LSTM Based Adaptive Filtering for Reduced Prediction Errors of Hyperspectral Images

Zhuocheng Jiang and W. David Pan
Dept. of Electrical and Computer Engineering
University of Alabama in Huntsville (UAH)
Huntsville, Alabama 35899

Hongda Shen
Bank of America Corporation
New York, NY 10020

THE UNIVERSITY OF
ALABAMA IN HUNTSVILLE

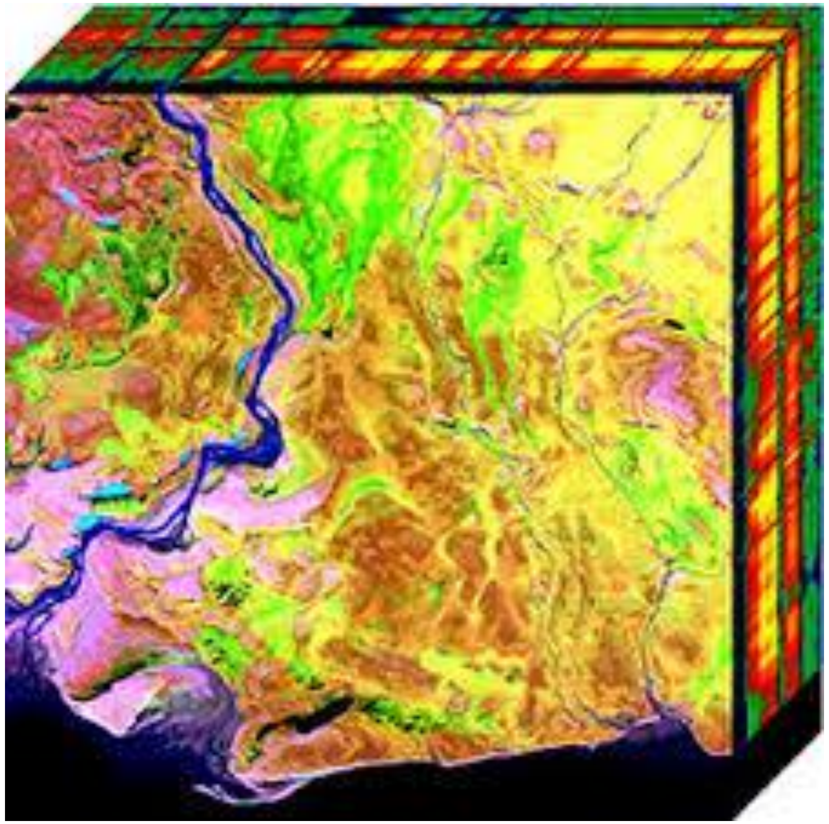
Outline

- Research Background and Motivation
- Famework of Predictive Lossless Compression
- Review of Traditional Adaptive Filter
- Long Short Term Memory (LSTM) Neural Network
- LSTM Neural Network for Weight Sequence Prediction
- The Proposed Famework
- Simulation Results
- Conclusions

THE UNIVERSITY OF
ALABAMA IN HUNTSVILLE

Hyperspectral Imaging

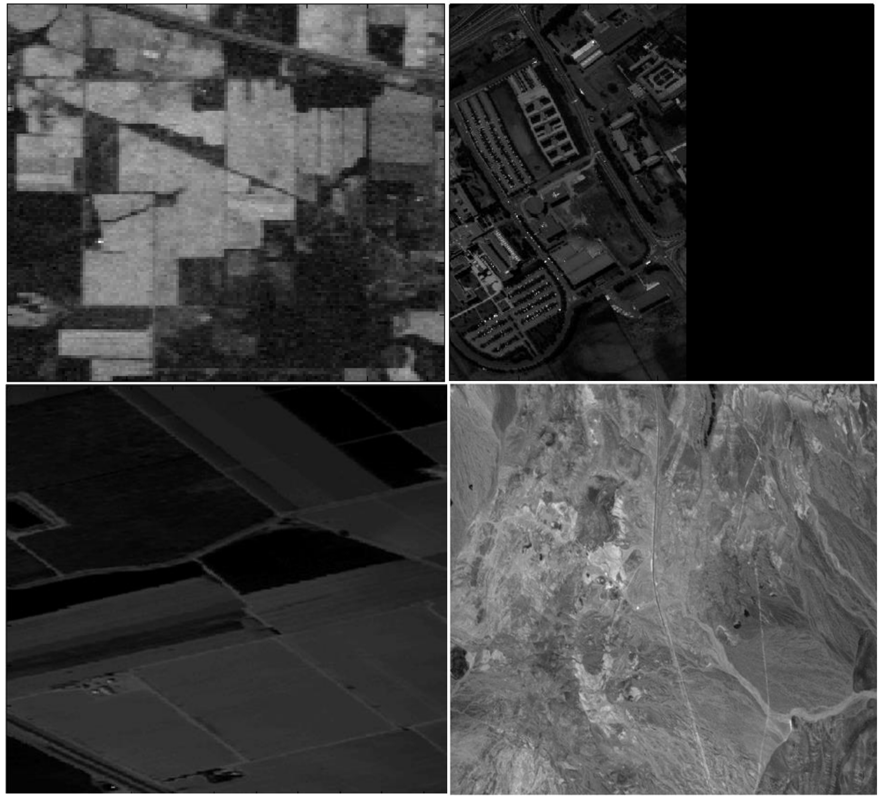
- Hyperspectral imaging technique is a combination of digital imaging and spectroscopy.
- Hyperspectral camera acquires the light intensity for a large number of contiguous spectral bands.
- Captured information can be used to characterize the objects in the scene with great precision and detail.



THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

Why Compression of Hyperspectral Images is Necessary?

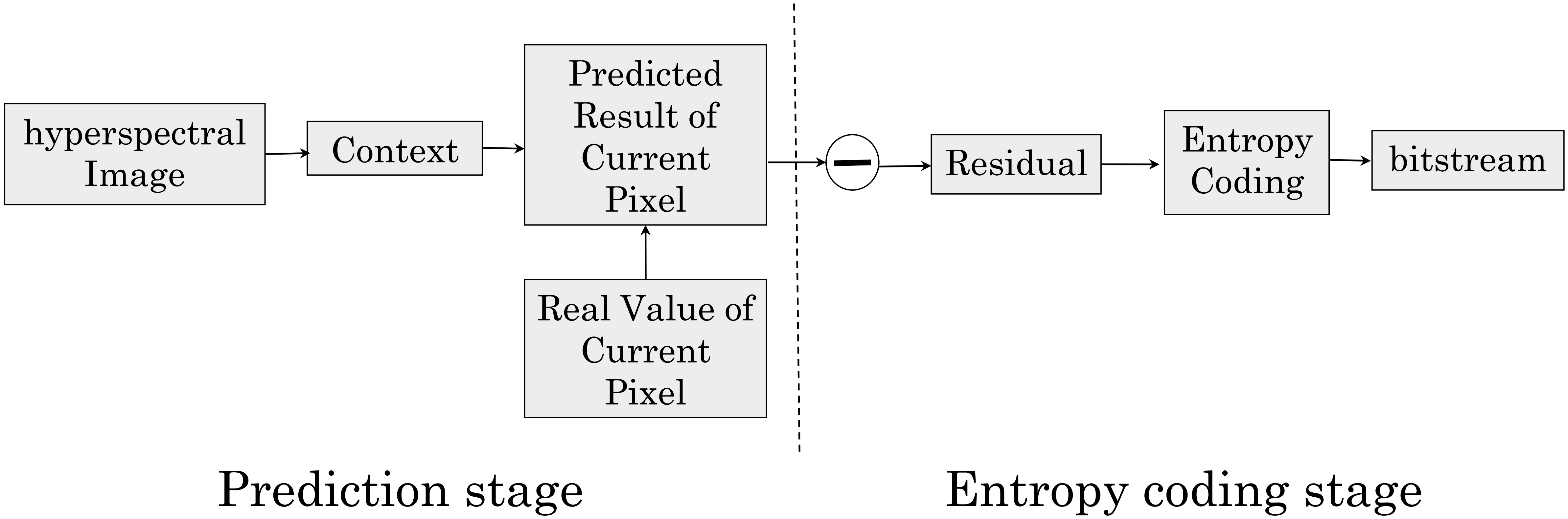
- Hyperspectral image sensor has limited memory capacity, thus storage of large images are challenging.
- Very large size of hyperspectral data makes transmission tasks very difficult.
- Compressing the hyperspectral images losslessly are highly valued in remote sensing applications.



THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

Predictive Lossless Compression

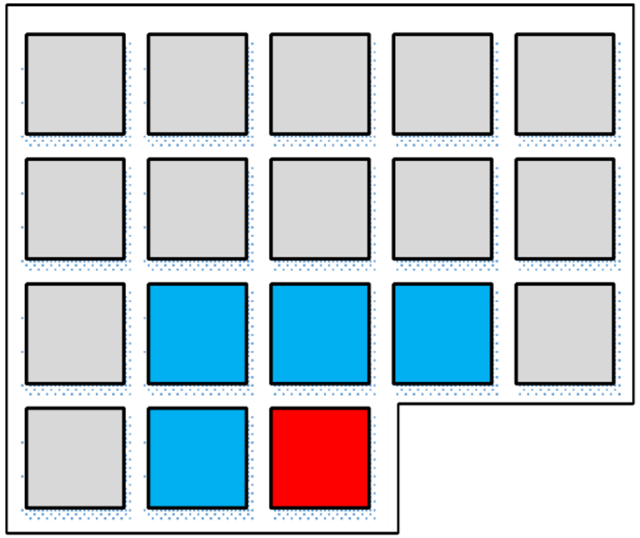
- Two-stage prediction framework



THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

Context Selection

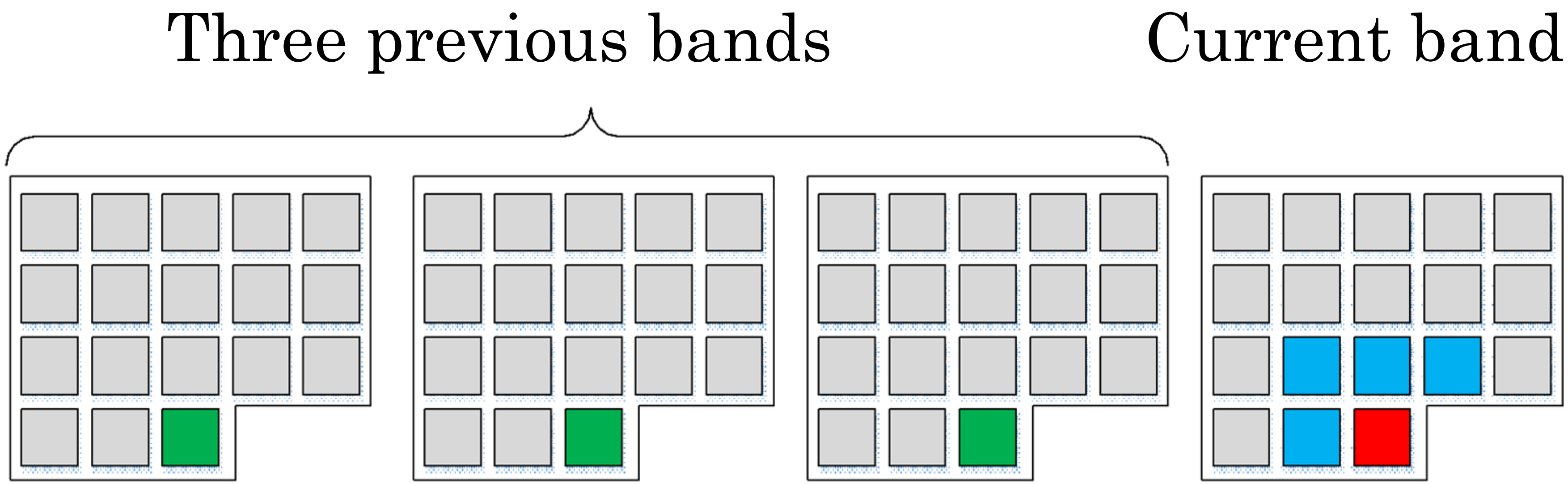
- Prediction-based lossless compression approaches take advantage of the strong correlations of image signals.
  - Spatial correlation
  - Spectral correlation
- To exploit the correlations, spatial context and spectral context are selected separately.
  - Current pixel (in red)
  - Spatial context (four nearest pixels colored in blue).



THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

Context Selection

- To expolit spectral correlations, three co-located pixels from three previous bands (colored in green in figure below) are chosen as spectral context.



THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

Traditional Adaptive Filter

- Since spectral correlations are much stronger than spatial correlations, we preform a context-based conditional average prediction (CCAP) [1] first to reduce the entropy.
- Let $s(x,y,z)$ be the pixel value at spatial location (x,y) in the spectral band z, the CCAP operation can be written as:

$$\bar{s}(x,y,z)=\frac{1}{|s(i,j,z)|}\sum_{(m,n,z)\in s(i,j,z)} s(m,n,z)$$

where $s(i,j,z)$ consists four neighborhood pixels (spatial context) in the current band, and $|s(i,j,z)|=4$ in our case.

[1] H. Wang, S. Babacan, and K. Sayood, "Lossless hyperspectral-image compression using context based conditional average," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4187-4193, Dec. 2007.

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

Traditional Adaptive Filter

- In adaptive filtering, the estimated pixel value is calculated as $W_n^T X_n$, where X_n and W_n^T are context vector and the corresponding weight vector.
- The prediction error is $e_n = d_n - W_n^T X_n$, where d_n is the actual pixel value. The error is used to adjust the filter weights interatively with a small learning rate μ:

$$W_{n+1} = W_n + \mu u_n e_n$$

- H. Shen proposed a maximum correntropy creteria (MCC) based LMS [2] by replacing the original mean square error with correntropy. The weight updating scheme can be written as:

$$W_{n+1} = W_n + \frac{\mu}{\sqrt{2\pi}\sigma^3} \exp\left(\frac{-e_n^2}{2\sigma^2}\right) e_n u_n$$

[2] H. Shen and W. D. Pan, "Prediction lossless compression of regions of interest in hyperspectral image via maximum correntropy criterion based least mean square learning," in P*roc. IEEE. Conf. Image Process.* Sep. 2016

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

Research Objectives and Novelty

- Traditional filtering methods do not take into account the longer-term dependencies of the data to be predicted.
- Motivated by the effectiveness of recurrent neural networks in capturing data memory for time series prediction, we design LSTM (long short-term memory) networks that can learn the data dependencies directly from filter weight variations.
- The trained networks are used to regulate the weights generated by conventional filtering schemes through a close-loop configuration.
- We compare the proposed method with two other memory-less algorithms, including
  - Least Mean Square (LMS) filtering method (widely used)
  - LMS variant based on the maximum correntropy criterion (MCC)

THE UNIVERSITY OF
ALABAMA IN HUNTSVILLE

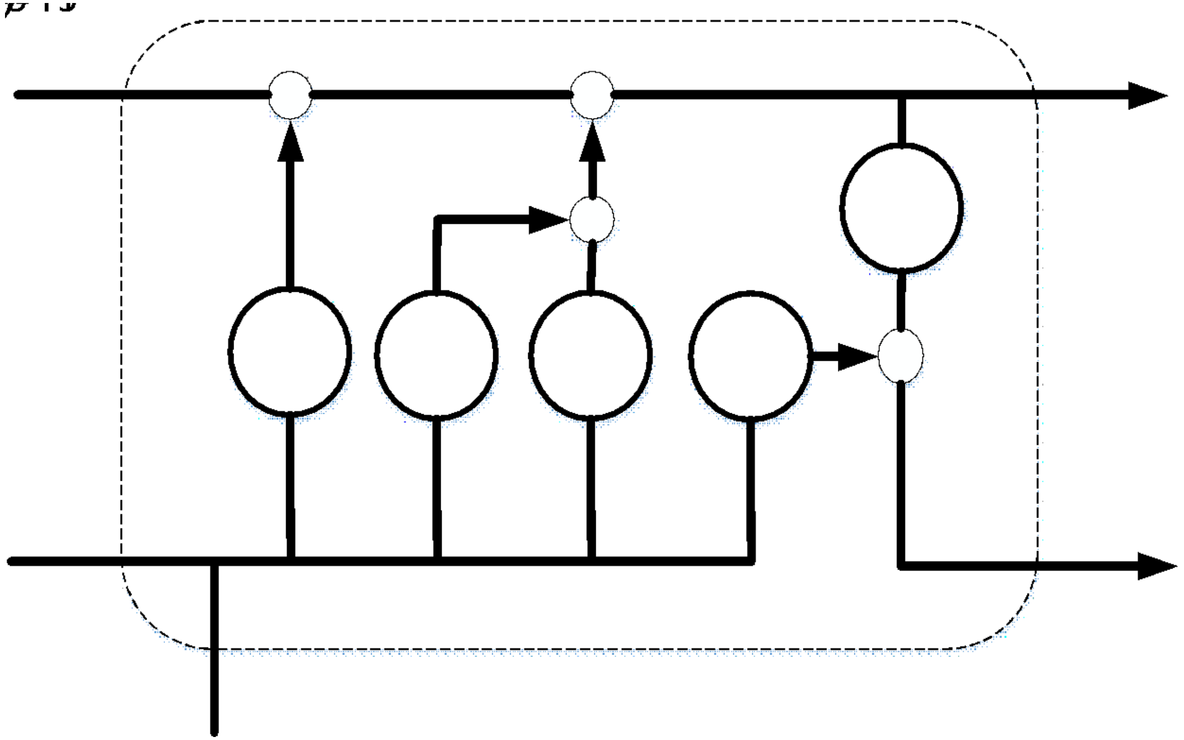
Long Short Term Memory (LSTM) Neural Network

- Learning to store information over extended time intervals via recurrent neural network (RNN) takes a very long time, due to vanishing gradient issue.
- The long short term memory (LSTM) network proposed in [3], addresses this problem effectively by introducing multiplicative gate units.
- By learning to open and close those gate units, the LSTM network can provide continuous analogues of write, read and reset operations for a cell in a digital cell.

[3] S. Hochreiter and J. Schmidhuber, Long short-term memory," *Neural computation*,
vol. 9, no. 8, pp. 1735{1780, Nov. 1997.

Long Short Term Memory (LSTM) Network

- A basic LSTM unit consisting of a self-connected memory cell with three multiplicative gates:
- The input gate i_t, output gate o_t, and forget gate f_t. The input data x_t.
- The output data from previous time step h_{t-1} are fed to each gate to determine the current cell state C_t , and the output h_t .



Long Short Term Memory (LSTM)

- where W_f W_i and W_o are weight matrices grouped with the corresponding gate, and the σ is the sigmoid function $\sigma(x) = 1/1 + e^{-x}$. Cell state $\tilde{C}_t$ and $\hat{C}_t$ are candidate values that can be added to the cell state and output, both of them are computed through a tanh layer:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t * \hat{C}_t,$$

$$g_{\tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

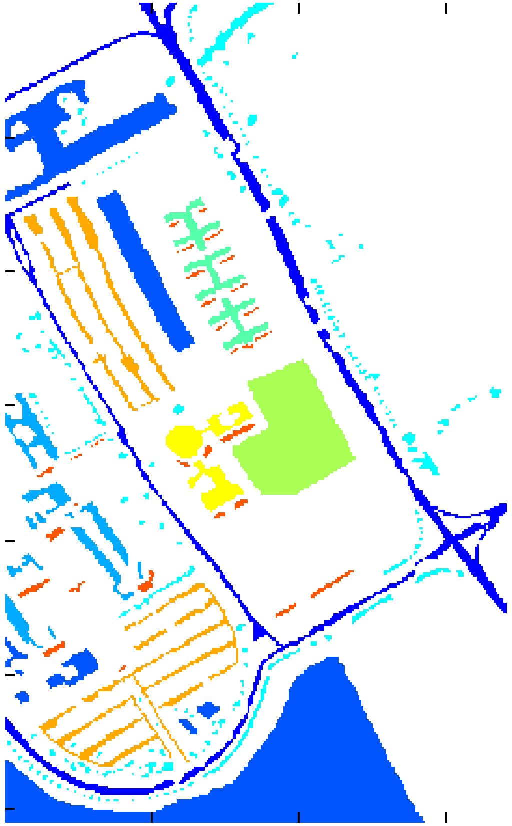
$$\tilde{C}_t = g(W_c \cdot [h_{t-1}, x_t] + b_c), \; and \;\; \hat{C}_t = g(C_t).$$

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

LSTM for Weight Sequence Prediction

Pavia University (PU) dataset

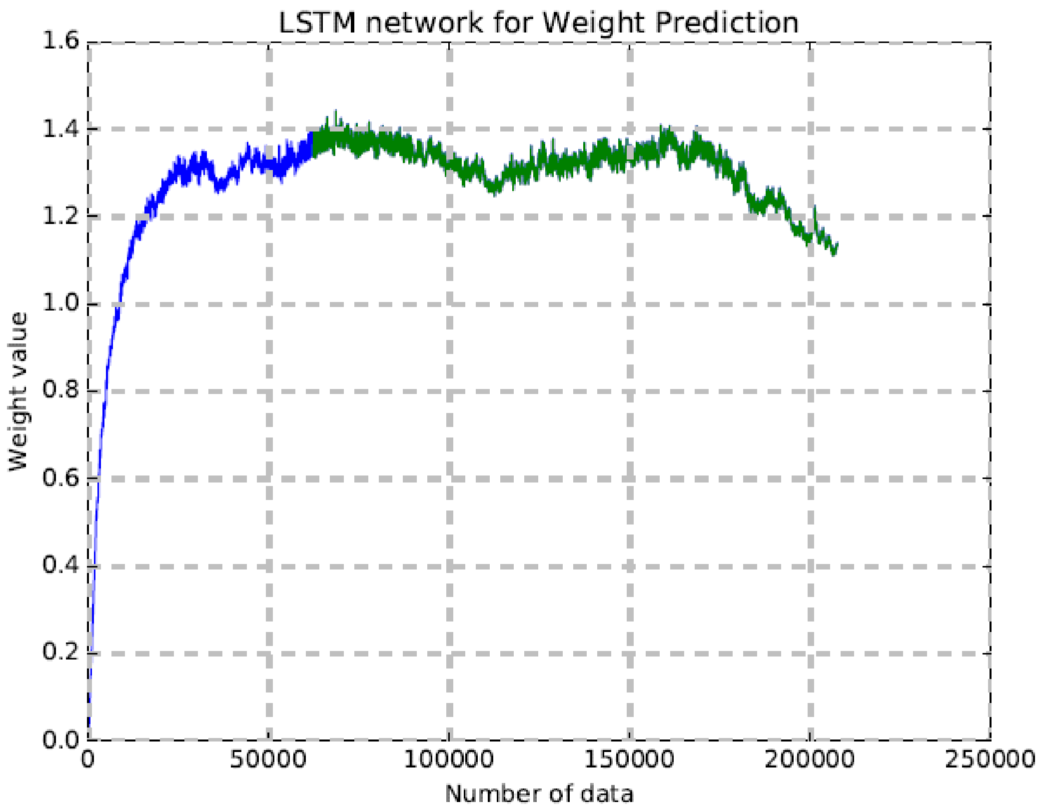
- Scene acquired by the ROSIS (Reflective Optics System Imaging Spectrometer) sensor during a flight campaign over Pavia University, in northern Italy.
- The PU dataset has 103 spectral bands, each band is a 610×610 pixel image.
- Ground truth of the PU dataset has 9 classes.



THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

LSTM for Weight Sequence Prediction

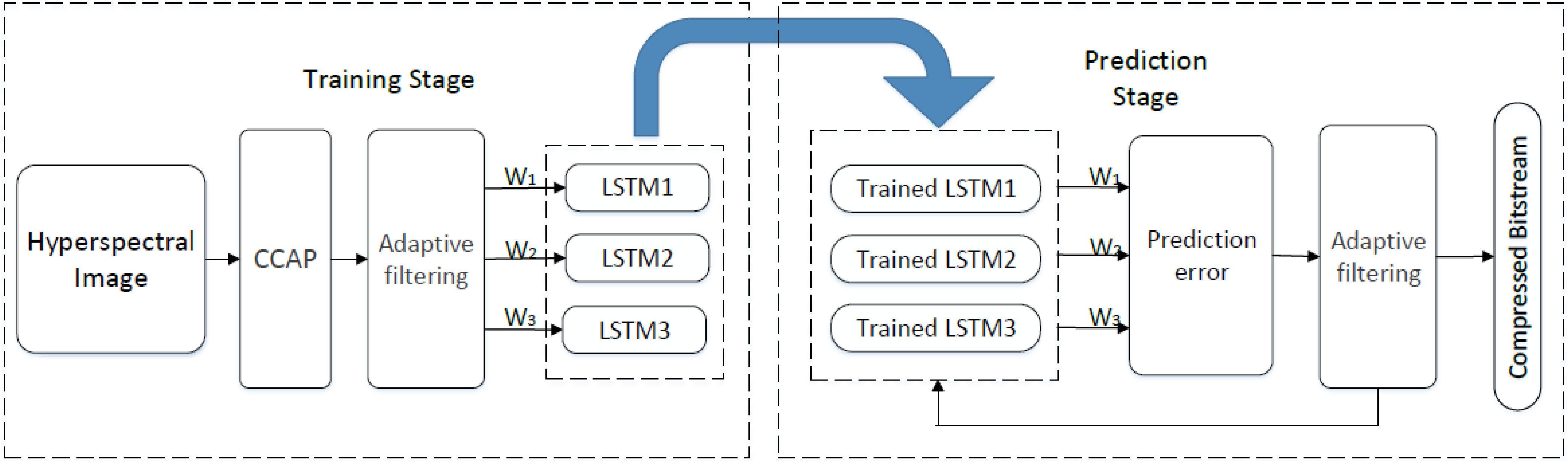
- Performance of LSTM network for weight prediction on PU data-set.
  - 30% of data for training.
  - 20% of data for validation.
  - 50% of data for testing.
- All the weights were colored in blue, and the prediction results were colored in green.



THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

The Proposed Famework

- LSTM neural networks learn the weight variations from the weight sequences directly.
- The trained networks are used to regulate the weights generated by conventional filtering schemes through a close-loop configuration.



THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

The Proposed Famework

The weight updating formula at the n^{th} time instant for the filtering operation can be written as:

$$W_{LSTM}(n) = LSTM(W(n)),$$

$$E_{LSTM}(n) = d_n - W_{LSTM}^T(n) X_n,$$

$$W_{LSTM}(n+1) = W_{LSTM}(n) + \frac{\mu}{\sqrt{2\pi\sigma^3}} \exp\left(\frac{-E^2{}_{LSTM}(n)}{2\sigma^2} \right) E_{LSTM}(n) X_n$$

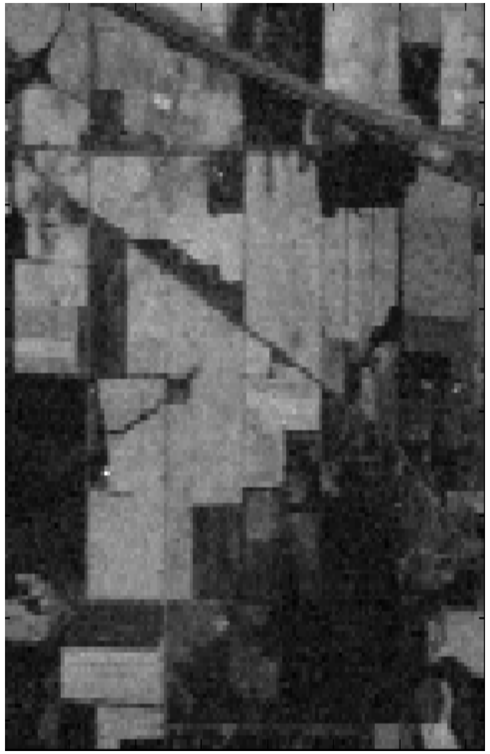
where $W(n)$ is the weight vector generated by adaptive filter,
W_{LSTM} is the weight vector predicted by the LSTM network,
and $E_{LSTM}(n)$ is the prediction error of the current pixel.

THE UNIVERSITY OF
ALABAMA IN HUNTSVILLE

Simulation Results

Indian Pines (IP) dataset

- Scene was gathered by the AVIRIS sensor over the Indian Pines site in north-western Indiana.
- IP dataset has 224 spectral bands, each band is a 145×145 pixel image.
- Ground truth of the IP dataset has 16 classes.
- The scene contains agriculture, forest and other natural vegetation.



Simulation Results

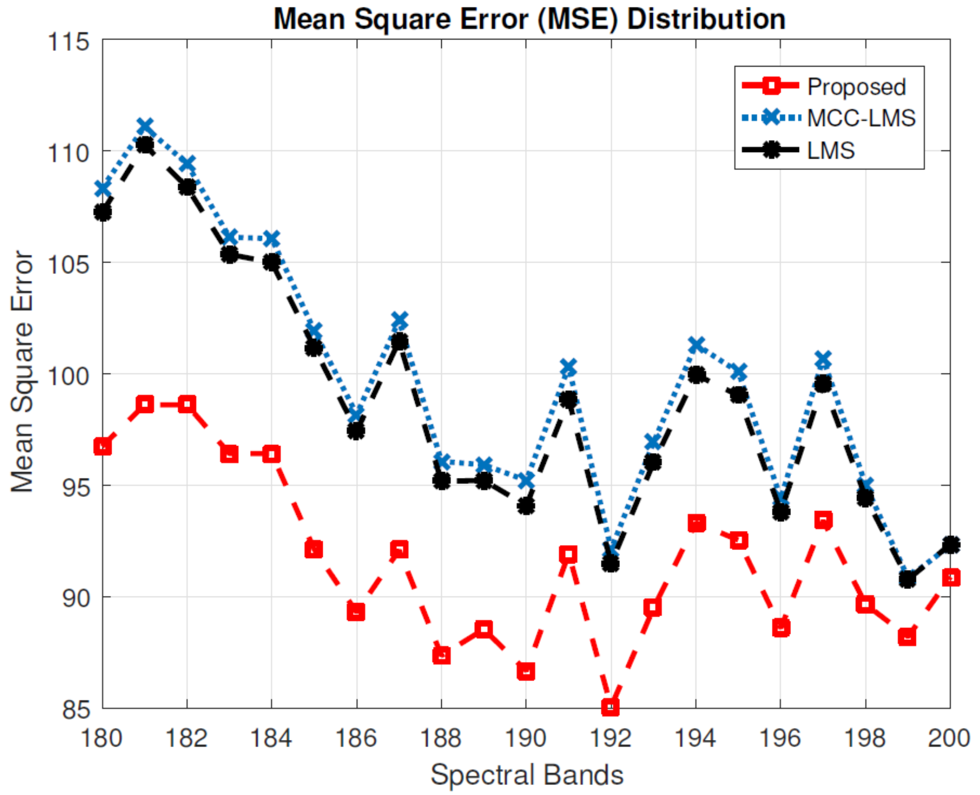
We compare our algorithm with two existing adaptive filtering methods:

- The adaptive LMS method as the new CCSDS standard for hyerpsectral data compression.
- MCC-LMS filtering based predictive compression algorithm, which replaced the cost function of LMS with correntropy.

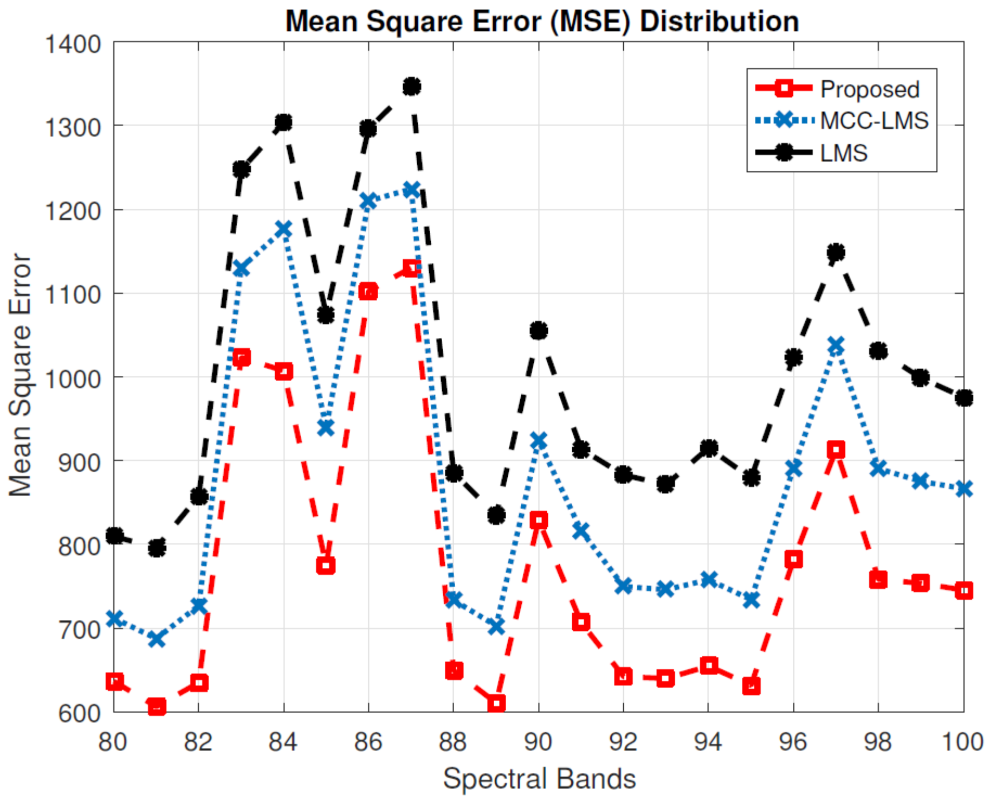
Datasaet	LMS	MCC-LMS	Proposed
Indian Pines	110.8	105.7	**104.6**
Pavia University	47.4	46.3	**45.7**

Simulation Results

Indian Pines (IP)



Pavia University (PU)



THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

Conclusions and Further Work

- We presented a novel adaptive filtering algorithm using LSTM network for hyperspectral images.
- LSTM networks appear to be effective in capturing the longer term dependencies of weight sequences.
- We proposed a two-stage framework by combining the trained LSTM networks with adaptive filters in a closed-loop configuration.
- To the best of our knowledge, this is the first attempt to model not only the correlations between pixels from different spectral bands, but also the temporal dependencies of the filtering weights.
- As future research, we will evaluate the impact of reduced prediction errors on predictive lossless coding performance.
- We will also analyze the long term dependencies in weight sequences.