High Speed Visual Motion Control Applied to Products with Repetitive Structures
Jeroen de Best, René van de Molengraft and Maarten Steinbuch

Abstract—This paper focusses on direct dynamic visual servoing at high sampling rates in machines used for the production of products that consist of equal features placed in a repetitive pattern. The word “direct” means that the system at hand is controlled on the basis of vision only. More specifically, the motor inputs are driven directly by a vision-based controller without the intervention of low level joint controllers. The product in view consists of a repetitive pattern, which is used as an encoder purely on the basis of vision. The considered motion task is to position the repetitive structure in order to align the tool with respect to the features. The vision based controller is designed using classical loop shaping techniques. Robustness with respect to imperfections of the repetitiveness is investigated. The combination of fast image processing and a Kalman-filter based predictor results in a 1 kHz visual servoing setup capable of exploiting the repetitive pattern as an encoder with an accuracy of 2 μm. The design approach is validated on an experimental setup.

I. INTRODUCTION

Many production processes take place on repetitive structures, for example in inkjet printing technology where droplets are placed in a repetitive pattern or in pick and place machines used in the production of solar cell arrays or in the manufacturing of LCD displays. In each of these processes one or more consecutive steps are carried out on the particular features of the repetitive structure to create the final product. Such production machines often consist of a tool, for example a printhead, and a stage or carrier on which the repetitive structure is to be processed. Key to obtaining a high product quality is to position the tool with respect to imperfections of the machine and product will be dealt with the features. In current industrial practice, local position sensors such as motor encoders are used to measure the tool position $x_t$ and the position of the stage $x_o$ separately as shown in Fig. 1(a). Often the absolute reference points of these measurements do not coincide. This is referred to as an indirect measurement of $x_t - x_o$. Using such local measurements in a closed-loop control approach leads to a collocated control design. The final accuracy of alignment in this case is directly dependent on the following machine properties:

- geometric accuracy of the mechanical construction,
- stiffness of the mechanical construction and
- thermal stability of the machine.

Furthermore, the final accuracy of alignment also relies on assumptions with respect to the repetitive structure:

- infinitely stiff connection between the supporting stage and the repetitive structure,
- constant and known alignment of the repetitive structure with respect to the actuation axes,
- infinite stiffness of the repetitive structure,
- constant and known pitch between successive features of the repetitive structure and
- thermal stability of the repetitive structure.

Some of the above issues may result in reproducible errors, especially geometric imperfections and dynamic flexibilities. These errors usually require expensive mechanical measures, with respect to both the machine itself and the repetitive structure. As an example, a priori unknown pitch variations in the repetitive structure will limit the attainable accuracy and prevent the use of absolute motion setpoints in high-accuracy applications. Such limitations due to imperfections can be overcome by adopting a different design paradigm where a camera is used for a direct measurement of the relative position between product and tool. In this paradigm the imperfections of machine and product will be dealt with by non-collocated feedback control. This paper exploits the potential of this approach by constructing a feature-based position measurement on the basis of camera images, such that motion setpoints can be defined from feature to feature without knowing the exact absolute position of the features beforehand, while achieving a high positioning accuracy.

Controlling a mechanical system by means of camera measurements is referred to as visual servoing [8], [10], [15]. Kinematic visual control [2], [8], [10] assumes rigid body dynamic behavior and can not be used in our dynamic, non-collocated
control approach. Indirect dynamic visual control [3], [4], [5], [18], [19] does account for dynamic effects but still relies on the presence of collocated position sensors. Therefore, we will adopt the concept of direct dynamic visual control [11], [12], [16] with eye-in-hand camera configuration. The main contributions of this work compared to the above literature are the following:

- feature-based position sensing enabling a direct dynamic visual control paradigm that is robust against machine imperfections and deviations in the pitch between successive features of the repetitive structure,
- stability analysis of the controlled system with respect to the allowable deviations in the pitch between features of the repetitive structure, and
- the validation of the proposed methods on a practical direct visual control setup using a commercially available and cost-effective camera with a 1 kHz update rate.

The rest of the paper is organized as follows. In Section II the measurement principle to create a feature-based position sensor using the repetitive structure in combination with a camera will be given followed in Section III by the design of a model-based predictor that is needed when moving at high velocities and for speeding up the image processing steps. The image processing algorithm will be discussed in Section IV. The practical setup used for validation of the proposed algorithm will be described in Section V, the system identification in Section VI, and Section VII will discuss the final integration. The stability analysis in combination with the controller design will be given in Section VIII, followed by the experimental validation in a closed-loop visual servoing control setting. Finally, conclusions and suggestions for future work will be given.

II. MEASUREMENT PRINCIPLE

Within this research we focus on machines used for the production of structures that inherently consist of identical features placed in a repetitive pattern such as OLED displays, see Fig. 2(a). At this point we restrict the focus of the paper to a one-dimensional repetitive structure for ease of explanation. In many manufacturing machines, production steps are carried out row by row or column by column, so in practice we need a two-dimensional position measurement. In our case the second dimension is however restricted by the field of view of the camera. The focus in this paper will be on the feature-based position measurement along the repetitive structure in order to apply feature-based control. For now we will consider the features to be circular objects as shown in Fig. 2(b), with a diameter of \( D \) pixels. The height and width of the image captured by the camera are \( I_h \) and \( I_w \) pixels, respectively, whereas the repetitiveness is characterized by the pitch between the features, which is denoted by \( P \). The number of features that are completely within the field of view for the presented method must be at least two, and must be located at different sides of the center of the field of view. Therefore, the required field of view is determined by the pitch of the repetitive structure together with the feature size. In the case of a different pitch either the height of the camera can be adjusted which influences the resolution, or a different sized area of pixels can be read out leading to different acquisition and processing times. Within the image, the horizontal pixel positions \( d_l \) and \( d_r \) of the two features that are located nearest to the opposite sides of the image center are measured, see Fig. 2(b). These features are labeled \( y_c(k) \) and \( y_c(k) + 1 \), with \( y_c(k) \in \mathbb{Z} \), irrespective of the mutual pixel distance. Here, the time step is indicated by \( k \). The measured position \( y_c(k) \) that will be used in the closed-loop visual control setting is now given by

\[
y_c(k) = y_c(k) + y_f(k),
\]

with \( y_c(k) \) being the coarse position, i.e., the integer feature label. The fine position \( y_f(k) \) is the linear interpolation between the left and right feature label and is calculated as

\[
y_f(k) = \frac{0.5I_w - d_l(k)}{d_r(k) - d_l(k)} \leq 1.
\]

The output \( y_c(k) \) indicates the position of the center of the image in features units. So, \( y_c(k) = 1.0 \) indicates that the feature labeled 1 is exactly in the center of the image, whereas \( y_c(k) = 0.5 \) indicates that the center of the image is exactly between the features with labels 0 and 1. Therefore, we define the feature label, denoted by \( f \), as a measurement unit. Comparing this approach with a classical incremental encoder an important difference is that with this method we can interpolate the position between features. Since we are only interested in the position of the feature with respect to the center of the image we have created a relative incremental encoder. Imperfections of the pitch \( P \) cause this measurement to become piecewise linear, i.e., the gain of the process varies along the structure. Section VIII will discuss this in detail, where the feature unit \( f \) also appears.

III. MODEL-BASED PREDICTION

Key to obtaining the correct position is determining the value of \( y_c(k) \) within the field of view. When for example, the velocity is one pitch per sample the camera will record identical images every time step. Based on that information
only, the measurement $y_v$ as described in the previous section gives the same value if $y_v$ is not incremented, i.e., we measure a velocity of zero while the structure is moving with the high velocity of one pitch per sample. If the velocity is increased further aliasing effects cause the features to appear to move slowly in the wrong direction. To tackle the problem of incrementing the value of $y_v$, a model-based solution will be applied. More specifically, we will design a stationary Kalman filter [14], from which the one step ahead prediction will be used to estimate the value of $y_v$ for the next time step. Moreover, next to incrementing the value of $y_v$, the one step ahead prediction will also be used to estimated where the features will be located in the field of view in the next time step. Therefore, we will model the input-output behavior of the motion drive carrying the repetitive structure as a mass system, see Fig. 3. The input of the system $u$ is the force applied to the mass and the output is the position $y$. The state space representation of the discrete time system is given by

\[
\begin{align*}
\hat{x}_{k+1} &= A \hat{x}_k + Bu(k) + w(k) \\
y_k &= C \hat{x}_k,
\end{align*}
\]

where $\hat{x} = (y \ y_v)^T$ is the state vector containing the position $y$ and the velocity $\dot{y}$, with $\hat{x}(0) = \hat{x}_0$, $u$ is the known applied force and $w$ is the process noise, being the unmodeled forces. The matrices $A$, $B$ and $C$ are the system, input and output matrices, respectively. The specific matrices for our model are straightforward, and expanded, time-delay versions are given in Section VI by (18). In this section a stationary Kalman filter will be given that estimates the output $y$ given the known input $u$ and the measurement $y_v$ given by

\[
y_v(k) = C \hat{x}_k + v(k),\]

where $v$ represents the measurement noise. For the process and measurement noise we assume

\[
E(B^T w^2) = E(w^2) = Q_w, E(v^2) = Q_v, E(wv) = 0,
\]

where $E(\cdot)$ is the expected value operator. The Kalman filter consists of a 1) prediction step

\[
\hat{x}(k+1|k) = A \hat{x}(k|k) + Bu(k),
\]

and a 2) 'no steps ahead' correction step

\[
\hat{x}(k|k) = \hat{x}(k|k-1) + M(y_v(k) - C \hat{x}(k|k-1)),
\]

where $M$ is the Kalman gain obtained from solving the steady state Riccati equation. Here, the prediction of the state at time

\[
\begin{align*}
\hat{x}(k-1|k-1) &= \hat{x}(k-1|k-2) + M \ldots \\
\hat{x}(k|k-1) &= A \hat{x}(k-1|k-1) + Bu(k-1)
\end{align*}
\]

step $k+1$ on the basis of measurements up to time step $k$ is denoted by $\hat{x}(k+1|k)$. The two steps are graphically depicted in Fig. 4. Combined the prediction and correction step lead to

\[
\hat{x}(k+1|k) = A(I - MC) \hat{x}(k|k) + Bu(k) + AMy_v(k).
\]

The one step ahead output prediction uses this one step ahead state prediction and is given by

\[
y_v(k+1|k) = Cy_v(k+1|k),
\]

where $y_v(k+1|k)$ is the estimate of $y(k+1)$ on the basis of measurements up to time step $k$. This prediction is used to get an estimate $\hat{y}_v$ of the position of the repetitive structure in the next time step $k+1$:

\[
\hat{y}_v(k+1|k) = \lfloor \hat{y}(k+1|k) \rfloor,
\]

where $\lfloor \cdot \rfloor$ is the floor function, which rounds $\hat{y}(k+1|k)$ to the nearest lower integer. In the prediction step explained above it is assumed that the pitch $P$ is constant. If this is not satisfied we cannot associate the right label to the feature if $n \Delta P > P$, where $\Delta P$ is the deviation of the nominal pitch $P$ and $n$ is the number of features that has passed within one time step.

IV. FAST IMAGE PROCESSING IMPLEMENTATION

This section discusses the image processing algorithm used for detecting the pixel positions $d_l$ and $d_r$, which in our case comprises thresholding and calculating the center of gravity. In case of more complex features, image registration techniques based on correlation or hough transform could be used which are more computationally demanding. At this point we will introduce search areas around each of the features within the field of view with a width and height of $S_w$ and $S_h$ pixels respectively, such that only one feature is completely present within one search area as shown in Fig. 5. In our case we have chosen $S_w = S_h = P$. The goal is to search for only one feature within one search area such that labeling implementations to distinguish between multiple features in the image processing algorithms, which cause overhead, can be eliminated. Furthermore, we introduce $\hat{d}$, which is a pixel position estimate of the feature that is closest to the image center. By using a better prediction the search area can be reduced, which in turn again leads to a smaller computation time of the image processing algorithms. Naturally, the size of the search area depends on i) the feature size $D$ ii) the variation of the feature position and iii) the quality of the prediction $\hat{d}$. Naturally, this size should be larger if i) the feature size is large, ii) the variation of the feature position is large or iii) the prediction quality is low. The size of the

![Fig. 3. Single mass system. The input is denoted by $u$. The output, denoted by $y$, is the position of the repetitive structure measured by the camera.](image-url)
features and the variation of the position are characteristics of the machine which cannot be altered. However, the estimate \( \hat{d} \) can be influenced. The pixel position estimation \( \hat{d} \) can be obtained from the one step ahead prediction discussed in the Section III as follows

\[
\hat{d}(k+1|k) = \begin{cases} 
0.5I_w + (1 - (\hat{y}(k+1|k) - \hat{y}_c(k+1|k)))P & \text{if } \hat{y}(k+1|k) - \hat{y}_c(k+1|k) \geq 0.5 \\
0.5I_w - (\hat{y}(k+1|k) - \hat{y}_c(k+1|k))P & \text{if } \hat{y}(k+1|k) - \hat{y}_c(k+1|k) < 0.5.
\end{cases}
\]

(12)

Given this estimate together with the search area, the position of the feature within the search area is calculated. This is done as follows.

First, the image is thresholded within the search area. Global optimal thresholding is performed using Otsu’s thresholding method [7], which determines the optimal threshold level \( TH \). The thresholding is done while reversing salient intensities as follows

\[
T(i,j,k) = \begin{cases} 
TH - I(i,j,k) & \text{if } I(i,j,k) \leq TH, \\
0 & \text{if } I(i,j,k) > TH.
\end{cases}
\]

(13)

Here, the image data is denoted by \( I(i,j,k) \), with indices \( i \in \{s_l, \ldots, s_l + S_h\} \), \( j \in \{s_l(k), \ldots, s_l(k) + S_w\} \) indicating the pixel elements and \( k \) indicating the time step. The position \( (s_l, s_l(k)) \) indicates the top left corner of the search area, see Fig. 5. This position is given by \( s_l(k) = \hat{d}(k) - 0.5S_w \) and \( s_l = 0.5(I_h - S_h) \). Therefore, we assume that the \( t_x \) positions of the features only vary within \( S_h - D \) with respect to the center of the image in \( t_x \) direction. As a result, we can also measure the \( t_x \) position within a limited range. This position can be used in a feedback loop to keep the features within the field of view. However in the remainder we will focus on the horizontal position measurement. The resulting thresholded image is given by \( T(i,j,k) \).

Secondly, the center of gravity in the \( t_y \) direction within the search area of the thresholded image \( T(i,j,k) \) is calculated as

\[
d(k) = \frac{s_l + S_h}{s_l} \frac{s_l(k) + S_w}{s_l(k)} \sum_{i=s_l}^{s_l+k} \sum_{j=s_l(k)}^{s_l(k)+S_w} iT(i,j,k).
\]

(14)

If \( d(k) \geq 0.5I_w \) we have found the center of the feature at the right of the center of the image and we call this distance \( d_r(k) = d(k) \). From Fig. 5 it can be seen that \( d_r(k) \) can be slightly different from \( \hat{d}(k) \) indicating the estimation error. Next, if the center of the feature is found at the right of the center of the image, the center of another feature is searched for at the left of the image center with an estimate given by \( \hat{d}_l(k) = d_r(k) - P \). Conversely, if \( d(k) \) was found to satisfy \( d(k) < 0.5I_w \) we have found the left feature with position \( d_l(k) = d(k) \) and we search for the right feature with an estimate given by \( \hat{d}_r(k) = d_l(k) + P \). We end up having two positions \( d_r(k) \) and \( d_l(k) \).

V. EXPERIMENTAL SETUP

The setup that is used for experimental validation is depicted in Fig. 6. It consists of two stacked linear motors forming an xy-stage. The data-acquisition is realized using an EtherCAT [13] data-acquisition system, where DAC, I/O, and ADC modules are installed, respectively, to drive the current amplifiers of the motors, to enable amplifiers and to measure the position of the xy-stage on the motor side. Hence, this position is only used for comparison and is not used in the final control algorithm as such. A Prosilica GC640M high-performance machine vision camera [17] with Gigabit Ethernet interface (GigE Vision) which supports jumbo frames and is capable of reaching a frame rate of 197 Hz full frame (near VGA, 659×493) is mounted above the stage. The GigE interface allows for fast frame rates and long cable lengths. The captured images are monochrome images with 8 bit intensity values. To obtain a frame rate of 1 kHz we make use of a region of interest (ROI); we read out only a part of the image sensor as large as 80×80 pixels. The GigE network controller is able to process frame sizes up to 9200 bytes. The frame size is the number of bytes per packet and the...
larger the frame size, the less the CPU will be loaded due to the processing of incoming packets. Since 80×80 pixels results in 6400 bytes fitting into a single packet, only one packet is needed. This region is centered at the principal point located near the center of the sensor. The used objective is a Fujinon DF6HA-1 lens, specified with a focal length \( f \) of 6 mm. According to the datasheet, the camera has a Micron MT9V403 sensor with a square pixel size \( p \) of 9.9 \( \mu \)m. The camera is calibrated [9] with a calibration grid, which first shows that distortion is neglectable due to the small ROI around the principal point. Therefore the pinhole camera model, which maps the metric positions \( X_i, Y_i, Z_i \) of the repetitive pattern to the image coordinates \( u_i, v_i \), can be applied

\[
\begin{align*}
    u_i &= \frac{f}{pZ_i} X_i \quad (15) \\
    v_i &= \frac{f}{pZ_i} Y_i. \quad (16)
\end{align*}
\]

The focal length \( f' \) which is expressed in pixel units, i.e., \( f' = \frac{f}{p} \) was calibrated and found to be 623 ± 11 pixels. The height of the camera \( h \) was also calibrated and is 0.1132 m. Therefore one pixel in the image coincides with 181.7 \( \mu \)m ± 3.3 \( \mu \)m on the repetitive structure. With a height \( h \) of 0.1132 m between the camera and the stage and a focal length \( f' \) of 623 pixels we can calculate the resulting field of view as

\[
\frac{I_{wh}}{f'} \times \frac{I_{hw}}{f'}, \quad (17)
\]

which in this case is 14.5 × 14.5 mm. Note that the region of interest is large enough such that at least two features are completely in the field of view. The exposure time is set to its minimum, which is 10 \( \mu \)s. The illumination is realized using power LEDs and set such that all pixel values are within the dynamic range, i.e., between 0 and 255, to avoid clipping. The data-acquisition is integrated in a Linux environment running a 2.6.28.3 preemptible low-latency kernel and the real-time executable is built using the real-time workshop (RTW) of Matlab/Simulink. The repetitive structure consists of circular black dots with a diameter \( D \) of 2 mm and a nominal pitch \( P \) of 4 mm.

**VI. SYSTEM IDENTIFICATION**

For the horizontal direction frequency response functions (FRFs) are measured. Two different FRFs are measured: one from the motor input \( u \) to the position output \( y_{mot} \) using the motor encoder (collocated control) and one from the motor input \( u \) to the position output \( y_{cam} \) using the camera with the position measurement as described in the previous sections (non-collocated control). The pitch was constant during the measurement and is 4 mm and the sampling frequency was 1 kHz in both cases. In the ideal case, both \( y_{mot} \) and \( y_{cam} \) would represent a measurement of the product position \( y \). The result is given in Fig. 7, where the camera measurement is scaled by the pitch for comparison. Different dynamics are present if the position measurement from the camera is used instead of the motor encoder. In the case of using the camera as sensor all relevant dynamics are measured including vibrations caused by the limited stiffness of the frame. Furthermore, it can be seen that different time delays are present when using the camera in the feedback instead of the motor encoder. The time delay when using the camera is larger, due to the necessary image acquisition and image processing time. When the camera is used the time delay is 3.5 ms, while the readout of the sensor according to the datasheet takes 816 \( \mu \)s, the image transport takes 1.1 ms, the image processing takes 60 \( \mu \)s, the data acquisition for driving the motors takes 1 ms and finally the zero order hold effect results in 0.5 ms delay.

At 70 Hz a resonance of the system is observed caused by the finite stiffness of the camera mounting. For frequencies below 50 Hz the two FRFs are quite similar and can be approximated by a single mass system as depicted in Fig. 3. The time delay is incorporated in the discrete time model of (3) and (4) by adding three additional states \( \bar{x} = (y(k - 3) \quad y(k - 2) \quad y(k - 1) \quad y(k) \quad \dot{y}(k))^T \) to model a 3 ms delay. The zero-order hold (ZOH) effect introduces another 0.5 ms delay, which in total leads to the 3.5 ms delay present in the system. The system, input and output matrices \( A, B \) and \( C \), are given as follows

\[
A = \begin{pmatrix}
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0
\end{pmatrix},
B = \begin{pmatrix}
0 \\
0 \\
0 \\
0.5T^2/m \\
T/m
\end{pmatrix},
C = \begin{pmatrix}
1 & 0 & 0 & 0 & 0
\end{pmatrix},
\]

where \( m \) is the dimensionless mass (including all motor and amplifier gains) of the system and \( T = 0.001 \) s is the sampling time. The value of \( m \) is dimensionless and is estimated to 0.184. For that the measured FRFs are approximated by a
single mass system $G(j \omega) = -\frac{1}{m \omega^2}$. The values of $G(j \omega)$ and $\omega$ are known, so $m$ can therefore be estimated.

VII. INTEGRATION

The integration of all blocks within the control scheme is depicted in Fig. 8. The steady-state Kalman filter is designed using the matrices $A$, $B$, and $C$ given in the previous section together with

$$Q_w = BB^T E(w^2) = BB^T q_w, \quad Q_v = E(v^2) = q_r,$$

with $q_r$, the variance of the measurement noise $v$. This variance is determined by measuring the position while the system is at a standstill. The noise of this measurement has a variance $2.8 \cdot 10^{-8}$, with $f$ being the unit features as explained in Section II. The value $q_w$, which is the variance of the unmodeled input $w$, is used as a tuning variable such that the innovation signal, defined as $y_r(k) - \hat{y}(k|k-1)$, is minimized. The optimal value for the variance of $w$ is found to be 0.07 V. We assume that there is no uncertainty in the delay of the system. In Fig. 8 the controller $K$ is connected to the system $G$ with motor input $u$ and image output $I$. This image is processed in the image processing block $IP$ using the estimates $\hat{y}_c(k)$ and $\hat{d}(k)$. These estimates are the previous outputs of the Kalman filter $KF$. The Kalman filter is used only for i) incrementing the value of $y_c(k)$ and ii) estimating the position of the feature closest to the image center in the next time step. The filtered position output of the Kalman filter $\hat{y}(k|k)$ is not used for feedback since in that case the dynamics of the Kalman filter would contribute to the dynamics to be controlled, i.e., by using $\hat{y}(k|k)$ for feedback, the plant as seen by the controller is the series connection of the system and the Kalman filter, see Fig. 8. Therefore, based on the Kalman filter tuning, a different plant is observed by the controller. In our case, we use the output $y_c$ of the system as feedback, such that the dynamics of the Kalman filter do not have to be considered in the controller design.

VIII. STABILITY ANALYSIS

When going from one feature to the next, a different pitch $d_r - d_l \neq P$ is considered after each transition in (2). This results in a switching gain of the system. In this section the switching behavior will be modeled. Furthermore, a stability analysis will be given that can be carried out to ensure stability of the closed-loop system if a single controller is used to control the switching system.

We model the switching behavior of the system $G(z) = R(z)I(z)$ as given in Fig. 9. The first system $R(z)$ relates the applied forces to metric velocities measured in m/s. The gain $1/P$ converts these metric velocities into feature-based velocities measured in $f/s$ according to the momentary pitch $P$. After integration the feature-based position $y$ is obtained.

![Fig. 9. Representation of switched system $G(z)$. The input to the system $u$ is mapped to the metric velocity $\tilde{u}$ by the block $R(z)$. In the block $I(z)$, this metric velocity is converted to its feature-based velocity according to the momentary pitch $P$. After integration the feature-based position $y$ is obtained.](image-url)
The switching results in a different gain of the above system, which is modeled by a switching $B_{t,i}$ matrix. It is chosen to have a switching $B_{t,i}$ matrix since in that case the feature position is continuous, whereas the feature velocity is not. For the switching plant $G$ mentioned above a nominal controller $K$ is designed based on the matrix $B_{t,i}$ containing the value $\bar{P}$. This controller is given in state space representation as

\begin{align}
\bar{z}_k(k+1) &= A_{K} \bar{z}_k(k) + B_{K} \varepsilon(k), \\
\varepsilon(k) &= C_{K} \bar{z}_k(k) + D_{K} \varepsilon(k),
\end{align}

(25a, 25b)

with $\bar{z}_k$ is the state of the controller. Since the time of switching is assumed to be unknown the question we want to answer is whether the (arbitrary) switching system $G$ controlled by the single controller $K$ is stable given a pitch deviation $\Delta P$. By defining the error $\varepsilon(k)$ as $\varepsilon(k) = r(k) - y(k)$ and $\bar{z}(k) = (\bar{z}_G(k)^T, \bar{z}_K(k)^T)^T$, substitution leads to the following closed-loop system

\begin{align}
\bar{z}_k(k+1) &= A_{cl,i} \bar{z}_k(k) + B_{cl,i} r(k), \\
y(k) &= C_{cl} \bar{z}_k(k),
\end{align}

(26a, 26b)

with

\begin{align}
A_{cl,i} &= \begin{pmatrix} A_{Gi,i} - B_{Gi,i} D_{K} C_{G} & B_{Gi,i} C_{K} \\ -B_{K} C_{G} & A_{K} \end{pmatrix}, \\
B_{cl,i} &= \begin{pmatrix} B_{Gi,i} D_{K} \\ B_{K} \end{pmatrix}, \\
C_{cl} &= (C_{G} 0) .
\end{align}

(27, 28, 29)

The above system is described by a linear differential inclusion (LDI) [1]. For two values of $P_1$, say $P_1$ and $P_2$, the stability of the system under arbitrary switching can be checked by solving the following linear matrix inequalities (LMIs) [1],

\begin{align}
E - A_{cl,i} T_{cl} \geq 0 \\
E - A_{cl,i} T_{cl} E \geq 0
\end{align}

(30, 31)

with the variable $E$ to be solved. However, in this case the stability is only checked for two fixed values of $P_1$. For any value in between $\bar{P} - \Delta P$ and $\bar{P} + \Delta P$ we proceed as follows. Define $A_{cl,min}$ containing $P_1 = \bar{P} - \Delta P$ as

\begin{align}
A_{cl,min} &= \begin{pmatrix} A_{G,i} - B_{G,i} D_{K} C_{G} & B_{G,i} C_{K} \\ -B_{K} C_{G} & A_{K} \end{pmatrix}.
\end{align}

(33)

Furthermore define $A_{cl,max}$ containing $P_1 = \bar{P} + \Delta P$ as

\begin{align}
A_{cl,max} &= \begin{pmatrix} A_{G,i} - B_{G,i} D_{K} C_{G} & B_{G,i} C_{K} \\ -B_{K} C_{G} & A_{K} \end{pmatrix}.
\end{align}

(34)

With these matrices we can write any matrix $A_{cl,i}$ as the convex combination of $A_{cl,min}$ and $A_{cl,max}$ by

$$A_{cl,i} = \alpha_1 A_{cl,min} + \alpha_2 A_{cl,max},$$

(35)

where $\alpha_1 > 0$, $\alpha_2 > 0$ and $\alpha_1 + \alpha_2 = 1$. Using this definition we can search for a common quadratic Lyapunov function $V(\bar{z}(k)) = \bar{z}(k)^T P \bar{z}(k)$, with $E = ET > 0$ such that $V(\bar{z}(k)) - V(\bar{z}(k+1)) > 0$, $\forall \bar{z}(k+1) = A_{cl,i} \bar{z}(k)$ or equivalently

$$\bar{z}(k)^T (E - A_{cl,i} T_{cl} \geq 0).$$

(36)

Since the inequality has to hold for every $\bar{z}$ it is sufficient to check

$$E - (\alpha_1 A_{cl,min} + \alpha_2 A_{cl,max})^T E (\alpha_1 A_{cl,min} + \alpha_2 A_{cl,max}) \geq 0,$$

(37)

where we substituted (35). This can be solved by checking

$$E - A_{cl,min}^T E A_{cl,min} \geq 0,$$

(38)

$$E - A_{cl,max}^T E A_{cl,max} \geq 0,$$

(39)

since $\alpha_1$ and $\alpha_2$ are always positive. The proof is given in Appendix A. Therefore, we can conclude that if the systems with $\bar{P} - \Delta P$ and $\bar{P} + \Delta P$ are stable, the arbitrary switching system is stable for any value $P_1$ within the bound given by $\bar{P} - \Delta P \leq P_1 \leq \bar{P} + \Delta P$.

In our case we investigate the stability for pitch deviations $\Delta P$ of up to 1 mm with respect to the nominal pitch $\bar{P}$ of 4 mm. The system under consideration is modeled with $I(z)$ defined by (21) with

$$A_I = 1, \quad B_{I,i} = \frac{T}{P}, \quad C_I = 1, \quad D_I = 0,$$

(40)

and with $R(z)$ defined in (20) with

$$A_R = \begin{pmatrix} 5.4 & -3.1 & 2.0 & -0.7 & 0.3 & -0.1 & 0 & 0 \ 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \ \end{pmatrix},$$

(41)

$$B_R = (0.1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0)^T, \quad C_R = 1 \cdot 10^{-3} \cdot (0 \ 0 \ 5 \ -8 \ 9 \ 1 \ -18 \ 15 \ -4),$$

(42)

$$D_R = 0.$$ (43)

These matrices are determined by fitting a transfer function on the measured frequency response. From the obtained model the integrator $I$ is extracted leading to the system $R$. This system is converted into a state space model leading to the matrices above. The result of the modeling is given in Fig. 10, where for $B_{I,i}$ the value $\bar{P}$ was used. The controller for this system (25) consists of a lead filter to create a phase lead with the zero of the lead filter at 7.5 Hz and the pole at 120 Hz. A second order low pass at 100 Hz with a damping of 0.6 is added such that high frequencies are not amplified. The discrete time controller is given by

$$A_K = \begin{pmatrix} 1.67 & -1.04 & 0.44 \\ 1 & 0 & 0 \ 0 & 0.50 & 0 \ \end{pmatrix}, \quad B_K = \begin{pmatrix} 64 & 0 & 0 \\ 0 \ \end{pmatrix}^T, \quad C_K = \begin{pmatrix} 26.86 & -9.09 & -31.54 \ \end{pmatrix}, \quad D_K = 0.$$ (45)

(46)

The open-loop is given in Fig. 11. The achieved closed-loop bandwidth is 20 Hz.

Given the matrices, $A_{cl,min}$ and $A_{cl,max}$ can be calculated. The LMIs can be solved efficiently using commercially
available software [6]. The LMIs were solved and a feasible solution was found, from which it can be concluded that the closed-loop arbitrary switching system is stable. Stated otherwise, the controlled system is robust against pitch deviations within 1 mm.

In practice a number of additional straightforward constraints should be satisfied, which are related to the pitch deviations in combination with the search area width. First, as mentioned before, there always have to be at least two features within the field of view. This constraint can be formulated as

$$2(\mathcal{P} + \Delta \mathcal{P}) + D < I_w.$$  (47)

The second obvious constraint that should hold is that features do not overlap

$$\mathcal{P} > 2\Delta \mathcal{P} + D.$$  (48)

For the width of the search region $S_w$, two more constraints are imposed

$$S_w > D + 2n\Delta \mathcal{P},$$  \hspace{1cm} (49)

$$S_w < \mathcal{P} - n\Delta \mathcal{P},$$  \hspace{1cm} (50)

where $n$ is the number of features that has passed the field of view within one sample. The first constraint states that there is always one complete feature within the search area, whereas the second one states that there may be only one feature within the search region. If the velocity is so high that more than one feature has passed within one sample the number $n$ gets larger than one, hence smaller pitch deviations can be allowed.

**Remark:** The above stability result could also be used to prove stability with respect to lens distortions. This makes the tedious task of camera calibration [20], [21] superfluous. It is known from literature that image-based visual servoing is more robust against lens distortions than position-based visual servoing [10]. Our analysis can also be applied for radial lens distortions. When a feature is traveling at a constant velocity radial lens distortions cause the feature to move at different velocities near the edges of the image than near the center of the image. The velocities can be higher or lower, dependent on whether a pincushion or barrel distortion is present. For that case, we can state that if a feature is near the edge of the image the gain of the system is larger or smaller than the gain near the center of the image. This gain varies during the movement of the feature. Therefore, the same type of analysis can be carried out where the value $\Delta \mathcal{P}$ depends on the distortion coefficients that describe the radial distortion.

**IX. Experiments and Results**

First, the proposed non-collocated visual servoing solution is compared to the classical collocated solution. Two experiments have been conducted for that, one with collocated feedback using the motor encoder and one with non-collocated feedback using the camera. The non-collocated controller of the previous section is used which has a bandwidth of 20 Hz, see Fig. 11. The collocated controller is tuned with a bandwidth of 40 Hz, which is twice as high as in the non-collocated case. In both experiments a repetitive structure is
placed on the stage, such that a feature is exactly centered in the field of view of the camera. The control task in both experiments is to simply keep the feature centered in the field of view. The setpoint is therefore zero. During the experiments a disturbance \( F \) is applied for one second, see Fig. 12. This disturbance might originate for example from other moving parts of the production machine. As a consequence of this disturbance and due to the finite mechanical frame stiffness the camera will move with respect to the feature. Hence, the feature will not be exactly in the center of the field of view anymore. Contrary to the collocated case, in the non-collocated case this displacement, denoted by \( x_c \) in Fig. 12, can be measured and compensated for using feedback control.

The results of the two experiments are given in Fig. 13. The measured motor encoder \( x_m \) is given on the left hand side. The camera measurement \( x_c \) is given on the right hand side. The black lines show these measurements under collocated control, whereas the gray lines show the measurements under non-collocated control. Under collocated control, the position error at the motor side is small. However, the performance at the camera side, which is our performance variable of interest, shows an error of up to 0.48 mm. There is even a steady state error of 0.1 mm. The performance at the camera side is much better under non-collocated control with maximum error values of 80 \( \mu \)m, which is a factor six better than in the collocated case. Also the steady state error is reduced significantly to less than 10 \( \mu \)m. The cumulative power spectrum (CPS) of the error signals measured by the camera, which are given in Fig. 14, also shows a significant error decrease in the case of non-collocated control.

Next, two non-collocated experiments have been carried out with a constant pitch \( P \) of 4 mm. Throughout the remainder of this section the measurement unit will be features and is denoted by \( f \). In the first experiment a reference with a constant velocity of 25 \( f/s \) is applied with the final position at feature 10 \( f \), which is well outside the field of view. In the second experiment the reference to be tracked is a sine wave with an amplitude of 4 \( f \) and a frequency of 2 Hz. The outputs \( y_v(k) \) are given in the top figures of Fig. 15 and Fig. 16. During these experiments a one step ahead prediction of the output \( y_v \) is calculated as \( \hat{y}(k+1|k) \), i.e., using the prediction given in Section III. In the lower figures of Fig. 15 and Fig. 16 this estimate is compared to the value of \( y_v(k+1) \) and this prediction error is depicted in gray. In Fig. 15 we see that during the movement there is an offset in the prediction error. Also a sinusoidal prediction error is present in Fig. 16. These errors are probably caused by the friction present in the system. This can be explained by the fact that the innovation signal of Fig. 15 is slightly negative. Hence, the prediction is ahead of the actual position, due to the unmodeled friction. This friction is not incorporated in the model of the Kalman filter, therefore leading to a prediction error. Still, in both figures it can be seen that the prediction error...
error is smaller than 0.005 f. Using the pitch $P$ of 4 mm, the prediction error can be calculated as $0.005P$ and is therefore at most $20 \mu m$. The quality of the position measurement is characterized by a $3\sigma$ of $3\sqrt{q_r}P = 2 \mu m$, with $\sigma$ the standard deviation of the measurement in meters.

To show the robustness against pitch imperfections of the closed-loop system a structure is placed on the stage with a nonconstant pitch. The pitch between subsequent features varies according to $3 \text{mm} \leq P \leq 5 \text{mm}$. The pitches are given in Table I. The experiment with the ramp reference is performed again with this structure. The pitch that was measured during the experiment is given in Fig. 17. As can be seen from this figure, the pitches coincide with the ones given in Table I. Furthermore, during the experiment the position error is measured and is given in Fig. 18. In this figure the vertical dotted lines indicate when a new feature is at the center of the field of view. Every time a new feature passes the center of the field of view the system switches and obtains a new gain as was explained in the previous section. The switching behavior can be recognized by the small transients in the error after each switching point. When larger

---

**Fig. 15.** Prediction results for ramp reference, black: $y_v$, gray: innovation signal $y_v(k + 1) - \hat{y}(k + 1|k)$.

**Fig. 16.** Prediction results for sine reference, black: $y_v$, gray: innovation signal $y_v(k + 1) - \hat{y}(k + 1|k)$.

**Fig. 17.** Non repetitive pitches.

**Table I**

<table>
<thead>
<tr>
<th>Number [-]</th>
<th>Pitch $P$ [m]</th>
<th>Measured pitch (mean) [m]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00376</td>
<td>0.00377</td>
</tr>
<tr>
<td>2</td>
<td>0.00359</td>
<td>0.00359</td>
</tr>
<tr>
<td>3</td>
<td>0.00470</td>
<td>0.00469</td>
</tr>
<tr>
<td>4</td>
<td>0.00377</td>
<td>0.00378</td>
</tr>
<tr>
<td>5</td>
<td>0.00366</td>
<td>0.00366</td>
</tr>
<tr>
<td>6</td>
<td>0.00436</td>
<td>0.00435</td>
</tr>
<tr>
<td>7</td>
<td>0.00411</td>
<td>0.00411</td>
</tr>
<tr>
<td>8</td>
<td>0.00347</td>
<td>0.00347</td>
</tr>
<tr>
<td>9</td>
<td>0.00388</td>
<td>0.00389</td>
</tr>
<tr>
<td>10</td>
<td>0.00472</td>
<td>0.00472</td>
</tr>
</tbody>
</table>
pitch imperfections are present, obviously these transients tend to increase leading to a larger position error. However from this figure, we observe that a stable closed-loop system is obtained as was expected.

X. CONCLUSIONS AND FUTURE WORK

In this paper a direct dynamic visual servoing setup has been created that controls a motion system with 1 kHz visual feedback, without the intervention of low level joint controllers. Different dynamics have been observed when using the motor encoder or measurements of the camera. By using the camera all the relevant dynamics between the camera and the features of the repetitive pattern have been measured. The control design accounts for these dynamics. Secondly, an algorithm has been developed that uses the repetitive pattern to create an optical encoder with an accuracy of 2 μm and capable of sampling at 1 kHz in combination with velocities up to 0.2 m/s. The sampling rate of 1 kHz has been realized by reading out only a part of the vision sensor to reduce the data flow. In the image processing steps a Kalman filter based prediction is used to further reduce the amount of data to be analyzed. The advantage of the proposed method is that feature-based control algorithms. Furthermore, the integration of metric movements of the tool with respect to each feature in combination with feature to feature movements based on the presented feature-based control approach will be investigated.

APPENDIX A

DERIVING THE STABILITY PROOF

In this section we prove that if the inequalities (38) and (39) are satisfied, the arbitrary switching closed-loop system is stable.

Using the Schur complement, (38) can be written as

$$\begin{pmatrix} E & A_{cl,\min}^T E \\ E A_{cl,\min} & E \end{pmatrix} \succ 0.$$  \hspace{1cm} (51)

Since $\alpha_1 > 0$ we are allowed to write

$$\begin{pmatrix} \alpha_1 E & \alpha_1 A_{cl,\min}^T E \\ \alpha_1 E A_{cl,\min} & \alpha_1 E \end{pmatrix} \succ 0.$$  \hspace{1cm} (52)

The same transformation can be applied to (39)

$$\begin{pmatrix} \alpha_2 E & \alpha_2 A_{cl,max}^T E \\ \alpha_2 E A_{cl,max} & \alpha_2 E \end{pmatrix} \succ 0.$$  \hspace{1cm} (53)

Summing up the two inequalities above results in

$$\begin{pmatrix} E & (\alpha_1 A_{cl,\min}^T + \alpha_2 A_{cl,max}^T) E \\ (\alpha_1 A_{cl,\min} + \alpha_2 A_{cl,max}) & E \end{pmatrix} \succ 0,$$  \hspace{1cm} (54)

which is the same as (37) using the Schur complement.

REFERENCES


**Jeroen de Best** was born in Wamel, the Netherlands, on June 30, 1982. He received his M.Sc. degree (cum laude) in mechanical engineering from the Eindhoven University of Technology in June 2006. Since September 2006, he is a Ph.D. student in the Control Systems Technology group at the department of Mechanical Engineering of Eindhoven University of Technology. The project he is involved in is called "Fast Focus on Structures", in which visual servoing is applied to high performance motion systems for the production of repetitive structures.

**René van de Molengraft** received the M.Sc. degree (cum laude) in mechanical engineering from TU/e in 1986, and the Ph.D. in 1990 on identification of mechanical systems for control. He is an associate professor in the Control Systems Technology group of the Department of Mechanical Engineering at TU/e. His current research interests are real-time control of embedded motion systems and the integrated design of intelligent mechatronic systems. He is an associated editor of IFAC Mechatronics.

**Maarten Steinbuch** (S’83-M’89-SM’02) received the M.Sc. degree (cum laude) and the Ph.D. degree from the Delft University of Technology, in 1984 and 1989, respectively. From 1987 to 1999 he was with Philips Labs., Eindhoven. Since 1999 he is full professor and head of the Control Systems Technology group of the Mechanical Engineering Department of Eindhoven University of Technology. His research interests are the modeling, design and control of motion systems and automotive powertrains. Since July 2006 he is also Scientific Director of the 3TU Centre of Competence High Tech Systems of the Federation of Dutch Technical Universities. He is Editor-in-Chief of IFAC Mechatronics.