

## Distinguished Lecture Series 2009

# From Text to Media: A Unified Approach to Multimedia Pattern Recognition

by

## Professor Chin-Hui LEE

Professor

School of Electrical and Computer Engineering  
Georgia Institute of Technology, USA



**Date:** 17 March 2009, Tuesday

**Time:** 4:30 p.m. – 6:00 p.m.

**Venue:** Room 222, 2/F, Ho Sin Hang Engineering Building, CUHK

### Abstracts

With an increasing amount of audio and video materials made available on the web, information extraction from multimedia documents is becoming a key area of growing business and technology interest. Research opportunities range from traditional topics, such as multimedia signal representation, processing, coding, modeling, authentication, and recognition, to emerging subjects, such as language modeling, semantic concept decoding, media data mining, and knowledge discovery. Conventional multimedia processing often focuses on techniques developed for an individual medium. However for multimedia pattern recognition purposes, a number of algorithms are well-positioned and applicable to many cross-media applications.

We present three families of such algorithms. The first, derived from speech and image coding, is *unsupervised tokenization of multimedia patterns* into a finite set of alphabets through segment or block quantization. Acoustic and visual lexicons can then be constructed. The second, derived from information retrieval, is a *vector space representation of multimedia documents* via extraction of high-dimensional salient feature vectors using co-occurrences statistics of acoustic and visual words. This can be accomplished through a feature extraction and feature reduction framework, known as latent semantic analysis (LSA), serving as a unified representation of multimedia patterns. This allows us to convert heterogeneous multimedia patterns into uniform text-like documents. Finally we discuss *decision-feedback discriminative learning*, derived from automatic speech and speaker recognition, for document classification, such as text categorization (TC) or topic identification. Machine learning techniques have been extensively used in the TC community to design discriminative classifiers. We present a recently developed maximal figure-of-merit (MFoM) learning framework for TC. It attempts to optimize parameters for any classifier with any feature representation on any desired performance metric, and was shown to outperform other well-known machine learning algorithms, such as support vector machine (SVM), especially for topics with only very few training documents.

The mathematical formulation of the above three sets of techniques will be described in detail first, followed by their applications to text categorization, automatic image annotation, video story segmentation, audio fingerprinting, and automatic language identification. The three frameworks, all derived from the speech and language processing community, provide a natural linkage to language characterization and concept modeling of multimedia documents and seem to serve as an ideal combination of tools for bridging the gap from conventional, low-level, content-based signal processing to high-level, concept-based processing of multimedia patterns.

### Biography of the Speaker

**Chin-Hui Lee** is a professor at School of Electrical and Computer Engineering, Georgia Institute of Technology. He received the B.S. degree in Electrical Engineering from National Taiwan University in 1973, the M.S. degree in Engineering and Applied Science from Yale University in 1977, and the Ph.D. degree in Electrical Engineering with a minor in Statistics from University of Washington, Seattle, in 1981.

Prof. Lee started his professional career at Verbex Corporation and was involved in research on connected word recognition. In 1984, he became affiliated with Digital Sound Corporation, Santa Barbara, where he engaged in speech research and product development of the DSC-2000 Voice Server. Between 1986 and 2001, he was with Bell Laboratories, Murray Hill, New Jersey, where he became a Distinguished Member of Technical Staff and the Director of the Dialogue Systems Research Department. His research interests include multimedia communication, multimedia signal and information processing, speech and speaker recognition, speech and language modeling, spoken dialogue processing, adaptive and discriminative learning, biometric authentication, and information retrieval. From 2001 to 2002 he was a visiting professor at School of Computing, The National University of Singapore. In September 2002, he joined the Georgia Institute of Technology.

Prof. Lee has participated actively in professional societies. In 1991-1995, he was an associate editor for IEEE Transactions on Signal Processing and IEEE Transactions on Speech and Audio Processing. During the same period, he served as a member of the ARPA Spoken Language Coordination Committee. In 1995-1998 he was a member of the Speech Processing Technical Committee and became the chairman in 1997-1998. In 1996, he helped promoting the SPS Multimedia Signal Processing Technical Committee in which he is a founding member. Prof. Lee is a Fellow of the IEEE, and has published more than 300 papers and 25 patents on the subject of automatic speech and speaker recognition. He received the SPS Senior Award in 1994 and the SPS Best Paper Award in 1997 and 1999, respectively. In 1997, he was awarded the prestigious Bell Labs President's Gold Award for his contributions to the Lucent Speech Processing Solutions product. Prof. Lee often gives lectures to a wide international audience. In 2000, he was named one of the six Distinguished Lecturers by the IEEE Signal Processing Society. He was also named one of the two ISCA's inaugural Distinguished Lecturers in 2007-2008. Recently he won the SPS's 2006 Technical Achievement Award for "Exceptional Contributions to the Field of Automatic Speech Recognition".

\*\*\*\*\*

**ALL ARE WELCOME**

\*\*\*\*\*

**For ENQUIRIES: Tel: 3163 4351**

\* Light refreshment will be served at 4:15 p.m. before the lecture \*