

Yuan Hao

ADDRESS:

1054 spruce street apt 89, Riverside, CA, 92507

Email: yhao@cs.ucr.edu

Cell phone: 951-880-8601

RESEARCH INTEREST

Data Mining/Machine Learning, Pattern Recognition and Information Retrieval techniques with applications to big data analytics (time series data, video, image and audio analytics).

TECHNICAL SKILLS

Computational skills: C/C++, MATLAB/Maple, MapReduce, Python, BASH, GDB Debugger, Visual Studio and Eclipse for IDE, SVN for version control, SimpleScalar, IC STATION, L-EDIT, AUTOCAD

Platforms: Windows, Unix/Linux, VMWare Workstation

EXPERIENCE

- **Software Engineer (Intern), Mountain View, Google, June 2013 to September 2013**

I implement PhotoDNA image hashing algorithm into Ocelot to detect child pornography uploaded in any Google product with hashes shared by the industry. I worked with different teams, including legal team, to design/implement a pipeline to evaluate coverage of different image hashing techniques.

- **Data Scientist (Intern), Mountain View, LinkedIn, September 2013 to Present**

I work in Content & Community (CNC) Decision Sciences team, which responsible for building the social component of LinkedIn that enables our users to create, share, discuss and distribute highly relevant content across the LinkedIn network.

EDUCATION

University of California, Riverside **Ph.D. Candidate** Sep 2009 - Present
Major: Computer Science **Advisor:** Dr. Eamonn Keogh **GPA:** 3.9/4.00

Rochester Institute of Technology (RIT) **Master of Engineering** Sep 2008 – June 2009
Major: Electrical Engineering **Advisor:** Dr. Karl D. Hirschman **GPA:** 3.68

Jilin University (JLU) **Bachelor of Science** Sep 2004 - June 2008
Major: Electrical and Microelectronics Engineering

PUBLICATIONS

- **Yuan Hao**, Yanping Chen, Jesin Zakaria, Bing Hu, Thanawin Rakthanmanon and Eamonn Keogh. Towards Never-Ending Learning from Time Series Streams. *SIGKDD'13: ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2013.*

- **Yuan Hao**, Mohammad Shokoohi-Yekta, George Papageorgiou, Eamonn Keogh. Parameter-Free Audio Motif Discovery in Large Data Archives. *ICDM'13: IEEE International Conference on Data Mining, 2013*
- **Yuan Hao**, Bilson Campana, Eamonn Keogh. Monitoring and Mining Insect Sounds in Visual Space. *SDM'12: SIAM International Conference on Data Mining, 2012.*
- **Yuan Hao**, Bilson Campana, Eamonn Keogh. Monitoring and Mining Animal Sounds in Visual Space. *Journal of Insect Behavior, 2012.*
- **Yuan Hao**, Lexiang Ye, Eamonn Keogh. Anyspace Indexing Algorithms with Application to Sensor Data Mining. (Under submission to Journal of Pattern Recognition)
- Bing Hu, Thanawin Rakthanmanon, **Yuan Hao**, Scott Evans, Stefano Lonardi, Eamonn Keogh. Discovering the intrinsic Cardinality and Dimensionality of Time Series using MDL. *ICDM'11: IEEE International Conference on Data Mining, 2011.*
- Bing Hu, Thanawin Rakthanmanon, **Yuan Hao**, Scott Evans, Stefano Lonardi, and Eamonn Keogh. Using the Minimum Description Length to Discover the Intrinsic Cardinality and Dimensionality of Time Series. *Journal of Data Mining and Knowledge Discovery (DMKD), 2013.*
- Gustavo Batista, **Yuan Hao**, Eamonn Keogh, Agenor Mafra-Neto. Towards Automatic Classification on Flying Insects Using Inexpensive Sensors. *ICMLA'11: IEEE International Conference on Machine Learning and Applications, 2011.*
- Yanping Chen, **Yuan Hao**, Thanawin Rakthanmanon, Jesin Zakaria, Eamonn Keogh, A General Framework for Never-Ending Learning from Time Series Streams. (Under Submission)
- Liudmila Ulanova, **Yuan Hao**, Eamonn Keogh. Generating Synthetic Data to Allow Learning from Single Exemplars. (Under submission)

ACADEMIC AND RESEARCH EXPERIENCE

- Grant writing for a funded Air Force SBIR on Desert Fauna Detecting and Tracking System. (AF 112-193 2011). This proposal was based on my work (SDM 2012) and beat out 23 competitors for funding.
- **Never-Ending Online Learning from Time Series Streams**
 - We proposed a never-ending learning framework for time series in which an agent examines an unbounded stream of data and occasionally asks a teacher for a label.
 - We demonstrated the utility of our ideas with experiments that consider real world problems in domains as diverse as medicine, entomology, wildlife monitoring, and human behavior analyses, etc.
- **Parameter-Free Audio Motif Discovery in Large Data Archives**
 - We introduced a novel technique for finding audio motifs. Our method does not require any domain specific tuning, is essentially parameter-free. We demonstrate our algorithm on diverse domains, finding audio motifs in laboratory mice vocalizations, wild animal sounds, music, and human speech.
 - Implemented a general framework for discovering audio motif with a probability early abandoning algorithm using C++ and Matlab. Explored standard MPEG-1 scheme, which implemented in C++.

- **Monitoring and Mining Animal Sound in Visual Space**
 - We introduced a novel bioacoustic recognition/classification framework that mitigates or solves issues, such as careful tuning many parameters, computational expensive for deployment in resource-limited sensors, specialized for a very small group of species, etc.
 - We proposed to classify animal sounds in the visual space, by treating the texture of their spectrograms as an acoustic fingerprint using a recently introduced parameter-free texture measure as a distance measure implemented in C++, Matlab.
- **Anyspace Indexing Algorithms with Application to Sensor Data Mining**
 - In this work we have shown that the Orchard's algorithm may be rescued from its relative obscurity by considering it as an anyspace algorithm and leveraging off of its unique properties to produce efficient sensor mining algorithms using C++ and Matlab.
- **MDL-SAX: Discovering the intrinsic cardinality and dimensionality of time series**
 - Proposed using MDL Model to find the intrinsic cardinality and dimensionality of time series to improve the work of Symbolic Aggregate approxImation (SAX).
 - Considering the combination of the complexities of the compressed data and the encoding to remove user defined parameters to prevent poor choice affects the efficiency of the algorithm.
 - Build a time series classifier into a small memory limited device such as a pacemaker.
- **Interactive and Intelligent Searching of Biological Images**
 - Build 3D model using multi-focus image processing to extract useful features and to find the most similar species for the given query biological image using Matlab.
- **UCR Time Series Classification/Clustering**
 - Work on different datasets for analysis of a Dynamic Time Warping algorithm, which can improve classification accuracy, and also the general Euclidean distance measure for time series. www.cs.ucr.edu/~eamonn/time_series_data/

ACADEMIC HONORS & SCHOLARSHIPS

- **Dean's Distinguished Fellowship Award**, University of California, Riverside, 09-10
- **RIT Graduate Scholarship & Assistantships**, Rochester Institute of Technology, 08-09
- **First Prize**, China Undergraduate Mathematical Contest in Modeling (CUMCM), 07
- **Second Prize** Undergraduate Research Training Project, Jilin University Undergraduate Research Opportunity Program, China, 07
- **Successful Participant** Mathematical Contest in Modeling Certificate of Achievement, USA(07')
- **Third Prize**, National English Contest for College Students (level A) China, 04, 05 (Top5%)
- **National scholarship**, China, 04-05 (Top 1%)
- **CHINA NATIONAL PETROLEUM CORPORATION-Scholarship**, China 05 (Top 3%)
- **Highest Honor Scholarship**, Jilin University, China, 05-06, 06-07
- **Outstanding Graduation with highest honor**, Jilin University, China, 08.