Vision-Guided Feedback Control of a Mobile Robot with Compressive Measurements and Side Information

Sourabh Bhattacharya  Garrett Warnell  Rama Chellappa  Tamer Başar

Abstract—In this work, we present feedback control laws for vision guided navigation of a mobile robot. The robot is modeled as a cart that can move along a straight line, and has two vision sensors onboard. The primary vision sensor is a high resolution single-pixel camera (SPC), based on principles of compressive sensing, for capturing images. Additionally, there is a low-resolution sensor that provides coarse measurements. In this work, we consider a simple scenario in which the target is modeled as a straight line segment on a plane. The main contribution of this work is the formulation of control laws directly from the compressed measurements, obtained from the SPC. Therefore, the reconstruction of the target image is sidestepped, leading to a reduction in the amount of data acquired for control.

I. INTRODUCTION

Using computer vision for mobile robot navigation has long been an active area of research [7], [12], [18]. However, most existing methods in this field require traditional image measurements which are often associated with large amounts of data, and can lead to demanding storage, communication, and processing requirements on the system. Compressive sensing (CS) [9] promises to overcome this problem of data deluge. CS is an emerging field based on the fact that a small group of non-adaptive linear projections of a compressible signal contains enough information for reconstruction and processing. One of the areas in which compressive sensing has shown considerable promise is in the reconstruction of sparse data [8] that is encountered in many important applications, for example, imaging systems [27], [28], [31]. In this work, we explore the possibility of controlling a mobile robot using measurements from onboard imaging sensors that work on the principles of compressive sensing.

There is a plethora of work regarding CS, and its applications [5]. In computer vision, current applications of CS theory are in the development of novel imaging systems, and reconstruction algorithms for processing compressed images and videos. For example, [30] provides a framework for performing CS on video streams by modeling the evolution of the scene as a linear dynamical system. This reduces the video recovery problem to first estimating the model parameters of the LDS from compressive measurements, from which the image measurements are then reconstructed. In [14], the author introduces the idea of block compressed sensing of natural images. The proposed scheme is simpler than the existing CS techniques, and effectively captures the complicated geometric structures of natural images. In [10], authors have shown that background subtraction can be done directly on compressive measurements. This is useful in applications related to communication constrained multi-camera computer vision problems. In [26], the authors describe a technique for combining compressive measurements from multiple cameras for the tasks of tracking and three-dimensional voxel reconstruction. In robotics, there have been some recent efforts by researchers to use CS theory in tasks related to robotic exploration. In [20], [21], the authors present a framework called compressive cooperative mapping for building an obstacle map or an aerial map of an environment. The framework relies on the sparse representation of the parameter of interest in order to build a map with minimal sensing.

Recently, CS theory has attracted a lot of attention among researchers in controls and dynamical systems. In [32], the authors address the problem of observability in high-dimensional systems. They use CS theory to provide concentration-of-measure bounds for the observability matrix of a specific class of linear systems. In [11], the authors address a similar problem for a more general class of systems in a deterministic as well as stochastic setting. This significantly reduces the number of measurements required to recover the initial state of a linear discrete time system. In [6], the authors present techniques to overcome difficulties related to optimal control and tracking in linear systems driven by feedback from sensors based on CS techniques. In [29], the authors use CS techniques to identify the exact topology of an interconnected dynamical network from a limited number of measurements of the individual nodes. In [22]–[25], the authors use CS techniques in packetized predictive control of networked control systems. In [33], the authors present a non-vector space control method for compressive feedback based on which control laws are designed on sets rather than individual points in the state space. The authors validate their theory by demonstrating position control of an Atomic Force Microscope using their technique.

The reduction in measurements provided by CS techniques can certainly provide relief with respect to storage and communication requirements, but the task of processing...
them remains complex. Usually, recovering the underlying image or video from the compressive measurements requires a time-consuming, iterative solver. However, in this work, we are not concerned with image reconstruction, but rather controlling the position of a mobile robot. The compressive measurements are used to infer the state of a system. Due to the inherent uncertainty in the observation model due to random encoding strategies, a robust estimator is designed to extract the state information. Therefore, we seek to sidestep the decoding process altogether and use the compressive measurements themselves as input to our controller-estimator pair. This is a key distinction between our setup and existing techniques in visual-servoing [16] which relies on actual images for deriving the control strategy.

The organization of the paper is as follows. Section II presents the problem formulation. Section III presents the observation model. Section IV presents robust control and estimation for tracking and stabilization. Section V presents the conclusions.

II. PROBLEM STATEMENT

In this section, we first describe the problem scenario. Then we present imaging model of the vision sensors. Finally, we present the motion model for the robot. The interested reader could refer to [17] which provides an excellent, comprehensive survey of important results in compressive sensing.

Consider a mobile target placed at the origin of the reference frame. For the sake of simplicity, let us assume that the target is a planar object in the xy plane. Consider a mobile robot, modeled as a cart, equipped with a SPC. The cart lies in the half-space containing the positive z-axis, and is restricted to move on a straight line parallel to the z-axis. In addition to the SPC, the cart has onboard a low-resolution camera (side sensor) that takes coarse level images of the target. We assume a frontal pinhole imaging model for both sensors [19]. The distance between the two camera centers is negligible compared to the distance between the cart and the target. Therefore, one can assume that the two cameras are collocated, and each collects measurements of the same underlying scene from an identical point-of-view. Next, we describe the orientation of the SPC. The optical axis of the camera is assumed to coincide with the z-axis of the reference frame. Since the cart is restricted to move on a straight line parallel to the z-axis, the position of the camera center always lies on the z-axis, and the optical center always coincides with the z-axis. The image plane of the camera is always assumed to lie between the camera center and the target, i.e., the camera always faces the target. Since the low resolution sensor is assumed to be collocated with the SPC, it has the same orientation. If the assumption regarding the collocation of both the cameras is relaxed, then one needs to compute the homography matrix [15] for describing the point correspondences between the images captured by the two sensors. In order to keep the scenario simple in this paper, we impose the colocation assumption. Finally, it is assumed that the distance between the cart and the target is sufficiently large to ensure that the target always lies entirely in the field-of-view of both sensors.

Let $X_{cs} \in \mathbb{R}^{N_{cs} \times N_{cs}}$ and $X_{ss} \in \mathbb{R}^{N_{ss} \times N_{ss}}$ denote the matrices corresponding to pixel intensities in the SPC and low-resolution camera, respectively. In this work, we assume that each pixel intensity lies in the interval $[0, 1]$. Let $X_{cs}^u \in \mathbb{R}^{N_{cs} \times N_{cs}}$ denote the upsampled version of $X_{ss}$ where the upsampling factor $(N_{cs}/N_{ss})$ is $D > 1$. The upsampling process replaces each pixel in $X_{ss}$ by a block of $D \times D$ pixels. Each pixel in the $D \times D$ block has the same intensity as the original pixel in the low resolution camera. Let $X_{cs} \in \mathbb{R}^{N_{cs}^2}$ and $X_{ss} \in \mathbb{R}^{N_{ss}^2}$ denote the vectorized version of $X_{cs}$ and $X_{ss}$, respectively.

An important function of the low-resolution sensor is to perform background subtraction. Real images contain a background and a foreground component. The foreground usually represents the interesting objects in the environment under observation, for example, the target. The background contains the irrelevant aspects of the image in addition to some noise. A rough estimate of the background can be obtained from the low-resolution image of the side sensor. This estimate can be upsampled, compressed and subtracted from the SPC measurement in order to remove the background component of the signal [10].

Next, we describe the motion model of the cart. We assume a discrete time model to describe the motion of the cart. Let $z_k \in \mathbb{R}$ denote the distance of the optical center of the onboard cameras from the xy plane at time $k \geq 1$. The cart is assumed to have a kinematic model governed by the following difference equation

$$z_{k+1} = z_k + u_k + w_k \quad k \geq 1, \quad z_1 \sim N(z_0, \sigma_0^2),$$

where $u_k$ is the control applied to the cart at time $k$, and $w_k$ is i.i.d. Gaussian random variable with variance $\sigma_w^2$. Based on the measurements $\Phi x_{cs}$, where $\phi$ is the sensing matrix, and $x_{ss}$, we want to design feedback control laws for driving the cart. For a stationary target, the objective is to stabilize the cart at a desired location. For a mobile target, the objective is to track a given trajectory so that a desired distance is maintained between the cart and the target.

III. FROM COMPRESSED MEASUREMENTS TO STATE INFORMATION

In this section, we present a technique to extract the information about $z_k$ from SPC measurements. We assume a discrete-time observation model i.e., the sensors can acquire measurements (images) of the target at discrete time instants, and the information can be processed instantaneously. Let us first consider the static problem of mapping the sensor measurements to the state information at time $k$. Therefore we do not attach any time indices to the sensor measurements $x_{cs}$ and $x_{ss}^u$.

For the sake of simplicity, we imagine the target to be a straight line segment of length $l_0$ placed in a white background. The mid-point of the line segment coincides with the origin. We assume that the thickness of the line segment is very small, and is confined to a single pixel in
the high resolution image compressed by the SPC. Based on the laws of perspective projection \[19\], the length of the line segment on the image plane is given by \( l_{z_k} = \frac{l_0 f}{z_k} \), where \( f \) is the focal length of the camera, and \( z_k \) is the distance of the camera center from the origin. In a similar manner, the upsampled, blurred image of the line segment in the low resolution camera is modeled as a blurred rectangle of size \( D \times l_{z_k} \). In general, real images contain substantially higher visual information than that provided by the simple target just described. However, the mapping from the visual information to the state depends on some of the system and the task.

In this section, we present the design of a controller and an estimator for the tracking and stabilization of the cart. Combining (3) and (1) leads to the following equations for the system and the observation models:

\[
\begin{align*}
z_{k+1} &= z_k + u_k + w_k, \\
y_k &= \Delta_k z_k + v_k, \\
\Delta_k &\in \left[ \frac{1}{l_0 f(D-1)(1+\delta^2)}, \frac{1}{l_0 f(D-1)(1-\delta^2)} \right],
\end{align*}
\]

where \( w_k \) and \( v_k \) are uncorrelated i.i.d. Gaussian random variables with variances \( \sigma_w^2 \) and \( \sigma_v^2 \), respectively. The control at time \( k \) is allowed to depend on measurement at time \( k \), but it does not have access to the true value of \( \Delta_k \), other than the fact that it takes values in the given interval. Given only this partial information, our approach is to look for a design for worst possible values of \( \Delta_k \), that is to view this problem as a game where the adversary controls the value of \( \Delta_k \). In the past, a similar approach has been adopted for problems in communication systems. For example, papers \[4\] and \[3\] consider a class of minimax decision problems which arise in the transmission of a Gaussian vector message over a vector channel with partially observed statistical description. The statistically unknown part of the channel is modeled as one which is controlled by a jammer who can corrupt the transmitted message by sending noise which may be correlated with the original message under a given power constraint.

The cost function we adopt, to be minimized by the controller, is the one given below, which is associated with tracking a desired trajectory, \( z^d_k \).

\[
L(u) = \frac{1}{2} \sum_{k=1}^{T} \left[ (y_{k+1}(z_{k+1} - z^d_{k+1})^2 + r_k v_k^2) \right]
\]

Because of the complexity of the underlying game, our approach is to use a Certainty Equivalent (CE) controller, which is obtained by first deriving an optimal controller for the following deterministic system which is a noise-free version of (7).

\[
\begin{align*}
\Delta_k &\in \left[ \frac{1}{l_0 f(D-1)(1+\delta^2)}, \frac{1}{l_0 f(D-1)(1-\delta^2)} \right],
\end{align*}
\]

From \[2\], we obtain the following structure for the optimal
controller for the above system and observation models.

\[ u_k^* = \gamma^*_k(z_k) = -P_kS_{k+1}z_k - P_k[s_{k+1} - S_{k+1}d^*_k] \]  

(6)

Next, we consider the problem of designing a linear least squares estimator for \( z_k \), which minimizes the following cost function at each stage:

\[
\max_{\Delta_k \in [a,b]} E[(\mu(y_k) - z_k)^2]
\]

\[ \Rightarrow \mu^*(y_k) = \arg \min_{\mu} \max_{\Delta_k \in [a,b]} E[(\mu(y_k) - z_k)^2] \]  

(7)

where

\[ a = \frac{1}{l_0f(D-1)(1+\delta)^2}, \quad b = \frac{1}{l_0f(D-1)(1-\delta)^2} \]

Since the estimator minimizes a stage-wise cost, the following static version of the original problem can be solved to obtain the estimator:

- Consider the following equation

\[ y = \Delta z + v \]

\[ z \sim N(m,\sigma^2_z), \quad v \sim N(0,\sigma^2_v), \quad \Delta \in [a,b] \]

Find a LLSE \( \mu^*(y) \) that minimizes the following cost

\[
\max_{\Delta \in [a,b]} E[(\mu(y) - z)^2]
\]

\[ \Rightarrow \mu^*(y) = \arg \min_{\mu} \max_{\Delta \in [a,b]} E[(\mu(y) - z)^2] \]  

(8)

**Theorem 1:** The solution to (8) above has the following form

\[
\frac{\sigma_v^2}{\sigma_z^2 + m^2} \geq \frac{a(b-a)}{2} \Rightarrow \left\{ \begin{array}{l}
\mu^*(y) = \alpha'y + \beta'm \\
\Delta^* = a \quad \text{w.p. 1}
\end{array} \right.
\]

\[
\frac{\sigma_v^2}{\sigma_z^2 + m^2} < \frac{a(b-a)}{2} \Rightarrow \left\{ \begin{array}{l}
\mu^*(y) = \alpha^*_v y + \beta^*_p m \\
\Delta^* = \left\{ \begin{array}{l}
a \quad \text{w.p. p} \\
b \quad \text{w.p. (1-p)}
\end{array} \right.
\end{array} \right.
\]

where

\[ \alpha^*_p = \frac{(pa + (1-p)b)[\sigma^2_v]}{(pa^2 + (1-p)b^2)[\sigma^2_v + m^2] + \sigma^2_v - m^2[pa + (1-p)b]^2]} \]

\[ \beta^*_p = \frac{(a-b)p(1-p)[\sigma^2_v + m^2] + \sigma^2_v}{(pa^2 + (1-p)b^2)[\sigma^2_v + m^2] + \sigma^2_v - m^2[pa + (1-p)b]^2]} \]

\[ p = \frac{2\sigma_v^2}{(\sigma^2_v + m^2)(b^2 - a^2)} + \frac{b}{(b+a)} \]  

(9)

For any fixed pair \((\alpha, \beta), E(\alpha, \beta, \Delta)\) is a strictly convex function of \( \Delta \). Therefore the function attains its maximum at its boundaries. Therefore, the optimal strategy of the maximizer belongs to the following class.

\[ \Delta^* = \left\{ \begin{array}{l}
a \quad \text{w.p. p} \\
b \quad \text{w.p. (1-p)}
\end{array} \right. \]

\[ E[(\alpha y + \beta m - z)^2] = \Delta^2 \sigma^2_v + 2\Delta \alpha \sigma^2_v m(\beta - 1) - \sigma^2_v \]

**Proof:** Let \( \hat{y} = \alpha y + \beta m \) denote the LLSE of (8).

\[ \frac{\partial E(\Delta)}{\partial \alpha} = 2[\alpha(p^2 + (1-p)b^2)[\sigma^2_v + m^2] + \sigma^2_v + \sigma^2_v] \]

\[ \frac{\partial E(\Delta)}{\partial \beta} = 2m^2(\beta - 1) \]

From the first order conditions for \((\alpha^*_p, \beta^*_p, \Delta^*)\) to be a saddle point [2] for \( E \), we obtain the following

\[ \Rightarrow \alpha^*_p = \frac{(pa + (1-p)b)[\sigma^2_v + m^2] + \sigma^2_v}{(pa^2 + (1-p)b^2)[\sigma^2_v + m^2] + \sigma^2_v - m^2[pa + (1-p)b]^2]} \]

\[ \beta^*_p = \frac{(a-b)p(1-p)[\sigma^2_v + m^2] + \sigma^2_v}{(pa^2 + (1-p)b^2)[\sigma^2_v + m^2] + \sigma^2_v - m^2[pa + (1-p)b]^2]} \]

If \( \Delta^* \) is a mixed strategy, imposing the condition \( E(\alpha^*_p, \beta^*_p, \alpha) = E(\alpha^*_p, \beta^*_p, \beta) \), leads to the following

\[ p = \frac{2\sigma_v^2}{(\sigma^2_v + m^2)(b^2 - a^2)} + \frac{b}{(b+a)} \]  

(9)

From the above expression, we can conclude that \( p > 0 \), since both the terms are always positive. If \( p \) is a valid probability mass function, then we obtain the following

\[ p \leq 1 \Rightarrow \frac{\sigma_v^2}{\sigma^2_v + m^2} \leq \frac{a(b-a)}{2} \]

Next, let us consider the case when \( \Delta^* = a \). A necessary condition for this is the following:

\[ E(\alpha^*_1, \beta^*_1, a) > E(\alpha^*_1, \beta^*_1, b) \Rightarrow \frac{a(b-a)}{2} < \frac{\sigma_v^2}{\sigma^2_v + m^2} \]
Therefore, for all values of the parameter \( \frac{\sigma^2}{\sigma^2 + m^2} \), we obtain the estimator given in the theorem.

Similar structures for an optimal encoder have been encountered in problems regarding transmission of a sequence of i.i.d. Gaussian random variables through a Gaussian memory-less channel, in the presence of an intelligent jammer [1].

Finally, the instants at which \( \Delta_k^* \) switches from a mixed strategy to a pure strategy or the other way round, can be computed a priori. From (9), we obtain the following form of the CE controller that uses \( \mu^*(y_k) \) as an estimate of \( z_k 
\)

\[ u_k = A_k \mu^*(y_k) + B_k \]

where

\[ A_k = -P_k s_{k+1}, \quad B_k = -P_k [s_{k+1} - s_{k+1} z_{k+1}^d] \]

The dynamics of \( m_k \) and \( \sigma_k^2 \) are given by the following difference equations:

\[ m_{k+1} = \left[ A_k \{ \alpha_k^2 [p a + (1 - p) b] + \beta_k^2 \} + 1 \right] m_k + B_k \]

\[ \sigma_{k+1}^2 = \left[ A_k \{ \alpha_k^2 [p a + (1 - p) b] + \beta_k^2 \} + 1 \right]^2 \sigma_k^2 + \sigma_w^2 \]

\[ m_0 = z_0, \quad \sigma_0^2 = \sigma_0^2 \]

The expressions for \( m_{k+1} \) and \( \sigma_{k+1}^2 \) in terms of \( m_0 \) and \( \sigma_0^2 \) are given by the following:

\[ m_{k+1} = \Pi_{i=1}^k A_i m_0 + \sum_{j=0}^{k} \Pi_{i=j}^{k-1} [A_i B_i] + B_k \]

\[ \sigma_{k+1}^2 = \Pi_{i=1}^k (A_i)^2 \sigma_0^2 + \sum_{j=0}^{k} \Pi_{i=j}^{k-1} [A_i + 1] \sigma_w^2 \]

The instants at which \( \Delta_k^* \) makes a transition from a pure strategy to a mixed strategy or the other way round can be obtained by computing the values of \( k \) at which \( (\sigma_k^2 + m_k^2) \) crosses the value of \( \frac{a(b - a) \sigma_0^2}{2} \).

Figures 1 and 2 present simulation results. The parameters of the simulations are:

- \( q_k = r_k = 100, \quad \forall k \)
- \( T = 100 \) steps.
- \( z_0 \sim N(0, 1), \quad \sigma_0^2 = 1, \quad \sigma_0^2 = 1 \)

Figure 1 shows the simulation results for a scenario when the desired position of the cart is at a distance of 20 units from the target. Figure 2 shows the simulation results for \( z_k^d = 20 + 10 \sin(k/5) \). This simulation models a scenario in which the cart is supposed to perform sinusoidal oscillations about the point \( z = 20 \). The simulations also show the comparison of the performance of the CE controller to that of the optimal controller under perfect state measurements. In all the three simulations, it can be seen that the state of the cart, and the CE control show large amplitude oscillatory behavior compared to the state and control that are obtained from an optimal controller with perfect state measurements. Therefore, the stagewise costs associated with the CE control are significantly higher compared to the stagewise costs of the optimal controller under perfect state measurements.

V. Conclusions

In this work, we presented feedback control laws for vision-guided navigation of a cart/robot. The novel concept we presented in this paper was that a set of control laws can be computed directly from compressed measurements. Hence, we were able to avoid the reconstruction of the high-resolution images while retaining the ability to control the system. This led to a reduction in the amount of data required for control.

In the future, we plan to extend our work for realistic targets that have interesting visual features, for example, texture, color, etc. This might involve significant amount of processing of the compressed measurements in order to extract the relevant information. Moreover, if the task is to recognize a target, then even the inference problem becomes very difficult since the features of the target are not known a priori. Another interesting direction of research is to investigate multi-vehicle team scenario. For vision-based control applications that involve a network of mobile cameras [13], the latency issues that arise due to large amounts of data acquired, processed and communicated among the cameras might be resolved using compressive imaging techniques.
Fig. 2: The simulation is for the case when $z_k^d = 20 + 10\sin(k/5)$. Plot (a) shows the desired trajectory $z_k^d$, state of the cart under perfect measurements, an estimate of the compressed measurements ($\mu^p_k$) and the state of the cart under compressed measurements ($z_k^p$) as a function of time. Plot (b) shows $p$ as a function of time. Plot (c) shows the cost incurred at each stage with compressed measurements and prefect state measurements. Plot (d) shows the output of the certainty equivalent controller and the optimal controller under perfect state measurement as a function of time.

REFERENCES


