MINIMAX GAME-THEORETIC APPROACH TO MULTISCALE H-INFINITY OPTIMAL FILTERING

Hamza Anwar and Quanyan Zhu

Electrical and Computer Engineering Department, New York University, Brooklyn, NY 11201, USA
Email: {ha1082, qz494}@nyu.edu

ABSTRACT

Sensing in complex systems requires large-scale information exchange and on-the-go communications over heterogeneous networks and integrated processing platforms. Many networked cyber-physical systems exhibit hierarchical infrastructures of information flows, which naturally leads to a multi-level tree-like information structure in which each level corresponds to a particular scale of representation. This work focuses on the multiscale fusion of data collected at multiple levels of the system. We propose a multiscale state-space model to represent multi-resolution data over the hierarchical information system and formulate a multi-stage dynamic zero-sum game to design a multi-scale H_∞ robust filter. We present numerical experiments for one and two-dimensional signals and provide a comparative analysis of the minimax filter with the standard Kalman filter to show the improvement in signal-to-noise ratio (SNR).

Index Terms— Multi-resolution analysis, dynamic games, state estimation, hierarchical systems, minimax techniques.

1. INTRODUCTION

Data processing of large-scale complex systems requires massive information exchange and communications that are often organized in a hierarchical manner in which data collected at multiple scales represent different resolutions. It is imperative that the fusion of the sensed data need to be cost-effective, energy-efficient, quality-ensuring, vulnerability-evasive, and overall robust in the process. Hence in this work, we aim to design a multi-scale minimax filter that allows state estimation which is robust to disturbances across multiple scales of the data collection.

To this end, we first build a multiscale state-space representation of the multiscale data collection process. Leveraging the state-space framework for modeling stochastic phenomenon at multiple scales in Chou et. al. [1], we formulate the robust multiscale state estimation problem using a zero-sum dynamic game approach. The development of multiscale state estimation algorithms provides a spatially dynamic approach towards fusing the multi-layer data and enables a scalable and implementable solution for the large-scale system.

To illustrate the multiscale model, we can consider crime incident reporting as an example. Crime incidents are reported at nearest precincts in the city. All precincts in a city have the aggregated criminal activity information of the whole city. Cities within a state can report this information at the state level. Similarly, the states of the country can aggregate criminal activity information from cities and report it to the federal government. This architecture can be considered as a pyramidal data structure with all precincts' data at the

finest level, city data at one coarser level, state data at one more coarser level, and so on. The data at each successive level is an aggregated representation of the same data, and thus data at each level is linearly related. Moreover, when the available data is noisy, a scheme for filtering out the unnecessary cofactors is often required. Optimal filtering minimizes the error between estimated and actual values across all levels while making the estimator robust to noise. Data is fused along coarser levels and interpolated along finer levels. This work formulates a filtering problem of this nature, where data are collected at multiple layers of different resolutions and the granularity of the information increases as the system zooms into finer levels.

We present a robust H_∞ filter that outperforms regular Kalman filter when the multiscale system is subject to additive noise in input and measurements. The filter operates in the direction of coarse-to-fine dynamics. Moreover, we extend our H_∞ filter to incorporate measurements at k-th level when estimating the state at k-th level (in contrast to a predictor-based design that relies on one previous step measurements). The use of game-theoretic H_∞ filter is not new in signal processing schemes [2]; however, up till now, it has not found applications in multi-resolution information systems because the tool of multiscale system modeling has not been widely used in all relevant domains owing to its complex nature [3]. In our work, we show the performance benefits of using the minimax robust H_∞ filter in contrast to Kalman filter for 1-D and 2-D signals.

Multiscale modeling has been well studied in the statistics and signal processing communities [4, 5], e.g., the development of wavelet analysis [3] and its applications in image compression and image restoration [6]. Among the relevant uses of multiscale models for estimation tasks are in geoscience and remote sensing especially when there are heterogeneous suites of sensors (infrared, visual, microwave, etc.) [7] and in hierarchical graphical modeling in machine learning [8]. In recent works for multi-layered systems, *transcale* optimal controller and robust state-estimator for discrete-time systems has been proposed by Zhao et. al. [9, 10]. Their theory uses wavelets decomposition to relate signals across resolution levels. The focus of our work, being starkly different, is in the fusion of sensor data of different resolutions as in [11]. Our work specifically deals with the multiscale data fusion problem that arises from the spatial hierarchical architectures of the sensors in contrast to time-domain multiscale filtering problems.

2. MULTISCALE STATE-SPACE ESTIMATOR DESIGN

Consider a dyadic tree structure of states: nodes are arranged in a layered setting. Fig. 1 represents $K + 1$ levels of a latent process on a dyadic tree. At the k-th level the (vectorized) process is denoted by x_k and its corresponding state-vector at m-th node is $x_k^m \in \mathbb{R}^n$. Thus, the multiple layers of the tree indicate different representations
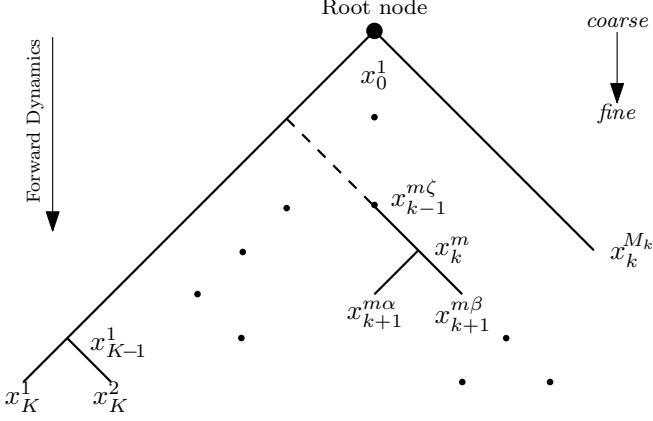
Fig. 1. Depiction of the dyadic tree structure for a multiscale linear system

of the complete latent process x. We assume that the nodes of the latent process at a given level are conditionally independent of each other given the state at an immediate coarser level. Our tree structure has the following form of the forward dynamics (coarse-to-fine):

$$
\begin{aligned}
x_{k+1}^{m\alpha} &= A_{k+1}^{m\alpha} x_k^m + B_{k+1}^{m\alpha} w_{k+1}^{m\alpha}, \\
x_{k+1}^{m\beta} &= A_{k+1}^{m\beta} x_k^m + B_{k+1}^{m\beta} w_{k+1}^{m\beta}, \\
y_k^m &= C_k^m x_k^m + v_k^m,
\end{aligned} \tag{1}
$$

$k \in \mathcal{K} := \{0, \ldots, K-1\}$; $m \in \mathcal{M}_k := \{1, \ldots, M_k\}$; $x_0^1 = x_0$,

where $v_k^m \in \mathbb{R}^p$ is the measurement noise with covariance given as $\mathbb{E}[v_k^m v_k^{m\top}] = R_k^m$, and $w_k^m \in \mathbb{R}^q$ the disturbance at $m(k)$-th node, having $\mathbb{E}[w_k^m w_k^{m\top}] = I$ w.l.o.g. Here, A_k^m represents the interpolation matrix, i.e., the relationship between system states at coarse and fine levels. B_k^m is the input matrix for disturbance signals. Along coarse-to-fine recursion, w signals actually represent the higher details added, but because coarse-to-fine recursion is analogous to the multiresolution synthesis of signals, w is rather interpreted as the disturbance signal corrupting our state trajectories x along resolution scales. For dyadic tree, the number of nodes at k-th level is $M_k = 2^k$. Since the coarsest level only has a single (root) node, we represent state of root node x_0^1 as merely x_0. For each m, $m\alpha := 2m-1$ and $m\beta := 2m$ indicate its two children nodes in the next finer level and $m\zeta := \lceil \frac{m}{2} \rceil$ is its parent node at a coarser level. We assume that $\{w^m\}$, i.e., the disturbance in m-th path along coarse-to-fine scales, is any l_2 sequence[1].

Ultimately, instead of estimating system states x_k^m, we are interested in estimating a linear combination of the system states x_k^m given by a general representation z_k^m such as

$$
z_k^m = L_k^m x_k^m. \tag{2}
$$

---

[1]The set representing the path from w_m^K to the root is $\{w^m\} = \{w_K^m, w_{K-1}^{m\zeta}, w_{K-2}^{(m\zeta)\zeta}, \cdots, w_0^1\}$. The sum of l_2 norms of each w_k^m along the path that leads to w_K^m from the root is bounded.

The measure of performance (objective) is given as follows:

$$
J =
$$
$$
\frac{\Sigma_{k=1}^K \Sigma_{m=1}^{M_k} ||z_k^m - \hat{z}_k^m||_{Q_k^m}^2}{||x_0 - \hat{x}_0||_{p_0^{-1}}^2 + \Sigma_{k=1}^K \Sigma_{m=1}^{M_k} \left[||w_{k+1}^{m\alpha}||_I^2 + ||w_{k+1}^{m\beta}||_I^2 + ||v_k^m||_{R_k^{-m}}^2 \right]}, \tag{3}
$$

where none of the denominator terms is exactly zero, $\hat{x}_0$ is an *a priori* estimate of x_0, $Q_k^m \geq 0$, $p_0^{-1} > 0$, I identity, and $R_k^m > 0$ are the weighting matrices. The optimal estimate of z_k^m, for a given attenuation $\gamma > 0$, should satisfy

$$
\sup J < 1/\gamma.
$$

We can rewrite the objective as a minimax problem

$$
\min_{\hat{x}_k^m} \max_{y_k^m, w_k^m, x_0} J = \frac{1}{2} \sum_{k=1}^K \sum_{m=1}^{M_k} \{ ||x_k^m - \hat{x}_k^m||_{\bar{Q}_k^m}^2 - \frac{1}{\gamma} (||w_{k+1}^{m\alpha}||_I^2
$$
$$
+ ||w_{k+1}^{m\beta}||_I^2 + ||y_k^m - C_k^m x_k^m||_{R_k^{m-1}}^2) \} - \frac{1}{2\gamma} ||x_0 - \hat{x}_0||_{p_0^{-1}}^2 \tag{4}
$$

subject to (1) where $\bar{Q}_k^m = L_k^{m^T} Q_k^m L_k^m$.

Theorem 1. *For noise attenuation $\gamma > 0$, an H_∞ filter for x_k^m exists if and only if there exists a stabilizing solution $P_k^m > 0 \ \forall k, m$, to the following coupled-pair of discrete-domain Riccati equations:*

$$
\begin{aligned}
P_{k+1}^{m\alpha} &= A_{k+1}^{m\alpha} P_k^m (I - \gamma \bar{Q}_k^m P_k^m + C_k^{m^T} R_k^{m-1} C_k^m P_k^m)^{-1} A_{k+1}^{m\alpha^T} \\
&\quad + B_{k+1}^{m\alpha} B_{k+1}^{m\alpha^T}, \\
P_{k+1}^{m\beta} &= A_{k+1}^{m\beta} P_k^m (I - \gamma \bar{Q}_k^m P_k^m + C_k^{m^T} R_k^{m-1} C_k^m P_k^m)^{-1} A_{k+1}^{m\beta^T} \\
&\quad + B_{k+1}^{m\beta} B_{k+1}^{m\beta^T},
\end{aligned} \tag{5}
$$

where $P_0 = p_0$. The H_∞ filter is given by

$$
\hat{z}_k^m = L_k^m \hat{x}_k^m, \quad k = 1, 2, \ldots, K
$$

where

$$
\begin{aligned}
\hat{x}_{k+1}^{m\alpha*} &= A_{k+1}^{m\alpha} \hat{x}_k^m + K_{k+1}^{m\alpha} (y_k^m - C_k^m \hat{x}_k^{m*}), \\
\hat{x}_{k+1}^{m\beta*} &= A_{k+1}^{m\beta} \hat{x}_k^m + K_{k+1}^{m\beta} (y_k^m - C_k^m \hat{x}_k^{m*}), \\
K_{k+1}^{m\alpha} &= A_{k+1}^{m\alpha} P_k^m (I - \gamma \bar{Q}_k^m P_k^m + C_k^{m^T} R_k^{m-1} C_k^m P_k^m)^{-1} \\
&\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad C_k^{m^T} R_k^{m-1}, \\
K_{k+1}^{m\beta} &= A_{k+1}^{m\beta} P_k^m (I - \gamma \bar{Q}_k^m P_k^m + C_k^{m^T} R_k^{m-1} C_k^m P_k^m)^{-1} \\
&\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad C_k^{m^T} R_k^{m-1}.
\end{aligned} \tag{6}
$$

Remark 1: The optimization problem (4) is a zero-sum game with $\hat{x}^m$ as the decision variable of the minimizer player, and x_0, w_k^m, y_k^m as the decision variables of the maximizer player. The game is subject to the dynamic constraint of (1), making it a dynamic game. In the proof for Theorem 1 (not shown here) we argue that the formulated problem is equivalent to a two-player zero-sum dynamic game, solving which gives us the saddle-point equilibrium solution in (6).

Remark 2: The proposed estimator is a predictor-based estimator; i.e., it makes use of observations at stage $k-1$ to estimate the states at stage k, while not using stage k observations. This approach

can degrade the performance substantially because the size of state vectors increases geometrically as the number of stages of the multiscale system increases. Kalman filter implemented in [12] implicitly uses the current measurements. As an extension here, we now incorporate the observations at the present scale. For linear time-domain systems, Green et. al. [13] have constructed the H_∞ filter in generalized frameworks. However, for multiscale settings, such extensions are unavailable, and thus we extend the generalized H_∞ filter by Green et. al. [13, B.3.1] done for discrete-time systems, to our multiscale state-space setting that also builds upon our main result of Theorem 1.

The numerator in our cost function (4) is modified for the estimator based on present scale measurements by making the per-node state-estimation error as follows:

$$e_k^m = ||\hat{z}_{k|k}^m - L_k^m x_k^m||_{Q_k^m}^2.$$

With the new cost function J, the optimization problem solution gives the estimator design of (7). For our experiments, we have used both the predictor-based and current-measurements-based estimators. If measurements at present scale are unavailable, then the results in Theorem 1 are reduced to:

$$\hat{z}_{k|k}^m = L_k^m \hat{x}_{k|k}^m,$$
$$\hat{x}_{k|k-1}^m = A_k^m \hat{x}_{k-1|k-1}^{m\zeta},$$
$$P_{k|k-1}^m = A_{k-1}^{m\zeta} P_{k-1|k-1}^{m\zeta} A_{k-1}^{m\zeta\top} + B_{k-1}^{m\zeta} B_{k-1}^{m\zeta\top},$$
$$\hat{x}_{k|k}^m = \hat{x}_{k|k-1}^m + K_k^m(y_k^m - C_k^m x_{k|k-1}^m),$$
$$K_k^m = P_{k|k-1}^m C_k^{m\top}(R_k^m + C_k^m P_{k|k-1}^m C_k^{m\top})^{-1},$$
$$P_{k|k}^m = P_{k|k-1}^m - P_{k|k-1}^m [C_k^{m\top} \; L_k^{m\top}][\tilde{R}_k^m]^{-1} \begin{bmatrix} C_k^m \\ L_k^m \end{bmatrix} P_{k|k-1}^m,$$
$$\tilde{R}_k^m = \begin{bmatrix} R_k^m & 0 \\ 0 & -I/\gamma \end{bmatrix} + \begin{bmatrix} C_k^m \\ L_k^m \end{bmatrix} P_{k|k-1}^m [C_k^{m\top} \; L_k^{m\top}].$$
(7)

3. NUMERICAL EXPERIMENTS

Fig. 2 shows various levels of a multiscale process on the interval $[0, 1]$. The process is a step function, and each sub-interval corresponds to a node of an autoregressive process on a dyadic tree. We show that as more stages are observed, the confidence interval around state estimates (red) becomes tighter giving more precision and they also tend to coincide with original signal (blue) showing accuracy. We let A, B and C be identity matrices in all our experiments for simplicity.

2-D signals have similar hierarchical modeling structure compared to 1-D signals, see Fig. 3. The computation overload is also not substantial because the way we model it requires more children per node. We test our algorithm on different RGB and grayscale images. The first step we take into account before estimation is the construction step that takes the finest image, i.e., the actual image, and generates its low-resolution versions step by step to yield images at all tree levels, Fig. 4. Unlike our 1-D examples, here, we assume that the original signal exists in the continuous real domain, i.e., as $K \to \infty$, x_K approaches the true signal.

A series of experiments have been performed with 2-D image data. To generate a layered data of this signal at lower resolution scales, we have averaged over a neighborhood of four pixels each recursively in a pyramid-like structure, see Fig. 3. Other methods such as frequency-domain representations can also be used, but they will not affect our results. Once we have the data at various levels,
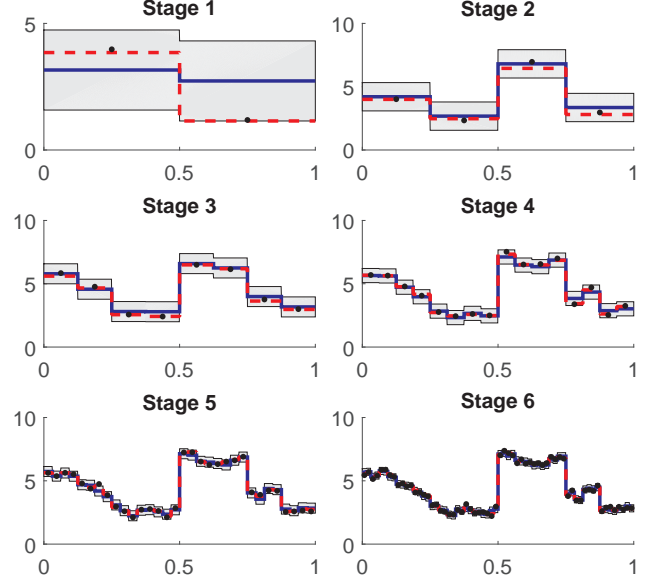
Fig. 2. Example of a one-dimensional multiresolution signal represented at six different levels. Gray 70% confidence bounds indicate the uncertainty in state transitions. Observations are indicated by black dots, while the original signal is in blue. State estimates recovered by proposed filter ($\gamma = 1$) from observations at all resolution levels are indicated by red dotted line.
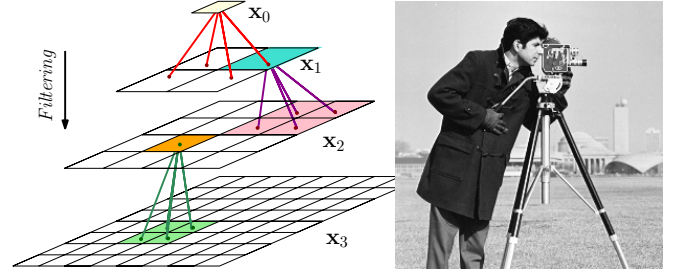
Fig. 3. Hierarchical multiscale structure for 2-D signals (left) and the cameraman image that we take as the finest resolution signal of our multiscale signal to use in experiments (right).

we use it to perform filtering by giving different noisy and corrupted observations to our system. In Fig. 5, we show that our method outperforms the standard Kalman filtering approach for multiscale models, giving higher SNR values and less block-like artifacts. Here, the cameraman image at 256×256 resolution serves as the finest resolution signal.

4. DISCUSSION AND CONCLUSION

Using the minimax game-theoretic approach to the optimal estimation problem, we have avoided solving the smoothing and filtering problems separately that has classically been considered in multiscale recursive estimation [12, 14]. For our estimator, we have avoided the reverse (fine-to-coarse) dynamics by only considering natural dynamics. Methodology employed by Chou et. al. [12] uses reverse dynamics as it involves two-sweep estimation process: first filtering and then smoothing. The reason for doing so is to allow

Fig. 4. Different resolution levels of a 2-D signal constructed by recursive averaging as shown in Fig. 3. These images after adding noise will serve as the observations for the state-estimator to recover original image.
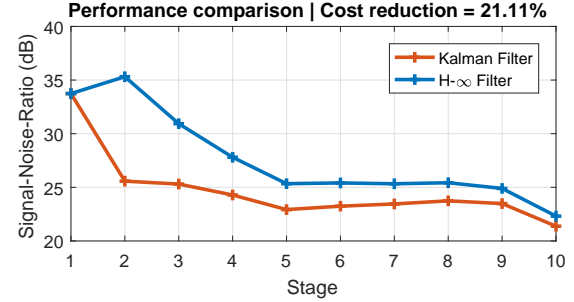

Fig. 5. Estimated image at finest resolution level using H_∞ filter with $\gamma = 185.44$ (left) and Kalman filter (right). High SNR for H_∞ estimate corresponds to a better image as is shown: block-like artifacts are visible near the edges in the image on right.

Fig. 6. Comparison of performance in cost reduction for standard Kalman and the proposed estimators ($\gamma = 84.44$).

Fig. 7. Trend of uncertainty (covariance) in state-estimates is shown. When there is no observation recorded at a given stage, uncertainty in the state estimate of that stage increases (left). As the system reaches the steady-state, the covariance is shown to decrease (right).

the measurement at one node to contribute towards the estimate at another node on the same resolution level. In our approach, this technicality is taken care of by the reverse Riccati recursion in (5).
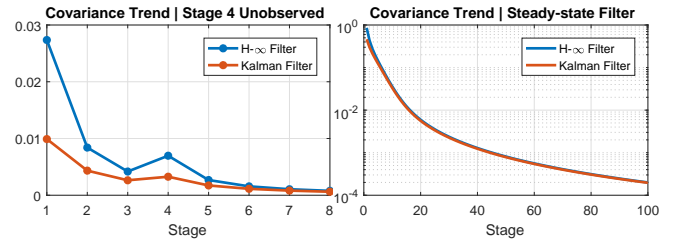
In Fig. 6, we show that by using H_∞ filter instead of standard Kalman filter, a large reduction in the estimation cost of 21% is achieved (i.e. calculated by accumulating squared differences between original and state estimates). Across all the stages, we see that H_∞ filter exhibits a higher value of signal-to-noise ratio in comparison with Kalman filter.

One important feature of the multiscale filter is the scenario of missing information at different nodes across the tree. The information at neighboring nodes can improve the state estimate of nodes where no information is recorded. We perform an experiment in which observations have not been recorded for any node of stage 4. The estimators can still give considerably good state estimates for nodes at stage 4, with a yet higher uncertainty. This increase in estimation uncertainty is shown in Fig. 7. Although we have observed a lower variance for Kalman filter, the variance converges in Fig. 7 as we move towards finer stages. Note that when the proposed H_∞ filter is used, the covariance update in (5) is not the mere Lyapunov update because of a nonzero γ. However, the mean update in (6) is predicted based on state matrix A only.

In this paper, we have presented an optimal robust filter design for a class of multiscale discrete domain systems that combats worst-case additive exogenous process and sensor noise. Our filter design gives the saddle-point solution which is the equilibrium strategy of the minimax two-player zero-sum dynamic game. We have given the conditions for an optimal solution both for the predictor-based estimator (6) and the current-measurements-based estimator (7) for a given disturbance attenuation level γ. Experimental results corroborate our theoretical findings, and we have shown our evaluations for 1-D and 2-D signal examples. Our work is not limited to the sce-

narios of images and one-dimensional signals, and in fact, we have argued that for the multiscale state-space modeling framework, the application of this work is not limited by computational or memory constraints. Moreover, the expressiveness of this model is used in various domains such as graphical topic modeling in machine learning [8]. Multi-resolution phenomena are widely found in numerous IoT-related applications. And this fact makes our work useful for different domains. We also show the convergence of state trajectories in steady-state and a comparative analysis with the standard Kalman filter. Lastly, dealing with missing information and the distributed nature of such models would be a future direction of our work.

5. REFERENCES

[1] Kenneth C Chou, *A stochastic modeling approach to multiscale signal processing*, Ph.D. thesis, Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1991.

[2] Xuemin Shen and Li Deng, "Game theory approach to discrete h-infinity filter design," *IEEE Transactions on Signal Processing*, vol. 45, no. 4, pp. 1092–1095, 1997.

[3] Marco AR Ferreira and Herbert KH Lee, *Multiscale modeling: a Bayesian perspective*, Springer Science & Business Media, 2007.

[4] SC Clippingdale and RG Wilson, "Least-squares image estimation on a multiresolution pyramid," in *Acoustics, Speech, and Signal Processing, 1989. ICASSP-89., 1989 International Conference on*. IEEE, 1989, pp. 1409–1412.

[5] Alan S Willsky, "Multiresolution markov models for signal

and image processing," *Proceedings of the IEEE*, vol. 90, no. 8, pp. 1396–1458, 2002.

[6] Mark R Banham and Aggelos K Katsaggelos, "Spatially adaptive wavelet-based multiscale image restoration," *IEEE Transactions on Image Processing*, vol. 5, no. 4, pp. 619–634, 1996.

[7] Sheng Zheng, Wen-zhong Shi, Jian Liu, and Jinwen Tian, "Remote sensing image fusion using multiscale mapped ls-svm," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 5, pp. 1313–1322, 2008.

[8] Amr Ahmed, Liangjie Hong, and Alexander J Smola, "Hierarchical geographical modeling of user locations from social media posts," in *Proceedings of the 22nd international conference on World Wide Web*. ACM, 2013, pp. 25–36.

[9] Lin Zhao and Yingmin Jia, "Transcale control for a class of discrete stochastic systems based on wavelet packet decomposition," *Information Sciences*, vol. 296, pp. 25–41, 2015.

[10] Lin Zhao and Yingmin Jia, "Robust transcale state estimation for multiresolution discrete-time systems based on wavelet transform," *IET Signal Processing*, vol. 7, no. 3, pp. 228–238, 2013.

[11] Lei Zhang, Xiaolin Wu, Quan Pan, and Hongcai Zhang, "Multiresolution modeling and estimation of multisensor data," *IEEE transactions on Signal Processing*, vol. 52, no. 11, pp. 3170–3182, 2004.

[12] Kenneth C Chou, Alan S Willsky, and Albert Benveniste, "Multiscale recursive estimation, data fusion, and regularization," *IEEE Transactions on Automatic Control*, vol. 39, no. 3, pp. 464–478, Mar 1994.

[13] Michael Green and David JN Limebeer, *Linear robust control*, Courier Corporation, 2012.

[14] Krishan M Nagpal and Pramod P Khargonekar, "Filtering and smoothing in an h setting," *IEEE Transactions on Automatic Control*, vol. 36, no. 2, pp. 152–166, 1991.