

# Vision-Based Control of Micro–Air–Vehicles: Progress and Problems In Estimation

A. Kurdila\*, M. Nechyba\*, R. Prazenica\*, W. Dahmen†, P. Binev‡, R. DeVore‡, and R. Sharpley‡

\* Department of Mechanical and Aerospace Engineering  
Department of Electrical and Computer Engineering  
University of Florida, Gainesville, FL 32603

† Rheinisch-Westfälische Technische Hochschule Aachen, Institut für Geometrie und Praktische Mathematik  
Templergraben 55, D-52056, Aachen, Germany

‡ Department of Mathematics  
University of South Carolina, Columbia, SC 29208

**Abstract**— Vision-based control of agile autonomous vehicles in complicated 3-D environments requires fundamental and ground breaking innovations in multiple, related disciplines. These disciplines include control theory, vision processing, signal processing, sensor development, micro-computer technology and the design and instrumentation of micro-air-vehicles (MAVs). Extremely agile small vehicles with acute situational awareness are required for flying through complex environments such as urban canyons confined by buildings, trees, ... etc. Vision-based navigation of such vehicles in the neighborhood of ground vehicles, civilians, as well as in poor weather likewise requires a host of innovations in robust vision estimation. Robust vision estimation includes tasks such as feature point extraction, feature point tracking, image registration, segmentation, object detection and object identification. Control of these agile autonomous vehicles requires innovative control methodologies that synthesize image plane inputs in real-time. This problem can be construed as a feedback control problem with millions of raw input channels, the pixels in the image plane. Observation aggregation techniques, rigorously stable “visual servoing” methods, robust sensor fusion methods and fundamental theoretical studies of the controllability and observability of such flight systems are just some of the control theoretic issues that are currently lacking. This paper outlines progress and several open problems in vision-based estimation for MAVs. It introduces and summarizes several critical technologies that are currently being applied to this overall objective.

## I. INTRODUCTION

The vital and increasing role of unmanned autonomous vehicles (UAVs) throughout the military is now well-appreciated. It has been increasingly apparent that small or “micro” UAVs are emerging in numerous applications of interest to the military. Fixed wing micro-air-vehicles, or MAVs, have been defined as those vehicles that have a wing span less than six inches in length. Perhaps most significant are those applications relating to observation, measurement or surveillance. Brief reflection suggests a host of applications of interest to the Department of Defense, Department of Homeland Security and civilian law enforcement.

### A. A Simple, Prototypical Vision-Based Control

A simple example easily illustrates the deep, diverse and numerous technical issues that must be resolved to enable autonomous flight of MAVs in urban environments.

Consider the MAV depicted in Fig. 1. This vehicle has a



Fig. 1. (a) six-inch UF MAV; (b) six-inch MAV in flight.

wing span of approximately six inches and a single forward looking camera. It is but one of a collection of carbon fiber composite, membrane wing, small UAVs that have been used as flight testbeds at the University of Florida. Topics such as inner loop control, real-time vision processing, aeroelastic response, aerodynamic characterization and autopilot design have been studied for this class of vehicle.

Consider what is perhaps one of the most simple, direct applications of vision-based control for this class of MAV: horizon-detection for feedback roll stabilization. Although this problem is remarkably simple in nature, it employs elementary versions of fundamental vision processing tasks that must be employed in general, 3-D flight vehicle control, to be discussed shortly. It includes rudimentary image segmentation, real-time optimization, elementary multilevel filtering and feedback stabilization. Fig. 2 depicts a single frame collected from the onboard camera in a MAV flight. As illustrated schematically in the figure, simple image

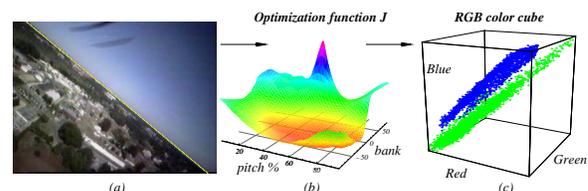


Fig. 2. (a) original image; (b) optimization criterion  $J$  as a function of bank angle and pitch percentage, that is, vertices along the perimeter of the image; (c) resulting classification of sky and ground pixels in RGB space.

segmentation is achieved via a real-time optimization that finds the best separating hyperplane that distinguishes the ground from the sky. This simple segmentation provides a one channel measurement for feedback to achieve roll stabilization. As summarized in [1], this inner loop control scheme provides an enormous practical advantage for guidance and navigation of remotely piloted MAVs.

Still, there are numerous ways in which such a rudimentary control structure can fail. Consider, for example, Fig. 3 which depicts the performance of the segmentation algorithm discussed above when the horizon is not a clean, isolated image feature. In this frame, at least three distinct image regions can be associated with the nearby field, distant mountains and clouds overhead. This simple example

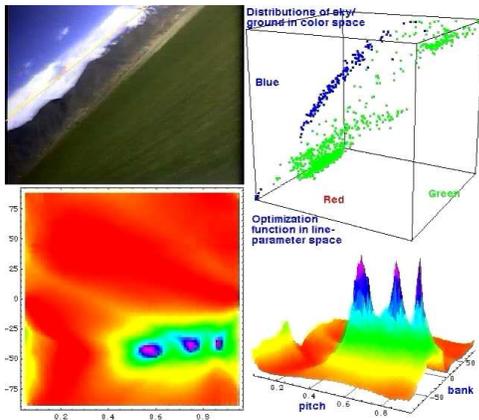


Fig. 3. Multiple peak, ambiguous horizon example.

illustrates the difficulty in processing complex imagery with multiple visual features, indistinct features or “cluttered environments.”

In addition, the real-time operation of the segmentation procedure relies on an inherent tradeoff in all such vision-based control applications. Real-time throughput constraints dictate that certain detailed features of the image are not employed. Image decimation, or more generally, image multiscale filtering, is required to match processor throughput to image granularity. In this case, image decimation or multilevel filtering has a direct consequence in terms of performance. The precision of the horizon angle estimate is quantized in direct relation to the granularity of the image.

Fig. 4 illustrates a distinctly different impediment to the functionality of the simple, prototypical control algorithm discussed above. Video noise and image artifacts can play havoc with the feedback control algorithm. Because analog transmission is employed for this particular hardware implementation, the real-time processing burden is passed to a computational ground station. Interference and corrupted transmission frames are common.

If the only application of vision-based control for MAVs is the simple, prototypical example just described, any number of *ad hoc* patches might address the summarized deficiencies. If the underlying assumption is that the scene

or environment is uncluttered, or that significant *a priori* knowledge of the operating environment is available, any number of existing vision and control algorithms can be expected to be viable.

The *Special Session on Vision-Based Guidance, Navigation and Control of Small and Micro Unmanned Autonomous Vehicles* must address numerous complex estimation and control problems that are based on vision-based measurement systems that operate in complex, “cluttered” environments. A number of challenges and open questions will need to be addressed in the special session. Some of these challenges are discussed in the following sections.

### B. Bandwidth Challenges

One of the foremost constraints in realizing the above research goal is bandwidth. In this context, bandwidth refers to the time to conduct hardware processing, sensor processing, signal processing and also computational throughput. Bandwidth constraints emerge in this problem in numerous fashions. In vision processing for flight control, many tasks must be undertaken including denoising, feature recognition, segmentation, registration, correlation, etc. The incorporation of complex vision processing tasks as part of a closed loop feedback control will, in many instances, dictate the complexity of the vision processing analysis. Scalable algorithms that trade speed for approximation error are desirable. Understanding the essential tradeoff between the complexity of the vision analysis task, the required bandwidth of the feedback controller, the stability of the closed loop and the performance of the closed loop system is a primary goal of this session.

### C. Robust and Real-Time Vision Estimation Challenges

It is now well-accepted that many classical vision-processing tasks such as structure from motion (SFM), denoising, segmentation, feature detection and identification can be computationally solved, at least in an academic sense. In some cases, the solution procedures require ideal image sequences, sufficient computational time, and perhaps, ad hoc or heuristic user input. However, flight testing with aircraft-borne CCD cameras makes it clear that images gathered under flight conditions are hardly ideal. Video dropout, video transmission noise, video artifacts, and rapidly varying scene sequences associated with forward flight velocity render many classical image processing tasks

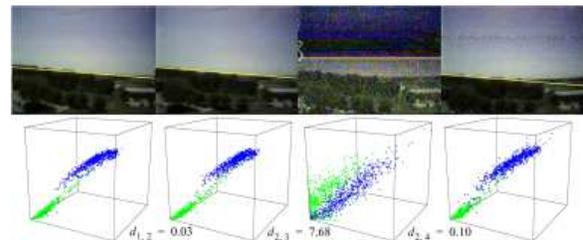


Fig. 4. Example of a corrupted frame.

inadequate in this context. Thus, at a primitive vision-processing level, novel methodologies for real-time denoising, segmentation, image registration, feature extraction, optic flow estimation and geometry estimation must be developed. These techniques must be real-time executable and robust with respect to noise and artifacts.

## II. ROBUST 3D VEHICLE STATE ESTIMATION IN COMPLEX ENVIRONMENTS

The robotics community has studied the problem of estimating camera kinematics from camera measurements for decades. Collectively, many of these pertinent algorithms are known as “Structure From Motion,” or SFM, techniques. Numerous excellent overview papers exist in the robotics literature that study advances in SFM techniques. Unfortunately, many of these algorithms are not well-suited for the task of real-time estimation from cameras fixed to flight vehicles that are rapidly navigating cluttered environments. The algorithms may be designed for the movie industry, in which the scene analysis can be carried out with an “unlimited” computational budget. Feature points, or persistent point landmarks between frames, are required for many SFM algorithms and must be identified during computation. The feature points may be selected based on *a priori* user input information which is not possible in our context. The issues associated with automatic registration, or tracking of feature points between frames, as well as numerous other properties of current generation SFM algorithms, are discussed in [2], also presented in this Special Session.



Fig. 5. Successive frames in vision sequence with automated feature points.

Fig. 5 depicts two frames from a sequence of flight video collected by the MAV in Fig. 1. These frames depict *automatically correlated* feature points obtained using analysis and tools provided by the authors of [2]. It must be recalled that, for our application, they must be generated in real-time for flight-relevant SFM algorithms. The problem of robust and real-time feature point tracking is also discussed in [2].

One class of SFM algorithms that may be amenable to general vision-based control in complex 3D environments

is based on Kalman filtering. This approach relies on correlated feature points as input, such as those depicted in Fig. 5. Significant progress has been made by authors that have derived and studied recursive Kalman filtering-based SFM in [3], [4], [5]. While these studies are exceptional in that they constitute foundational work in the field, they treat “academic problems” that do not include the complex imagery as viewed in Fig. 5. For the most part, the model, or propagator, in the extended Kalman filter in these papers has been a simple random walk. For the treatment of state estimation for flight vehicles, the formulation is considerably more complex. Extensions of the work in [3], [4], [5] can be found, for example, in [6]. The formulation therein employs the full, nonlinear 3D aircraft state equations as the state propagator, and the subsequent corrections are based on the so-called “subspace constraint” discussed in [3], [4], [5]. The assumption that detailed aerodynamic models are available for MAVs does not produce a robust algorithm: MAV aerodynamics are notoriously difficult to characterize. In recent work by some of the authors, aircraft equations with simplified aerodynamics are used as the state propagator in [7] and subsequent corrections are derived from the epipolar constraint.

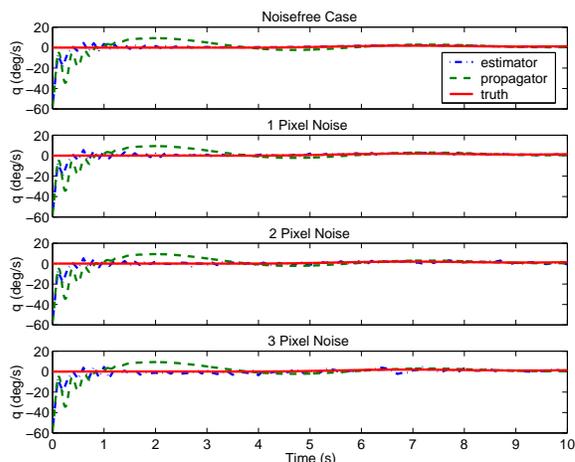


Fig. 6. Performance of the vision-based state estimator for varying levels of measurement noise in the pixel plane.

Fig. 6 depicts the performance of the methodology presented in [7]. In this example, the full, nonlinear, 3D aircraft equations are used to generate the measurements of simulated feature points in the pixel plane of a CCD camera. The propagator in the implicit extended Kalman filter does not incorporate the full nonlinear aerodynamic terms that appear in the equations of motion. Instead, simplified nonlinear aerodynamic terms are incorporated into the propagation model. This simplified model seeks to avoid the use of detailed aerodynamics, and consequently, provide for an estimation algorithm that is robust with respect to modeling error. In [7], the performance of this vision-based Kalman filter is evaluated for a simulated roll doublet maneuver. Eight aircraft states are estimated for

a case where the initial conditions provided to the filter are in significant error. Furthermore, the robustness of the algorithm with respect to measurement noise is explored by adding varying amounts of Gaussian white noise to the simulated feature point positions in the pixel plane. Fig. 6 depicts, as an example, the estimation of one component of the body angular velocity for noise levels varying from 0 to 3 pixels standard deviation. For this state, the true initial condition is 0 deg/s while the estimator has been given an erroneous initial condition of 57.3 deg/s. The figure shows the estimates provided by the Kalman filter as well as those obtained using the propagator with no measurement updates. In this case, while the propagator eventually converges to the true angular velocity, the vision-based estimator does so significantly faster. As the noise level increases, the estimator still converges to the vicinity of the true state but exhibits increasing oscillations. Therefore, it is clear that sensitivity to noise is a critical concern in vision-based, implicit extended Kalman filtering approaches to SFM. This problem is discussed in more detail in [7].

### III. MULTISCALE ANALYSIS FOR ROBUST AND SCALABLE ALGORITHMS

As noted in the introduction, it is believed that multilevel filtering can play an essential role in vision-based GNC algorithms that seek to respect the tradeoff between computational burden and image detail. This section will additionally illustrate that multilevel methods often exhibit an “inherent” robustness in vision estimation. This is due to the fact that multilevel filtering often induces implicit denoising in transforming high fidelity images to a lower granularity. In this section and the next, we discuss two such approaches. This section presents recent progress in Multiscale Linear Discriminant Analysis (MLDA) for robust, multiscale vision estimation tasks. Details of this approach can be found in [8] and [9].

An MLDA atom  $w$  is a piecewise constant function on either side of a linear discriminant that intersects a square in vertices  $v_i$  and  $v_j$ , as illustrated in Fig. 7. The discriminant  $(v_i, v_j)$  divides the square into two regions characterized by  $\mu_0$  and  $\mu_1$ , which represent the mean vectors of RGB pixel color values for the two regions. Decomposing the image into a number of dyadic squares and finding their corresponding MLDA atoms, we obtain the MLDA dictionary over a finite range of locations, orientations and scales.

MLDA image-feature extraction is performed by searching through the MLDA dictionary for the atoms that best represent the analyzed image with respect to two competing criteria: *discrimination* and *parsimony*. As the size of squares decreases, we achieve better piecewise linear approximation of boundaries between regions in the image, as illustrated in Fig. 7. Thus, the analyzed image is decomposed into dyadic squares organized in the MLDA tree  $\mathcal{T}$ . The MLDA tree  $\mathcal{T}$  is *incomplete*, because atom generation stops at different scales for different locations

in the image. The leaf nodes of  $\mathcal{T}$  store the final MLDA image representation.

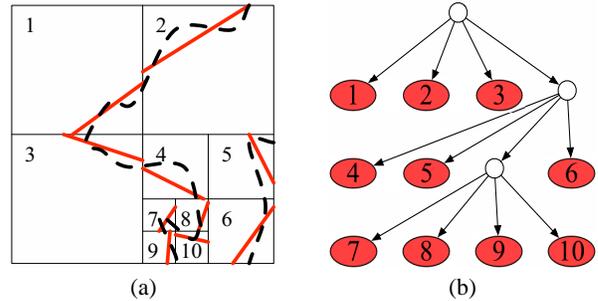


Fig. 7. (a) MLDA graph: the dashed line depicts the actual curve; (b) corresponding MLDA tree: ellipsoid nodes represent the leaf MLDA atoms.

To control the generation of children dyadic squares, we impose the second optimization criterion (i.e. parsimony), as a counter-balance to accuracy. Essentially, a penalty term is added to the optimization criteria used to achieve discrimination. The penalty term may consist of the number of terminal nodes multiplied by a scale factor, for example. Clearly, an exhaustive search in tree space for the minimum cost function is computationally prohibitive. Therefore, we implement a one-step optimization procedure, as in [10].

Instead of stopping at different terminal nodes, we continue the MLDA decomposition until all leaf squares are small in size, resulting in a large tree. Then, we selectively prune this large tree upward. We can regulate the pruning process by increasing the penalty to obtain a finite sequence of subtrees with progressively fewer leaf nodes. This process is repeated until the actual number of leaf nodes is equal to or less than the desired number of leaf nodes. We note that the pruning process described above is computationally fast and requires only a small fraction of the total tree construction time. Starting with a complete tree, the algorithm initially trims off large subtrees with many leaf nodes. As the tree becomes smaller, the procedure tends to cut off fewer nodes at a time.

MLDA implicitly encodes information about spacial frequencies in the image (i.e. texture) through the process of tree pruning. Furthermore, the MLDA tree can easily be examined for spacial interrelationships of its linear discriminants, such as connectedness, collinearity and other properties of curves in images. Therefore, the MLDA representation is appropriate for computer-vision tasks where both color and texture are critical features.

In benign flight video, the horizon-detection problem can be adequately solved by computing only the root atom of the MLDA tree, where the linear discriminant of the root is the optimal solution for the horizon estimate, as illustrated in Fig. 2; in fact, the work presented in [1], [11] is equivalent to this special case.

There are, however, unfavorable image conditions (e.g., when the horizon is not a straight line and/or an image is corrupted by video noise), when it is necessary to examine

discriminants at finer MLDA scales. For example, in Fig. 8, we show that the discriminant of the root MLDA atom does not coincide with the true horizon due to video noise.

Expanding the MLDA tree corresponds to image filtering and leads to more accurate positions for the linear discriminants, which then give more accurate evidence of the true horizon’s location in the image. Also, in Fig. 9, we show an example where MLDA detects the true horizon more correctly as compared to a single-line discriminant.

This general framework has many extensions. It can be used to construct multilevel and scalable vision-based estimation sub-tasks including feature detection and feature tracking. As an example, Fig. 10 illustrates the use of MLDA for object recognition in successfully identifying the car in the image. A detailed discussion can be found in [8] and [9].

#### IV. LEARNING AND GENERALIZATION FOR OBSTACLE ESTIMATION

In the last section, it was shown that multilevel techniques lend themselves to scalable and robust algorithms for segmentation, feature identification, ... etc. In this section, we will outline a class of multilevel techniques applicable to geometric estimation. This approach is based on recent developments that synthesize progress in learning theory [12] and multilevel approximation [13]–[15].

To motivate the presentation, consider the task of rapid path planning. While excellent algorithms for path planning have been developed over the years, the constraint that the method be implementable in real-time is severely restrictive. The authors of [16] have derived a rapid planning algorithm that solves sequences of small, local optimization problems to traverse a 3D urban environment. While there are interesting and nontrivial open problems associated with the stability and convergence of the technique, the formulation is quite amenable to rapid execution. Also, the technique assumes that estimates of the urban environment in the form of “blocky geometric primitives” are available, as depicted in Fig. 11. Thus, the context of the estimation problem is clear. From an incoming, real-time collection of scattered data (such as the global coordinates of feature points obtained from an SFM algorithm), we seek real-time estimates of “blocky geometry.” We wish our algorithm to

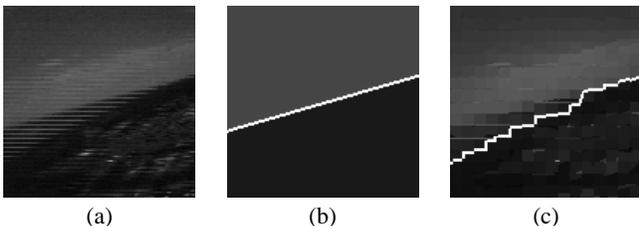


Fig. 8. MLDA efficiently filters video noise: (a) noise degraded original image; (b) MLDA root atom with the discriminant not equal to the true horizon; (c) MLDA atoms at finer scales as clues for the true horizon position.

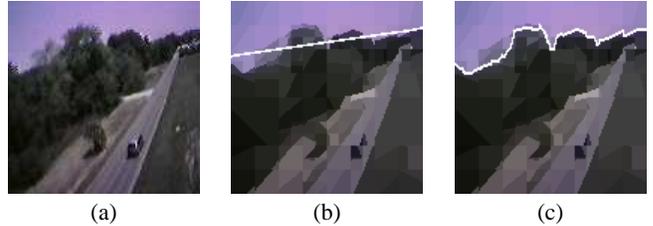


Fig. 9. Non-straight-line horizon: (a) original image; (b) the root MLDA atom with the discriminant not equal to the true horizon; (c) MLDA atoms at finer scales as clues for the true horizon position.

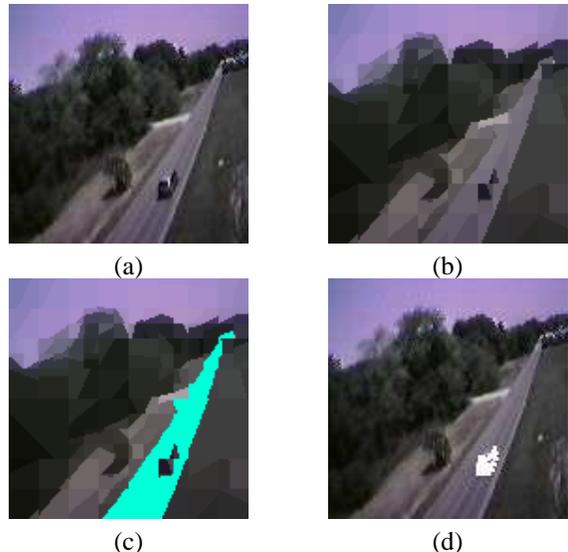


Fig. 10. The hierarchy of visual contexts conditions gradual image interpretation: (a) a  $64 \times 64$  MAV-flight image; (b) MLDA image representation; (c) localization: recognition of global objects; (d) car recognition using CWT and HSI features.

be scalable and robust. The analysis that follows describes an adaptive, multilevel wavelet method that is designed to have these properties. Details may be found in [14].

For Banach spaces  $X, Y$  let  $\rho$  denote a probability measure on the Cartesian product  $Z := X \times Y$ , write  $d\rho = d\rho(y|x)d\rho_X(x)$  and define the regression function

$$f_\rho(x) := \int_Y y d\rho(y|x) \quad (1)$$

that gives the conditional expectation of the random variable  $y$  at  $x$ . Here and in what follows it will always be assumed that for some constant  $M$  one has  $\|y\|_Y \leq M$ . It is easy to check that  $f_\rho$  is the minimizer of the functional

$$\mathcal{E}(f) := \int_Z (y - f(x))^2 d\rho \quad (2)$$

over the weighted space  $L_2(\rho_X, Y)$  which consists of all functions from  $X$  to  $Y$  which are square integrable with respect to  $\rho_X$ . The objective is to find an estimator  $f_{\mathcal{Z}}$  for  $f_\rho$  based on the sample vector  $\mathcal{Z} = (z_1, \dots, z_m) \in Z^m$  where each  $z_i = (x_i, y_i), i = 1, \dots, m$ . A common approach is to obtain  $f_{\mathcal{Z}}$ , in analogy to (2), as

$$f_{\mathcal{Z}} := \operatorname{argmin}_{f \in \mathcal{H}} \mathcal{E}_{\mathcal{Z}}(f) \quad (3)$$

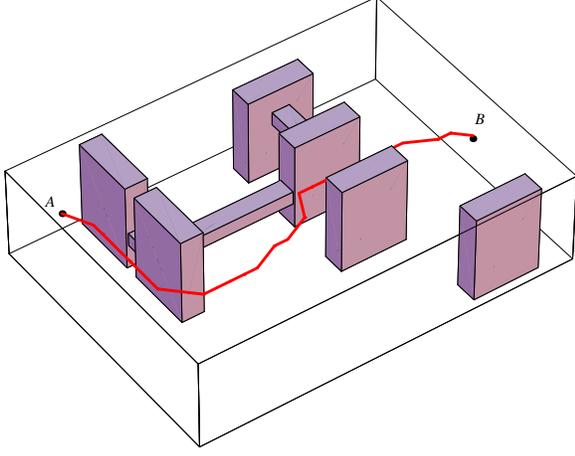


Fig. 11. Example path through urban environment, local optimization of [16].

where

$$\mathcal{E}_{\mathcal{X}}(f) := \frac{1}{m} \sum_{i=1}^m (y_i - f(x_i))^2 \quad (4)$$

The *hypothesis class*  $\mathcal{H}$  typically depends on a finite number  $N$  of parameters that are unknown. The key issue is then to obtain error estimates for the probability

$$\text{Prob}_{\mathcal{X}} \{ \|f_{\rho} - f_{\mathcal{X}}\|_{L_2(\rho_X, Y)} \geq \eta \} \quad (5)$$

or the expectation

$$E \left( \|f_{\rho} - f_{\mathcal{X}}\|_{L_2(\rho_X, Y)}^2 \right) \quad (6)$$

In this paper we propose a class of constructive schemes that will be applicable to the case when  $Y = \mathbb{R}$  and  $X$  is a compact subset in  $\mathbb{R}^d$ . In fact, we will organize our presentation by choosing  $d = 1$  and  $X = [0, 1]$ . This choice enables a succinct presentation of the key results, and at the same time, makes it clear that the results are easily extended to the case when  $d > 1$ .

Consider a partition  $\Lambda$  of  $[0, 1]$  into  $N = |\Lambda|$  intervals and denote by  $S_{\Lambda}$  the space of all piecewise constant functions with respect to  $\Lambda$ . The elements in  $\Lambda$  are allowed to have arbitrary size. We adopt the above assumptions on the probability space  $Z = [0, 1] \times Y$  and the corresponding distribution  $\rho$ , i.e.,  $\|y\|_{L_{\infty}} \leq M$ . Since there is no more risk of confusion we shall briefly write in a slight abuse of notation  $L_2(\rho)$  instead of  $L_2(\rho_X, Y)$  and  $\rho_I$  instead of  $\rho_X(I)$  for any  $I \subset [0, 1] = X$ . If the distribution  $\rho$  is known, it would be possible to compute

$$P_{\Lambda} f_{\rho} := \text{argmin}_{f \in S_{\Lambda}} \mathcal{E}(f) \quad (7)$$

which is the  $L_2(\rho)$ -projection into  $S_{\Lambda}$ . Since we are focused on applications where  $\rho$  is *not known*, we are limited to the use of empirical data. Define a minimization over the empirical data

$$f_{\mathcal{X}, \Lambda} := \text{argmin}_{f \in S_{\Lambda}} \mathcal{E}_{\mathcal{X}}(f) \quad (8)$$

$$= \text{argmin}_{f \in S_{\Lambda}} \frac{1}{m} \sum_{i=1}^m (y_i - f(x_i))^2 \quad (9)$$

i.e.,

$$f_{\mathcal{X}, \Lambda} = \sum_{I \in \Lambda} c_I(\mathcal{X}) \chi_I \quad (10)$$

where

$$c_I(\mathcal{X}) = \text{argmin}_{c \in \mathbb{R}} \sum_{i \in I} (y_i - c)^2 \quad (11)$$

The geometric estimation algorithm presented in this system relies on the development of wavelet bases that are defined on a partition that is adapted to the empirical data. Let  $\mathcal{D}^o$  denote the set of all dyadic subintervals of  $[0, 1]$ , including the interval  $[0, 1]$  itself. The set  $\mathcal{D}$  can be viewed as a tree with root  $[0, 1]$ . Each  $I \in \mathcal{D}$  has two children  $I_-, I_+$ , the left and right dyadic subintervals of  $I$ . Note that together with the characteristic function  $\chi_{[0, 1]}$ , the functions  $\psi_I := \frac{1}{\rho(I_-)} \chi_{I_-} - \frac{1}{\rho(I_+)} \chi_{I_+}$  for  $I \in \mathcal{D}$  form an orthogonal system in  $L_2(\rho)$ . For convenience we extend  $\mathcal{D}^o$  to  $\mathcal{D}$  by an “index”  $I^*$  and write  $\psi_{I^*} := \chi_{[0, 1]}$ . The whole orthogonal basis can be written as  $\Psi = \{\psi_I : I \in \mathcal{D}\}$ . Normalizing these functions in  $L_2(\rho)$ , i.e. setting

$$\psi_{I, 2} := \psi_I / \|\psi_I\|_{L_2(\rho)} \quad (12)$$

one can check that the corresponding *wavelet coefficients* are given by

$$\begin{aligned} d_I &= d_I(f_{\rho}) = \int_0^1 f_{\rho}(x) \psi_{I, 2}(x) d\rho_X \\ &= \sqrt{\frac{\rho(I_-)\rho(I_+)}{\rho_I}} (c_{I_+} - c_{I_-}) \end{aligned} \quad (13)$$

for each  $I \in \mathcal{D}^o$ . We define  $d_{I^*} := c_{[0, 1]}$ . It is easy to show that

$$\sum_{I \in \mathcal{D}} d_I^2 = \|f_{\rho}\|_{L_2(\rho)}^2 \quad (14)$$

The goal of defining such a wavelet basis adapted to a non-uniform partition is to obtain a good approximation to  $f_{\rho}$  from a few terms in the infinite wavelet expansion of  $f_{\rho}$ . It is well-known that such approximations can be obtained by thresholding the wavelet coefficients. For a positive threshold  $\tau$  define the set of intervals

$$\mathcal{I}_{\tau} := \{I \in \mathcal{D} : |d_I| \geq \tau\} \quad (15)$$

We want to associate a partition of  $[0, 1]$  to this set of “large wavelet coefficients,” which may be quite small depending on the value of  $\tau$ . Clearly, the elements of  $\mathcal{I}_{\tau}$  may not cover  $[0, 1]$ . Let  $\mathcal{T}_{\tau} \subset \mathcal{D}$  denote the smallest tree containing  $\mathcal{I}_{\tau}$ . Moreover, let us define for any tree  $\mathcal{T} \subset \mathcal{D}$  that has as its root the interval  $[0, 1]$  the set of *outer leaves* by

$$\partial \mathcal{T} := \{I \in \mathcal{D} : I \notin \mathcal{T} \text{ but the parent of } I \text{ belongs to } \mathcal{T}\} \quad (16)$$

The set of outer leaves defines a partition of  $[0, 1]$ . It can be shown that a good approximation to  $f_{\rho}$  would then be given by  $P_{\partial \mathcal{T}_{\tau}} f_{\rho}$  for a suitable selection of  $\tau$ .

With these definitions in mind, the crucial choice that must be made is the selection of the set of functions  $\mathcal{H}$

over which the approximations will be made. Fix any small number  $\gamma > 0$ , say  $\gamma < 1/2$ , and define for  $0 < s \leq 1$  the positive number  $\mu$  via the equation

$$\frac{1}{\mu} = s + \frac{1}{2} - \gamma \quad (17)$$

Let  $d_I(f)$  denote the wavelet coefficients with respect to the  $\rho_X$ -normalized Haar basis of a given function  $f \in L_2(\rho)$ . Now let

$$B^s := \left\{ f \in L_2(\rho) : \sum_I |I|^{-\gamma\mu} |d_I(f)|^\mu := |f|_{B^s} \leq \infty \right\} \quad (18)$$

and define

$$\|f\|_{B^s}^\mu := \|f\|_{L_\mu}^\mu + |f|_{B^s}^\mu \quad (19)$$

This set of functions will play the role of the *prior class* in the theoretical development that follows. For the development of a constructive approximation, it is useful to note that the wavelet coefficients  $d_I(f)$  have for  $f \in B^s$  a scalewise decay

$$|I|^{-\gamma} |d_I(f)| \leq C \quad (20)$$

whenever  $s > \gamma$ . This property can be viewed as a statement that such a function  $f$  so selected exhibits some minimal smoothness in  $L_2(\rho_X)$ .

We now concentrate on developing in more detail what might be called the *empirical wavelet coefficients*. We cannot compute the partitions discussed above because we do not know the probability measure  $\rho$ . Hence, we likewise do not know the basis  $\Psi$  and cannot compute the wavelet coefficients  $d_I$ . However, we can expect to guess good partitions from the observed data  $\mathcal{X}$ . We cannot expect useful information from such partitions when the threshold becomes too small relative to the sample size. We shall essentially restrict the discussion to

$$\tau \geq \tau_m := \log m / \sqrt{m} \quad (21)$$

To describe a computational scheme that is realizable, let us fix a small parameter  $\gamma > 0$  and choose a limiting threshold  $\tau_m = \log m / \sqrt{m}$ . From (20), we can expect to find significant wavelet coefficients only for intervals of the size  $|I| \geq C\tau_m^{1/\gamma}$ . Define the empirical measure of the interval  $I$  as

$$\rho_I(\mathcal{X}) := \frac{1}{m} \sum_{i=1}^m \chi_I(x_i) \quad (22)$$

The empirical wavelet coefficients are then simply given by

$$d_I(\mathcal{X}) := \begin{cases} \sqrt{\frac{\rho_{I_-}(\mathcal{X})\rho_{I_+}(\mathcal{X})}{\rho_I(\mathcal{X})}} (c_{I_+}(\mathcal{X}) - c_{I_-}(\mathcal{X})) & \text{if (C1)} \\ 0 & \text{otherwise} \end{cases}$$

The condition (C1) is defined to be true whenever  $\rho_I(\mathcal{X}) \neq 0$  and  $|I| \geq C\tau_m^{1/\gamma}$ . Of course, they are obtained just as the  $d_I$  by forming an orthonormal basis with respect to the discrete measure.

In analogy with the definition of partitions discussed above in terms of the measure  $\rho$ , a similar procedure can be carried out for the empirical measure. Define for any threshold  $\tau > 0$  the set of  $\tau$  significant empirical coefficients

$$\mathcal{I}_\tau(\mathcal{X}) := \{I \in \mathcal{D} : |d_I(\mathcal{X})| \geq \tau\} \quad (23)$$

and let  $\mathcal{T}_\tau(\mathcal{X}) \subset \mathcal{D}$  denote the smallest tree containing  $\mathcal{I}_\tau(\mathcal{X})$ . For  $\mathcal{T}_\tau(\mathcal{X})$  defined above and any  $\tau \geq \tau_m$  let

$$\Lambda_\tau(\mathcal{X}) := \partial \mathcal{T}_\tau(\mathcal{X}) \quad (24)$$

and define

$$f_{\mathcal{X},\tau} := \operatorname{argmin}_{f \in S_{\Lambda_\tau(\mathcal{X})}} \mathcal{E}_{\mathcal{X}}(f) = \sum_{I \in \Lambda_\tau(\mathcal{X})} c_I(\mathcal{X}) \chi_I \quad (25)$$

We can now state the main result of this section.

*Theorem 1:* There exists a constant  $c$  depending on  $\gamma > 0$  (but not on  $m, \rho, \tau \geq \tau_m$ ) such that whenever  $f_\rho \in B^s$  one has for

$$r := \frac{1}{\gamma} > \frac{1}{2s} + 1 \quad (26)$$

that

$$\operatorname{Prob}_{\mathcal{X}} \left\{ \|f_\rho - f_{\mathcal{X},\tau}\|_{L_2(\rho)} \geq \tau^{\frac{s}{2s+1}} \right\} \leq \tau^{-r} e^{-cmr^2} \quad (27)$$

In particular, defining

$$f_{\mathcal{X}} := f_{\mathcal{X},\tau_m}, \quad \tau_m := \log m / \sqrt{m} \quad (28)$$

one has

$$\operatorname{Prob}_{\mathcal{X}} \left\{ \|f_\rho - f_{\mathcal{X}}\|_{L_2(\rho)} \geq \left( \frac{(\log m)^2}{m} \right)^{\frac{s}{2s+1}} \right\} \leq Cm^{r/2} e^{-c(\log m)^2}$$

and hence

$$E(\|f_\rho - f_{\mathcal{X}}\|_{L_2(\rho)}) \leq \left( \frac{(\log m)^2}{m} \right)^{\frac{s}{2s+1}} \quad (29)$$

In reviewing the preceding analysis, it is important to realize several features of this formulation.

- In contrast to much of the foundational work in learning theory, including for example [12], the approach above is constructive.
- As discussed more fully in [14] and [15], the methodology above attains an *optimal* rate of convergence that is again, an improvement over that attained in [12].
- The algorithm described above exhibits a *universality* in that it is optimal for a range of smoothness spaces (or hypothesis spaces).
- The methodology summarized in Theorem 1 can be implemented in a recursive, or block recursive, manner. As such, it has excellent potential to realize the vision-estimation problem required for the rapid path re-planning as illustrated in Fig. 11.

A piecewise linear version of the wavelet-based learning algorithm has been implemented, but theoretical convergence results are still under development. The performance of the piecewise constant and piecewise linear learning

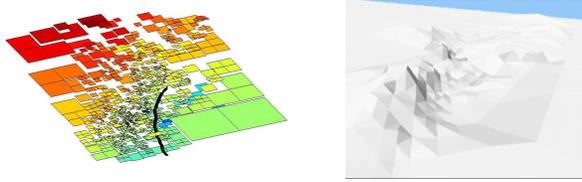


Fig. 12. Piecewise constant and piecewise linear terrain mappings from images taken during the MAV flight shown in Fig. 5.

algorithms are depicted in Fig. 12. Using the methodology in [2], a set of 3D feature points was obtained from images during a MAV flight, as shown in Fig. 5. The piecewise constant approximation of the terrain is shown on the left (with the flight path shown in black) and the piecewise linear reconstruction is shown on the right. Fig. 13 and 14 show results from a simulated MAV flight over mountainous terrain obtained from satellite imagery. Fig. 13 shows a single image frame with identified context-dependent feature points. The red points corresponding to the clouds are detected as outliers and removed. The estimated 3D positions of the feature points are subject to error in the SFM algorithm, and the deviations of the points from the true surface are depicted in the lower left plot. Finally, the figure shows the piecewise linear reconstruction of the terrain. Fig. 14 shows an overhead view of the simulated MAV flight (shown in yellow) with the full spread of identified feature points. Also shown is an overhead view of the piecewise linear approximation of the terrain. Note that, in 2-D, the piecewise linear reconstruction is defined over triangular elements, and the high density of triangles in certain regions denotes locations where feature points are plentiful.

## V. CONCLUSION

This paper has given an overview of many of the challenges associated with the realization of vision-based control of autonomous micro-air-vehicles in complex environments. Representative results from several enabling technologies have been presented, including vision-based state estimation, multiscale linear discriminant analysis (MLDA) for horizon and object detection, and wavelet-based learning algorithms for terrain reconstruction and obstacle estimation.

## REFERENCES

- [1] S. M. Ettinger, M. C. Nechyba, P. G. Ifju, and M. Waszak, "Vision-guided flight stability and control for Micro Air Vehicles," in *Proc. IEEE Int'l Conf. on Intelligent Robots and Systems*, Lausanne, Switzerland, 2002.
- [2] T. Kanade, O. Amidi, and Q. Ke, "3d and real-time vision systems for autonomous unmanned air vehicles," in *Proceedings of the IEEE 43rd Conference on Decision and Control*, 2004.
- [3] S. Soatto and P. Perona, "Reducing structure from motion part 1: Modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 9, pp. 933–942, September 1998.
- [4] —, "Reducing structure from motion part 2: Implementation and experimental assessment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 9, pp. 943–960, 1998.
- [5] S. Soatto, P. Perona, R. Frezza, , and G. Picci, "Motion estimation via dynamic vision," *IEEE Transactions on Automatic Control*, vol. 41, no. 3, pp. 393–413, March 1996.

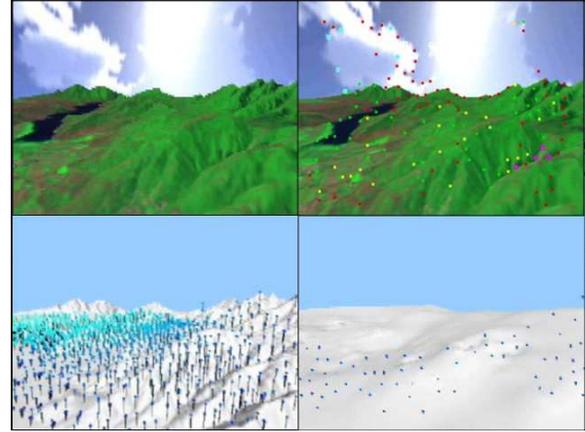


Fig. 13. Piecewise linear terrain reconstruction from a single frame: original image (upper left), image with feature points (upper right), deviation of 3D feature point estimates from the true surface (lower left), and reconstructed terrain with feature points (lower right).

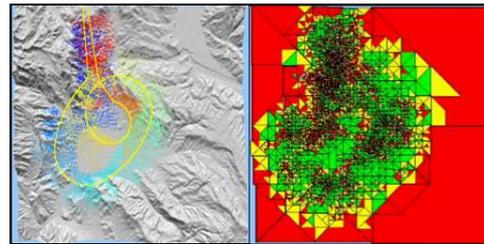


Fig. 14. Simulated MAV flight over mountainous terrain: identified feature points (left) and piecewise linear terrain reconstruction (right).

- [6] P. Gurfil and H. Rotstein, "Partial aircraft state estimation from visual motion using the subspace constraints approach," *Journal of Guidance, Control and Dynamics*, vol. 24, no. 5, pp. 1016–1028, September–October 2001.
- [7] T. Webb, R. Prazenica, A. Kurdila, and R. Lind, "Vision-based state estimation for autonomous micro-air-vehicles," in *Proceedings of the 45th AIAA Guidance, Navigation and Control Conference*, 2004.
- [8] S. Todorovic and M. Nechyba, "Detection of artificial structures in natural-scene images using dynamic trees," in *preprint*, 2004.
- [9] —, "Intelligent missions for MAVs: Visual contexts for control, tracking and recognition," in *preprint*, 2004.
- [10] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Belmont, CA: Wadsworth, 1984.
- [11] S. M. Ettinger, M. C. Nechyba, P. G. Ifju, and M. Waszak, "Vision-guided flight stability and control for Micro Air Vehicles," *Advanced Robotics*, vol. 17, 2003.
- [12] F. Cucker and S. Smale, "On the mathematical foundations of learning," *Bulletin of the American Mathematical Society*, vol. 39, no. 1, pp. 1–49, 2001.
- [13] S. Konyagin and V. Temlyakov, "Some error estimates in learning theory," in *preprint*, 2004.
- [14] R. DeVore and W. Dahmen, "Notes on learning theory," *preprint*, 2004.
- [15] R. DeVore, G. Kerkycharian, D. Picard, and V. Temlyakov, "On mathematical methods of learning," *preprint*, 2004.
- [16] M. Zabaranin, A. Kurdila, O. Prokopyev, A. Goel, V. Boginski, R. Causey, S. Uryasev, and P. Pardalos, "Vision-based trajectory planning for autonomous micro-air-vehicles," *preprint*, 2004.