Curriculum Vitae - Po-Jen Hsu / 許伯任

Name: Po-Jen Hsu / 許伯任

Email: clusterga@gmail.com

Tel: +886-952-335534

(Click here to see the on-line version)

EDUCATION

2014: Ph.D., Molecular Science and Technology, TIGP, Academia Sinica/Physics, National Central University. Supervised by Dr. Siew Ann Cheong, Dr. Arnaldo Rapallo, and Dr. San-Kiong Lai (primary thesis supervisor) [1-5].

2003: M.S. in Physics, National Central University [9-11]

2000: B.S. in Physics, National Central University

RESEARCH INTERESTS

Machine Learning

- Time series clustering [1, 6]
- Pattern recognition [1, 2]
- Open data mining

Pattern recognition and time series clustering techniques were applied to the mechanistic study of protein folding and unfolding [1]. Using pattern recognition, the shape of the composing residues of a protein can be converted into multiple time series functions of shape similarity. Therefore, one can perform time series clustering and segmentation analysis to study the correlation between residues, in order to understand the underlying mechanism of amyloid formation that are believed to be associated with brain disorders such as Alzheimer's disease, Parkinson's disease, and Bovine Spongiform Encephalitis. In future research this method will be used to scrutinize signatures from Open Data such as medical data, earthquakes, weather, traffic, economy, and particle physics data (CERN Open Data).

Pattern recognition is also popular in computer vision. To achieve precise and real-time hand gesture control, a histogram-based method with graphics processing units (GPU) acceleration technique was developed (GestureCV). The goal is to understand how intelligence recognizes objects and makes decisions. More statistical and dynamical methods will be integrated into this system for more general recognition.

High-speed Computation

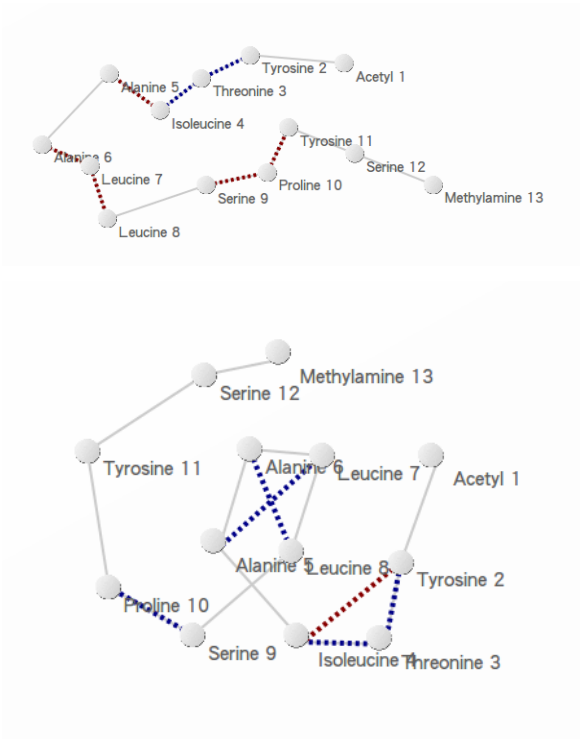
- General-purpose computing on graphics processing units (GPGPU) [2, 4, 5, 7, 8]
- PC cluster [1-11]
- Cloud computing

Complexity and veracity are two major issues in Big Data processing. To facilitate this challenging task, one seeks for any possible solution from conventional PC clustering to modern cloud computing. For example, a time series analysis can be accelerated by more than 10 times by utilizing PC cluster computation (PTMD).

Moreover, the GPGPU technique can dramatically boost the performance to more than 100 times (CL-VAF). Ultimately it is hoped to adopt cloud computing techniques such as Hadoop and a distributed file system to extend our scope and ability for Big Data analysis.

Data Visualization

Data visualization aims to communicate information clearly, efficiently, and interactively to common users using effective visual diagrams. Currently, I'm using NumPy and Matplotlib as the tools for data representation. In the future, web-based visualization technique such as HTML5 and JavaScript (ex: D3.js) will be implemented.



Visualization of strong (red dashed) and weak (blue dashed) correlations of the time series clustering results [1].

(Click here to see the on-line demonstration)

OPEN SOURCE PROJECTS

I'm interested in building tool chains from statistical sampling to model simulation. I enjoy programming very much and have spent a lot of time in software engineering. Smart phones and tablets are currently equipped with powerful CPUs and GPUs, but most of them are only used for gaming and web browsing. However, I saw the possibility of high-speed computation on these devices. This year I joined the IT industry to work on GPGPU techniques for mobile devices and embedded systems. I hope that one day these devices can become new computing resources to benefit science and education.

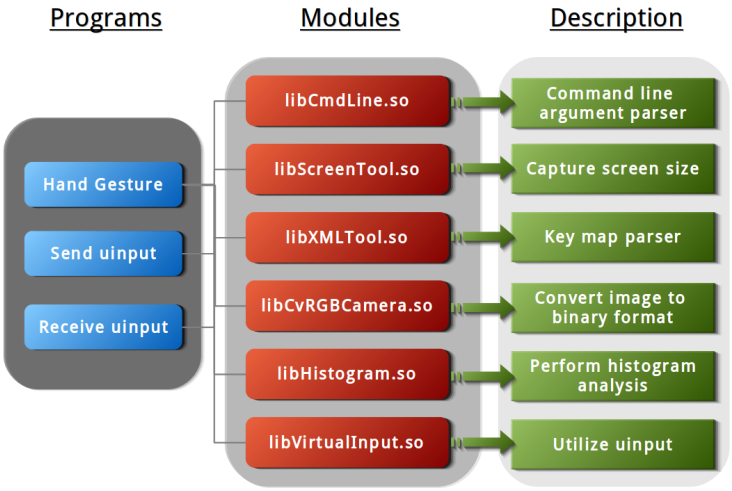
All my programs were published by Open Source licenses, which are listed as follows:

(Click on each subtitle to view the source code repository on GitHub)

GestureCV

Hand gesture control based on histogram analysis (C++/OpenCL/OpenCV)

GestureCV combines image filtering and histogram analysis to accomplish precise real-time hand gesture control on laptops or embedded systems. It is a machine learning prototype for computer vision.



Framework of GestureCV.

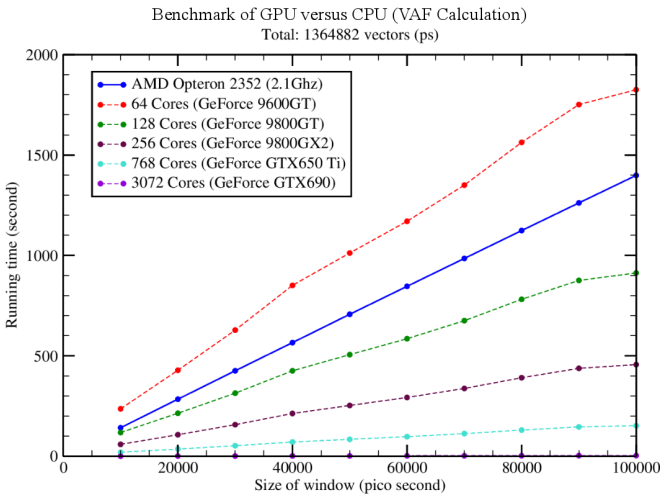
Application switcher and presentation software control.

(Click here to view the demonstration)

CL-VAF

Vector Autocorrelation Function with GPGPU (C++/OpenCL)

CL-VAF [3, 4, 5, 7, 8] implements GPGPU techniques to calculate the autocorrelation function of multi-dimensional vectors.



Performance of CL-VAF.

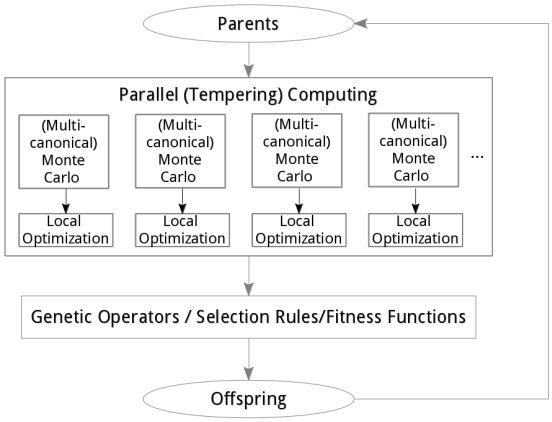
MPI-Tool

MPICH utility for PC cluster (Shell Script)

MPI-Tool allows users to deploy and monitor jobs created by MPICH on PC clusters. It is written in shell script language.

PTMBHGA

Parallel Tempering Multicanonical Basin-hopping Plus Genetic Algorithm (Fortran/MPICH)

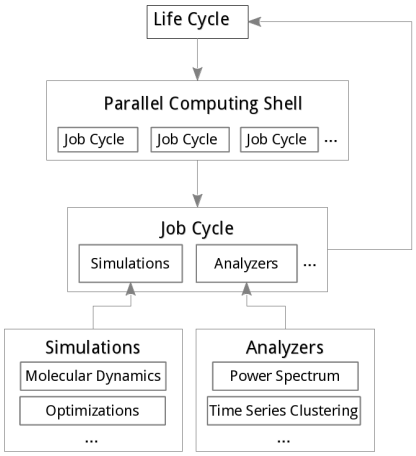


Framework of PTMBHGA.

PTMBHGA [2-5, 9-11] combines several state-of-art optimization techniques such as the genetic algorithm, parallel tempering Monte Carlo method, simulated annealing, basing-hopping method, and the multicanonical Monte Carlo method. It is flexible and reliable for searching global optimized results. This program has been used by research groups in Japan and Malaysia.

PTMD

Parallel Tempering Molecular Dynamics Simulation (Fortran/MPICH)



Framework of PTMD.

PTMD [2, 4-8] is designed for model simulation and several statistical analyses including moments, Fourier transformation, and nearest neighbor analysis. It integrates a task schedule system so that users can perform

multiple simulations and analyses in parallel.

D-Tool

A utility for Mode-coupling Diffusion Theory (C/C++)

The D-tool [3] reduces tedious preparation for diffusion theory calculation. The research was published with Dr. Arnaldo Rapallo from ISMAC, Italy.

Matlab and Python codes

- Time Series Clustering Method [1, 6]
- Time Series Segmentation Method [1]
- Sliding Window Method [1, 6]
- Pattern Recognition Method [1, 2]
- Power Spectrum Density Calculation [4, 5, 7, 8]
- Nearest Neighbor Analysis [6]
- Auto-correlation Calculation [3-5, 7, 8]

Others

Chinese translations of PhET education project in Physics (EzGo, OSSACC, Ministry of Education)

- Davission-Germer Experiment
- Stern-Gerlach Experiment
- Quantum Wave Interference
- Quantum Tunneling
- Quantum Bound States
- Covalent Bonds
- Band Structure

WORKING EXPERIENCE

2014 - Present: Senior Engineer, Innovation Digital System, System Software Development Division, Hon Hai Precision IND. CO., LTD. --Responsible for Image recognition algorithms, hand gesture control algorithms, HTML5 acceleration (WebCL), GPGPU applications and development of Android input systems.

2005 - 2008: Research assistant in Physics department, National Central University. --Built and maintained PC clusters and Linux web/mail servers. Built algorithms and models from scratch to simulate chemical and biological materials under statistical mechanics circumstances. [6-8]

2003 - 2005: Military service

SPECIALTIES

- Applied Statistics, Machine Learning, and Mathematical Modeling

    1. Time Series Analysis
    2. Genetic Algorithm
    3. Monte Carlo Method
    4. Optimization methods
- Theoretical and Computational Physics
    1. Statistical Physics
    2. Strongly-correlated Systems
    3. Long-time Dynamics
    4. Molecular Dynamics
    5. Molecular Biology
    6. Condensed Matter Physics
    7. Complex Systems
- Reduced Statistical Methods and Diffusion Theory
    1. Mode-coupling Approaches for long-time behaviors
    2. Generalized Diffusion Equation (GDE)
    3. Optimized Rouse-Zimm Local Dynamics (ORZLD)
- Parallel Computing Techniques
    1. Open Computing Language (OpenCL)
    2. PC Cluster Computing using Open Message Passing Interface (OpenMPI/MPICH)
- Web-based Technologies
    1. HTML5 and CSS
    2. Data Visualization using JavaScript (D3.js)
    3. XML (libxml2)
- Computer Vision, Hand Gesture Control, and Image Recognition
    1. Open Source Computer Vision (OpenCV)
    2. Linux/Android Camera and Input Subsystems
    3. Arduino and Embedded System programming
- Open Source Projects for science and education
- Statistical Mechanics, Quantum Mechanics, Chemical Physics, and Applied Mathematics,
- Software Management and Engineering
    1. Git
    2. SVN
    3. Unified Modeling Language (UML)
- Algorithms and programming language teaching
    1. C++ and C
    2. Python, NumPy, and Matplotlib

3. Regular Expression
4. Matlab
5. GNU Scientific Library (GSL)
6. Linux System Programming (IPC, thread, socket...)
7. Linux Shell Script
8. Fortran
9. Latex

- System Administrator with experience in Linux

1. PC Clusters
2. Web Servers
3. Mail Servers

AWARD

- Best Team Presentation Award, 4th Hope Meeting, 2012, Japan.

INVITED TALK

“ Open Source in Physics", International Conference on Open Source 2009, Taiwan (download slides)

PUBLICATIONS

(Download the paper by clicking the title)

1. Precursory Signatures of Protein Folding/Unfolding: From Time Series Correlation Analysis to Atomistic Mechanisms, P. J. Hsu, S. A. Cheong, and S. K. Lai, J. Chem. Phys. 140, 204905 (2014).

2. A new perspective of shape recognition to discover the phase transition of finite-size clusters, P. J. Hsu, J. Comput. Chem. 35, 1082 (2014).

3. Peptide dynamics by molecular dynamics and diffusion theory methods with improved basis sets, P. J. Hsu, S. K. Lai, and A. Rapallo, J. Chem. Phys. 140, 104910 (2014).

4. Melting behavior of Ag14 cluster: An order parameter by instantaneous normal modes, P. H. Tang, T. M. Wu, P. J. Hsu, and S. K. Lai, J. Chem. Phys. 137, 244304 (2012).

5. Comparative study of clusterAg17Cu2by instantaneous normal mode analysis and by isothermal Brownian-type molecular dynamics simulation, P. H. Tang, T. M. Wu, T. W. Yen, S. K. Lai, and P. J. Hsu, J. Chem. Phys. 135, 094302 (2011).

6. Dynamical study of metallic clusters using the statistical method of time series clustering, S. K. Lai, Y. T. Lin, P. J. Hsu, and S. A. Cheong, Compt. Phys. Commun. 182,1013(2011).

7. Melting behavior of noble-metal-based bimetallic clusters, T. W. Yen, P. J. Hsu, and S. K. Lai, e-J. Surf. Sci. Nanotech.7, 149-156 (2009).

8. Melting scenario in metallic clusters, P. J. Hsu, J. S. Luo, S. K. Lai, J. F. Wax, and J-L Bretonnet, J. Chem. Phys.129, 194302 (2008).

9. Structure of bimetallic clusters, P. J. Hsu and S. K. Lai, J. Chem. Phys.124, 044711 (2006).

10. Multi-canonical basin-hopping: a new global optimization method for complex systems, L. Zhan, B. Piwowar, W. K. Liu, P. J. Hsu, S. K. Lai, and Jeff Z. Y. Chen, J. Chem. Phys.120, 5536 (2004).

11. Structures of metallic clusters: mono- and polyvalent metals, S. K. Lai, P. J. Hsu, K. L. Wu, W. K. Liu, and M. Iwamatsu, J. Chem. Phys.117, 10715 (2002).

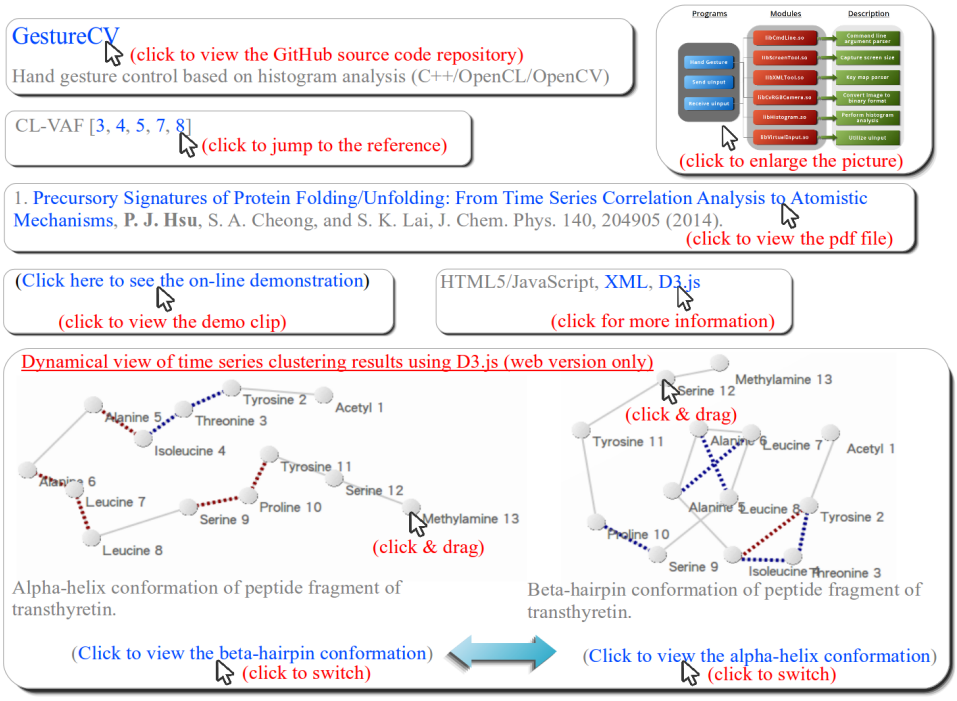
OTHER INFORMATION

- PhD certificate: link 1 and link 2.
- PhD transcripts: link.

RECOMMENDED BY

- 鄭王曜 教授(中央大學物理系，Email: wycheng@ncu.edu.tw /電話:(03)4227151#65337)
- 陳永富 教授(中央大學物理系， Email: yfuchen@cc.ncu.edu.tw /電話:(03)4227151#65375)
- Dr. Arnaldo Rapallo (ISMAC, Italy. Email: rapallo@ismac.cnr.it)

ABOUT THIS DOCUMENT



Click on the links for more information and dynamical demonstrations