LUIS PEDRO COELHO
Curriculum Vitæ

Instituto de Medicina Molecular
Av. Prof. Egas Moniz
1649-028 Lisboa (Portugal)
phone : (+351)-917-814-123
email : luis@luispedro.org
Citizenship : U.K. and Portugal

Education

2011 **PhD in computational biology, Carnegie Mellon University**

Dissertation topic: *Modeling the Space of Subcellular Location Patterns Using Images and Other Sources of Information*, advised by Prof. Robert F. Murphy.

2006 **MS in computer science, Instituto Superior Técnico (Technical University Lisbon)**

Dissertation topic: *Bayesian Network Parameter Estimation Using Noisy Observations or Soft Evidence*, advised by Prof. Arlindo Oliveira.

2004 **BS in computer science, Instituto Superior Técnico (Technical University Lisbon)**

Finished top of my class.

Research Highlights

2011 **Proteome-wide Study of Subcellular Location in Mouse Fibroblasts**

I am developing new methods for learning from both a very large collection of images of fluorescently tagged proteins and other forms of data (sequence information, database information). This work builds on my past work. A manuscript describing it is in preparation.

2010 **Unsupervised Subcellular Pattern Unmixing**

Many proteins are located in more than one organelle simultaneously, a phenomenon known as a mixed pattern (the pattern corresponding to a single organelle being a pure pattern). Recovering the organelle structure from image data where mixed patterns are present is known as the unsupervised unmixing problem in subcellular location. We developed new models for solving this problem. Our solution was chosen as one of the highlights of 2010 in computational biology by the journal *Nature Biotechnology*.[1]

2008–2009 **Structured Literature Image Finder (SLIF)**

SLIF is a search engine that indexes biomedical papers with their images. In this project, I was responsible for the computer vision aspects of the processing pipeline. I also handled much of the integration effort of the several components and coordinated the preparation of the multi-author publications that resulted from the effort. This project was a finalist in the Elsevier Grand Challenge (4 teams out of 70 were chosen for the final). I represented our team in both the semi-final and the final of this competition.

[1] *Trends in computational biology–2010* by H. Craig Mak, Nature Biotechnology, vol. 29(1), Jan. 2011, pp. 45–45 [DOI]

Teaching Experience

2011 Practical Tutorial in Lisbon Machine Learning Summer School

2009 Programming for Scientists

I designed and taught a semester-long course on computer programming for scientists at Carnegie Mellon University. This course was not based on any existing course at CMU, but resulted from a need I identified in biomedical or physical scientists who, as part of their job, have a need to program.

2008 Teaching Assistant for Laboratory Methods for Computational Biologists (CMU)

2005 Introduction to Computers Course in Cacém

I co-developed and co-taught an introductory course in computer usage in Cacém, an underprivileged neighborhood near Lisbon. This was pro-bono work.

2005 Teaching Assistant for Decision Support Systems (IST)

Other Research Projects

2006 Parameter Estimation in Bayesian Networks from Noisy Data

A Bayesian network is a fundamental class of graphical probabilistic models and learning parameters from data is a fundamental problem. In this work, I investigated the theoretical limits of learning from noisy data.

2005 Approximate String Indexing

An approximate string index is a data structure that allows for fast string queries where the matches may be approximate. I developed one such index, the *dotted suffix tree*.

Scholarships & Awards

2012 Siebel Scholar

Awarded annually for academic excellence and demonstrated leadership to 85 top students from the world's leading graduate schools

2007–2011 PhD. Scholarship from Portuguese Science Foundation

2009 Joint CMU-U. of Pittsburgh PhD. in Computational Biology Research Excellence Award

2008 Joint CMU-U. of Pittsburgh PhD. in Computational Biology Academic Excellence Award

2006 Fulbright Fellow

2005 Scholarship from Portuguese Science Foundation

2004 Second Prize in *Lisboa à Letra* short story competition

2001 Instituto Superior Técnico (IST) Academic Excellence Award

Organisational & Mentoring Experience

2010–2012 **Beira Project & Rabbit Bounce**

With Rita Reis, I started the Beira Project. We spent two months as volunteers in Beira (Mozambique) working with organisations that provide public health information and services, mostly relating to HIV. We had also held organised successful fund-raising activities for our partner organisations in Mozambique. In 2012, we started a non-profit association called *Rabbit Bounce*, which will raise money for education in Mozambique.

2008–2011 **Mentoring Junior Members of Murphy Lab**

During my PhD. studies, I had the opportunity to directly supervise and mentor several junior members of the Murphy Lab. This experience includes working with **paid undergraduate programmers** (Nathan Herzing and Jephthah Liddie), one **MSc. student** (Shannon Quinn), **undergraduate students** performing lab work for credit (Jimmy Xu), and **high school students** volunteering over the summer (Peter Webb and Robert Webb).

2010 **Local Committee for Portuguese-American Postgraduate Society National Forum**

I headed the local organising committee for the 2010 edition of this annual event. It took place in Pittsburgh and included, as speakers, cabinet-level Portuguese government officials, renowned researchers, artists, as well as participants from all around the US.

2002–2004 **Producer for IST Theatre Group**

I served as the producer for the IST Theatre Group, which is one of the top university theatre groups in Portugal. We participated in several festivals, including international festivals. As producer, my activities included fund-raising and management.

Other Experience

Open Source Software

I have released several important open source packages related to my research work, namely in the areas of **machine learning**, which is addressed by the packages *milk* and *elgreco*; and **computer vision**, which is addressed by the package *mahotas*. Previously, I was a KDE developer. The KDE project is one of the largest open-source projects in the world with several hundred developers and over one million lines of code.

Skills

Computer Programming

I am an experienced programmer in several languages including **C**, **C++**, **Python**, **Matlab**, and **Haskell**.

Cell Culture & Fluorescent Imaging

As part of my doctoral research, I was responsible for **fluorescent imaging** of **cultured cells**.

Language Skills

I am bilingual in **English** and **Portuguese**. I speak and write fluent **German** (I attended a German high-school and the Technical University of Vienna) and speak fluent **French** (I have lived in France). I have a basic understanding of **Luxembourgish**.

Full List of Publications

Peer-Reviewed Journal Articles

1. **Luis Pedro Coelho**, Tao Peng, and Robert F. Murphy, *Quantifying the distribution of probes between subcellular locations using unsupervised pattern unmixing* in Bioinformatics, vol. 26 (12), pp. i7–i12, 2010 [DOI]

2. **Luis Pedro Coelho**, Amr Ahmed, Andrew Arnold, Joshua Kangas, Abdul-Saboor Sheikh, Eric P. Xing, William W. Cohen, and Robert F. Murphy, *Structured Literature Image Finder: Extracting Information from Text and Images in Biomedical Literature* in Lecture Notes in Bioinformatics, vol. 6004, pp. 23–32, 2010 [DOI]

3. Amr Ahmed, Andrew Arnold, **Luis Pedro Coelho**, Joshua Kangas, Abdul-Saboor Sheikk, Eric P. Xing, William W. Cohen, *Structured Literature Image Finder: Parsing Text and Figures in Biomedical Literature* in Web Semantics: Science, Services and Agents on the World Wide Web, vol. 8, pp. 151–154, 2010 [DOI]

Review Articles

1. **Luis Pedro Coelho**, Estelle Glory-Afshar, Joshua Kangas, Shannon Quinn, Aabid Shariff, and Robert F. Murphy; *Principles of Bioimage Informatics: Focus on machine learning of cell patterns* in Linking Literature, Information, and Knowledge for Biology Lecture Notes in Computer Science, vol. 6004, pp. 8–18, 2010 [DOI]

2. Aabid Shariff, Joshua Kangas, **Luis Pedro Coelho**, Shannon Quinn, and Robert F. Murphy; *Automated Image Analysis for High Content Screening and Analysis* in Journal Biomolecular Screening, August 2010, pp. 726–734 [DOI]

Peer-Reviewed Conference Papers

1. Amina Chebira, **Luis Pedro Coelho**, Aliaksei Sandryhaila, Stephen Lin, William G. Jenkinson, Jeremiah MacSleyne, Christopher Hoffman, Philipp Cuadra, Charles Jackson, Markus Püschel, Jelena Kovacevick; *An Adaptive Multiresolution Approach to Fingerprint Recognition* in Proceedings of IEEE International Conference on Image Processing, 2007 [DOI]

2. **Luis Pedro Coelho** and Robert Murphy; *Identifying Subcellular Locations from Images of Unknown Resolution* in Bioinformatics Research and Development Communications in Computer and Information Science, vol. 13, pp. 235–242, 2008 [DOI]

3. **Luis Pedro Coelho** and Robert F. Murphy; *Unsupervised Unmixing of Subcellular Location Patterns*, Proceedings of ICML–UAI–COLT 2009 Workshop on Automated Interpretation and Modeling of Cell Images, 2009

4. Amr Ahmed, Andrew Arnold, **Luis Pedro Coelho**, Joshua Kangas, Abdul-Saboor Sheikk, Eric P. Xing, William W. Cohen, and Robert F. Murphy; *Structured Literature Image Finder*, Proceedings of BioLINK (ISMB Special Interest Group), 2009

5. **Luis Pedro Coelho**, Aabid Shariff, and Robert F. Murphy; *Nuclear segmentation in microscope cell images: A hand-segmented dataset and comparison of algorithms* in Proceedings of IEEE International Symposium in Biomedical Imaging, 2009 [DOI]

6. Taraz Buck, Arvind Rao, **Luis Pedro Coelho**, Margaret Fuhrman, Jonathan W. Jarvik, Peter B. Berget, and Robert F. Murphy; *Cell Cycle Dependence of Protein Subcellular Location Inferred from Static, Asynchronous Images* in Conference Proceedings of the IEEE Engineering in Medical and Biology Society, pp. 1016–1019, 2009 [DOI]

7. **Luis Pedro Coelho** and Arlindo Oliveira; *Dotted Suffix Trees: A Structure for Approximate Text Indexing* in String Processing and Information Retrieval Lecture Notes in Computer Science, vol. 4209, pp. 329–336, 2006 [DOI]

Invited Talks

1. *Proteome-scale analysis and modeling of subcellular location*, 4th CeBiTec Symposium BioImaging, Bielefeld (Germany), 25–27 August 2009
2. *Unsupervised Mixture Pattern Unmixing*, University of Bielefeld International Graduate School of Bioinformatics and Genome Research, July 2008
3. *Studying the subcellular location space with bioimages and other data modalities*, University of Delaware, Computer and Information Sciences Department, September 2011
4. *Bioimage Informatics: Computer Vision for Biology*, EMBO Practical Course on Microscopy: from single molecules to animals, Pretoria, November 2011
5. *Learning Subcellular Location from Images and Other Sources of Information*, KDBIO (Knowledge Discovery and Bioinformatics) seminar, Lisbon, February 2012
6. *Modeling Subcellular Location from Images and Other Sources of Information*, EAO Seminar (Instituto Gulbenkian da Ciência), Oeiras, March 2012

Other Talks

1. **Luis Pedro Coelho** and Robert F. Murphy; *Determining Resolvable Subcellular Location Categories as a Function of Image Resolution*, 24th ISAC Congress, Budapest, May 2008
2. Rita Reis and **Luis Pedro Coelho**; *Using Theatre to Fight HIV/AIDS in Mozambique*, National Conference of the Association for Theatre in Higher Education, Chicago 2011