

**THURSDAY
FEBRUARY 22, 2007**

**Scaife Hall Auditorium
Room 125**

4:30 p.m.
Refreshments—4:00 p.m.

venu govindaraju

PROFESSOR
UNIVERSITY OF BUFFALO



Dr. Venu Govindaraju is a Professor of Computer Science and Engineering at the University at Buffalo (UB), State University of New York. He received his B-Tech (Honors) from the Indian Institute of Technology (IIT), Kharagpur, India in 1986, and his Ph.D. degree in Computer Science from UB in 1992.

Govindaraju has co-authored more than 225 scientific papers. He is the PI on current NSF sponsored projects on International Digital Libraries and Digital Government. He is the founding director of the Center for Unified Biometrics and Sensors.

Govindaraju serves on the editorial boards of several leading Pattern Recognition journals including the IEEE Transactions. He is a Fellow of the IEEE and the IAPR.

ECE Seminar Hosts:

Radu Marculescu,
radum@ece.cmu.edu

Ken Mai,
kenmai@ece.cmu.edu

Marios Savvides,
Marios.Savvides@ri.cmu.edu

Paradigms in Handwriting Recognition for Varying Degrees of Domain Constraints

Recognition of handwritten documents has been a success in well-constrained domains, such as automatic recognition of handwritten postal addresses and validating bank check amounts. However, unconstrained offline handwriting recognition is still a very challenging task when lexicon size is greater than 1K. Thus, the large amount of handwritten material of the world, historical manuscripts and contemporary documents, are still inaccessible by search engines.

We will describe the seminal work in real-time handwriting recognition which paved the way for the first deployment of end-to-end handwritten address interpretation system for the USPS. This will include the several paradigms of handwriting recognition which contributed to this pioneering work, in particular an adaptive character recognition algorithm that actively determines the features, their resolution, and the length of the feature vector depending on a dynamically generated lexicon. The problem of combining the various recognizers is viewed as construction of a post-processing classifier operating on the matching scores of the recognizers. We will show how accounting for score dependencies between the lexical classes is essential for developing an optimal combination methodology.

We will also present our recent extension of the postal work to loosely constrained applications such as indexing and searching hand-filled medical forms for the NY Department of Health. We take a topic categorization approach based on latent semantic modeling to reduce the lexicon and boost the recognition accuracy to levels adequate for search engines.

We will conclude with an overview of our current handwriting research in processing historical manuscripts, developing handwritten CAPTCHAs, and multilingual OCRs.