  $T_z$  on feature  $f_x$  such that image and task space error in the  $x$ -component disappear simultaneously after 12 iterations. Even though there is no initial displacement along the  $y$ -axis, the inherent coupling with  $T_z$  induces an undesired motion in task space  $\Delta y \approx 5\text{mm}$ . Again, the decoupled controller eliminates this detrimental disturbance and  $\Delta y$  is not effected by the motions  $T_x$  or  $T_z$ . The residual task space error of the decoupled controller is less than 0.5mm for the position and  $0.5^\circ$  for the rotation. Figure 4 shows the task space error for visual servoing for 6DOF in a virtual reality simulation. Notice, the substantial coupling among the different degrees of freedom in the task space for the conventional controller (left). The translational motions in  $x$ ,  $y$  and  $z$  demonstrate a significant overshoot caused by the couplings of the conventional centroids  $f_x, f_y$  with  $T_z$ ,  $\omega_\alpha$  and  $\omega_\beta$ . The task space errors in  $x$  and  $z$  initially increase and only start to converge after stabilization of the other errors. For the weighted centroid control (right) with the partially decoupled Jacobian the disturbances are significantly reduced and the six pose errors converge smoothly and largely independent of each other. The residual overshoot in the  $x$  and  $y$  components is caused by the remaining coupling with  $\omega_\beta$  and  $\omega_\alpha$ , respectively. The results clearly demonstrate that the weighted features result in a more favorable task space motion of the camera.

## 6 Conclusions and Future Work

This contribution presented a novel approach for visual servoing based on decoupled moments of SIFT-Features. The properties of SIFT-features render the approach universally applicable for manipulation of daily-life objects that exhibit some texture. SIFT-features for visual servoing applications offers the further advantage that object recognition and pose alignment of the manipulator rely on the same representation of the object. For 4DOF visual servoing a set of completely decoupled visual features is introduced, that results in robust and independent convergence of the corresponding task space errors. Problems of the classical Jacobian based visual servoing scheme like the camera retreat problem and local minima are resolved. A novel sensitivity matrix for 6DOF visual servoing is introduced, which has only four off-diagonal couplings between the visual features and the degree of motion. The visual control with the novel causes the pose errors to converge largely independent of each other resulting in a smoother task space motion of the camera. In future work we intend to accomplish large view visual servoing. Due to the limited visibility and perceptibility of SIFT-features across different views, it becomes necessary to introduce additional intermediate reference views to navigate across the entire view hemisphere. Future research is dedicated to identify intermediate views and to define an appropriate cost function that governs the transition between overlapping reference views in a stable, robust and time-optimal manner.

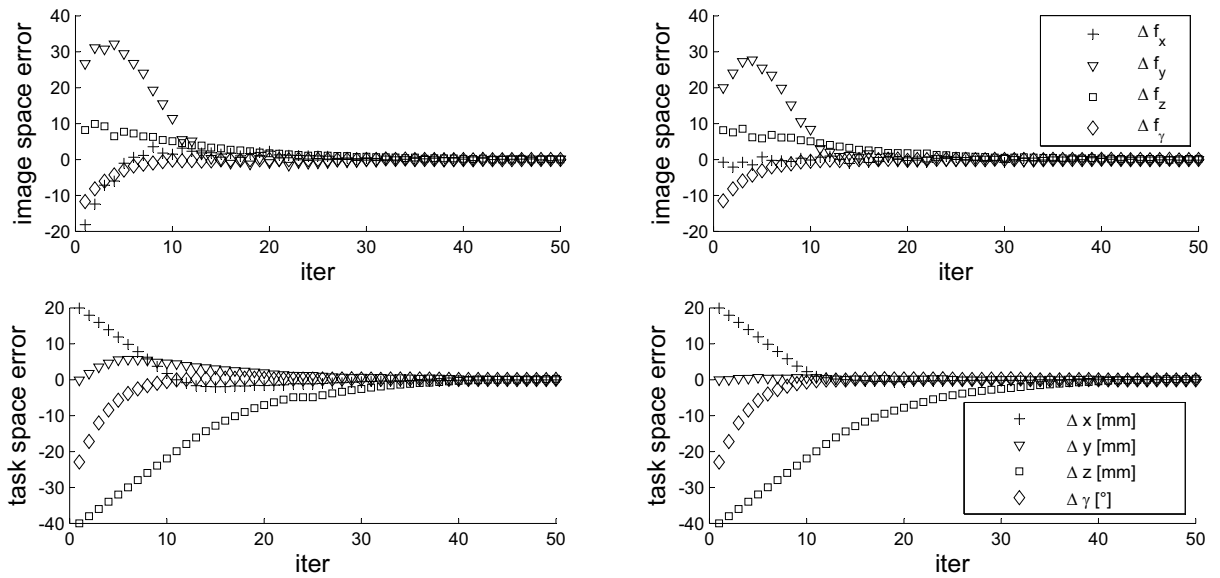


Figure 3: Left top & bottom) Image and task space error for visual servoing in 4DOF, Right top & bottom) Image and task space error for decoupled visual servoing in 4DOF

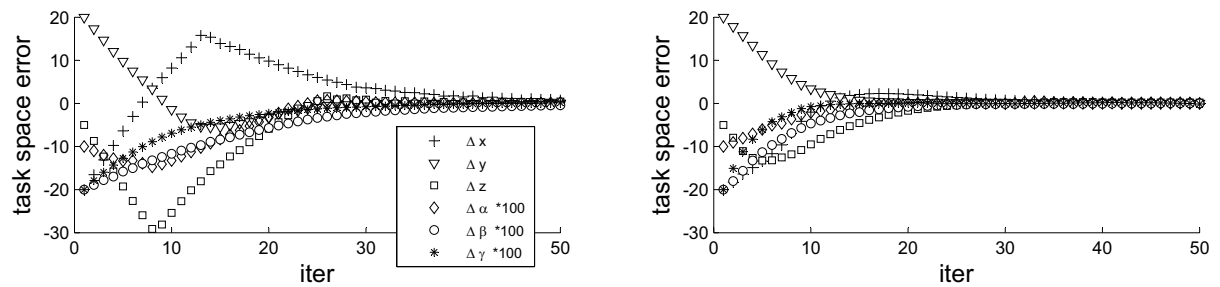


Figure 4: Left ) Task space error for visual servoing in 6DOF, Right) Task space error for decoupled visual servoing

## Acknowledgments

This research has been supported by the German Research Foundation (DFG) through the collaborative research center "Computational Intelligence" (SFB 531)

## References

- [1] D. G. Lowe, *Distinctive image features from scale-invariant keypoints*, International Journal of Computer Vision, 60, 2 (2004), pp. 91-110.
- [2] S. Hutchinson, G. D. Hager, I. P. Corke, *A Tutorial on Visual Servoing Control*, IEEE Transactions on Robotics and Automation, vol.12, pp.651-668, 1997.
- [3] A.C Sanderson and L.E Weiss, *Image-based visual servo control using relational graph error signals*, Proceedings of the IEEE, pp. 1074-1077, 1980
- [4] K. Deguchi, *Optimal motion control for image-based visual servoing by decoupling translation and rotation*, in IEEE Int. Conf. Intell. Rob. and Syst., vol.2, pp. 705-711, Oct. 1998.
- [5] Francois Chaumette, Ezio Malis: *2 1/2 D Visual Servoing: A Possible Solution to Improve Image-Based and Position-Based Visual Servoings*, ICRA 2000.
- [6] O. Tahri and F. Chaumette, *Point-based and region-based image moments for visual servoing of planar objects*, IEEE Transactions on Robotics and Automation, Vol. 21, No. 6, 2005.
- [7] A. Shademan and F. Janabi-Sharifi, *Using scale-invariant feature points in visual servoing*, Optomechatronic Sensors, Actuators, and Control, 2004, Proceedings of the SPIE, Volume 5603, pp. 63-70.
- [8] F. Hoffmann, T. Nierobisch, T. Seyffarth and G. Rudolph, *Visual Servoing with Moments of SIFT Features*, IEEE International Conference on Systems, Man, and Cybernetics, SMC 2006