

# Omar U. Florez

204 Greaves Hall  
Logan, UT-84321, USA

## Resume

### Contact

**Name:** Omar U. Florez

**Webpage:** <http://omarflorez.info/>

**Email:** Omar.Florez@usu.edu

**Cellphone:** (435) 757-9354

### Education

2007 – present **PhD in Computer Science**, (*Utah State University, USA*),

Graduation expected: **December 2012**

Research Advisor: Curtis Dyreson (Curtis.Dyreson@usu.edu)

GPA: 3.6

2001–2007 **Bachelor in Computer Science**, (*San Agustin University, Peru*),

Placed in top 10 after graduation

GPA: 4.0

### Technical Achievements

- December 2010 **Winner of an IBM Innovation Award on Scalable Data Analytics** valued in \$20,000 that supports his PhD thesis on the *automatic rules extraction to explain vehicle interactions from traffic cameras*
- October 2011 **Three information disclosures and one patent filed** from two Summer internships with **IBM Research**. Topics: content protection for multimedia data (healthcare and engineering documents)
- October 2011 **Founder of PrimerosPuestos**, a startup that exploits a **crowdsourcing algorithm** to **automatically** discover talented undergrads in Latin America to give them opportunities to get higher education in the US. (**Semifinalist** of the **Intel Challenge 2011** and **Finalist** of **Wayra-Peru 2011** out of 1200 projects)
- October 2011 **Published** 18 conference papers, journals, and book chapters since undergrad
- October 2011 **Scholarship** to attend to the **Society of Hispanics Professional Engineers**, Anaheim, CA, USA
- August 2011 **Scholarship** to attend to the **International Conference on Very Large Data Bases (VLDB'11)**, Seattle, WA, USA
- August 2011 **Scholarship** to attend to the **Summer School** on Parallel Programming, UC Berkeley, CA, USA
- April 2011 **Scholarship** to attend to the **Doctoral Consortium** of the **2011 Richard Tapia Celebration of Diversity in Computing Conference**, San Francisco, CA, USA
- February 2011 **Scholarship** to attend to the **NSF Academic Workshop for Underrepresented Assistant & Associate Professors and Senior Doctoral Students**, Los Angeles, CA, USA
- November 2005 **Best paper** in the **Fifth South American Congress on Informatics and Systems (COISIS 2005)**
- December 2000 **Second** out of over **1500 participants** in the undergrad admission exam to the department of **Computer Science** at **San Agustin University** in Peru

### Research Interest

- **Multimedia Information Retrieval:** Indexing/similarity search of very large images and videos databases containing real human activities via Locality Sensitive Hashing
- **Machine/Statistical Learning:** Classification/clustering of words and images with a Hierarchical Dirichlet Process
- **Large Scalable Systems:** Sketching stream data via Map-Reduce techniques implemented on hash tables
- **Data Mining:** Real-time co-occurrence analysis of spatio-temporal interactions between vehicles for traffic cameras
- **Image Processing/Computer Vision:** Scene understanding via real-time tracking and event detection of moving objects for traffic road and human activity videos

## Work Experience

- May 2011 – August 2011 **Research Intern, IBM Almaden Research Center**
- Worked in the selective de-identification of **Medical** and **Aerospace/Defense** documents (PDF files, images, and videos), which contain **sensitive information** embedded as text
  - The **goal** is to provide automatic **sharing of confidential documents** between organizations while preserving **intellectual property**
  - Programmed the **editing of embedded text** with **Optical Character Recognition**
  - Programmed the **generation** of a **compact representation** for the de-identified multimedia document. This file is then transmitted as a small sequence of bytes over a network
  - Programmed a **hierarchy of access levels**. Thus, users with different compact files will de-identify different parts of the original document within a safe server
- (\*) Achievements:**
- Wrote two **Information Disclosures** based on these algorithms, a requirement to initiate the process of getting two patents
  - Implemented a demo to illustrate the **automatic translation** (from English to Spanish) of medical terms embedded in **Echocardiogram videos**
  - Collaborated with code and demo in the obtaining of a **FOAK** (First of A Kind) program for the Content Protection Group at IBM. The FOAK is valued in **\$800K**, involved the collaboration between **IBM Almaden** and **IBM Tokyo**, and last for 2 years. Target clients are *Airbus* and *Raytheon*
- Contact information:** Dulce Ponceleon, Research Manager (Dulce@almaden.ibm.com).
- July 2010 – October 2010 **Research Intern, IBM Almaden Research Center**
- Internship at **IBM Research** in California in the **Health Care** and **Content Protection** group
  - Programmed the **Statistical Learning of regions** that contain **patient's sensitive information** embedded as text within **Echocardiogram** images and videos
  - Programmed the de-identification of this data with k-anonymity **Data Mining techniques**
- (\*) Achievements:**
- Wrote two information disclosures needed to initiate the process of getting two patents
  - The **first patent** is about **privacy preserving** in echocardiogram images/videos
  - The **second patent** is about the **automatic mixing of diverse multimedia** data under **fair use copyright** principle. This patent has recently been **filed** in the US Patent and Trademark Office
- Contact information:** Dulce Ponceleon, Research Manager (Dulce@almaden.ibm.com)
- August 2007 – present **Graduate Research Assistant, Utah State University**
- Developed, made experiments, and wrote articles on Discovery of Rules to Explain Vehicle Traffic in Video data, Large Scale Clustering/Similarity Search of Unstructured data (Video, Music, Human Motion, and Text), Information Retrieval, and Parallel Computing.
- (\*) Achievements:**
- **Twelve publications** in refereed conferences, journals, and book chapters
  - One **IBM innovation award** valued in \$20,000 for our study on the **efficient detection of interesting patterns in congested roads from streams of video data**
  - **Four scholarships** to attend conferences, doctoral workshops, and Summer schools.
- (\*) My Work:**
- **Very fast retrieval of human motion in video data**  
Web: [www.omarflores.info/index.php?id=very-fast-retrieval-video](http://www.omarflores.info/index.php?id=very-fast-retrieval-video)
  - **Vehicle traffic understanding in video sequences**  
Web: [www.omarflores.info/index.php?id=traffic-understanding](http://www.omarflores.info/index.php?id=traffic-understanding)
  - **Sublinear search of time series and human motion via hashing**  
Web: [www.omarflores.info/index.php?id=retrieve-human-motions](http://www.omarflores.info/index.php?id=retrieve-human-motions)
  - **Extracting timeseries from video data to automatically categorize activities**  
Web: [www.omarflores.info/index.php?id=timeseries-video](http://www.omarflores.info/index.php?id=timeseries-video)
  - **HRG: Voronoi-based graph for efficient indexing of multi-dimensional data**  
Web: [www.omarflores.info/systems/hrg/](http://www.omarflores.info/systems/hrg/)  
Demo: [omarflores.info/systems/hrg/demo.html](http://omarflores.info/systems/hrg/demo.html)
  - **GPS and Cellphone applications: Mobile browsing of USU campus with cellphones**  
Web: [www.youtube.com/watch?v=X1omPC5RtZM](http://www.youtube.com/watch?v=X1omPC5RtZM)
  - **Voice2SQL: Automatic translation of commands from voice to SQL sentences**  
Web: [www.omarflores.info/index.php?id=voice2sql](http://www.omarflores.info/index.php?id=voice2sql)
- Contact information:** Curtis Dyreson, Academic Advisor (Curtis.Dyreson@usu.edu)
- January 2007– August 2007 **Software Engineer, zAgile**
- Software Engineer in a team with people from different countries (Argentina, Brazil, Peru, and Chile)
  - Experimented in person the problems associated to work in a distributed team: different cultures, time zones, geographical positions, and level of knowledge
  - Integrated opensource tools from a software lifecycle (design, implementation, testing, and documentation) into one general framework See <http://www.zagile.com/>

---

## Leadership Skills

- December 2010 – present
- Founder of PrimerosPuestos.org**
- This is an **startup** that gives equal opportunities to **undergrad students** in Latin America to get assistantships in American universities, internships, and mentoring.
  - We developed the **TalentScore** algorithm to let users of **Facebook** choose their most talented friends via **Crowdsourcing**
  - 10 students with **highest TalentScore** are introduced to universities and companies every week.
  - This social effort is aimed to break the networking barrier that hinders young students to go abroad, get a higher degree, and come back to play a positive role in Latin American countries.
- (\*) Achievements:**
- Semi-finalist of the **Intel Challenge 2011!**, which selects innovative startup projects around the world
  - Signed a partnership with **Fulbright-Peru** to find low-income and talented students in Peru
  - Signed a partnership with **ObiettivoLavoro** (an Italian HR company) to let them use our **TalentScore** algorithm to recruit talent in their Latin American offices.
  - This project is in the annual agenda of the **ACM Computing Alliance for Hispanic-Serving Institutions** (CAHSI)
- January 2010 – April 2011
- Vice-president in the USU chapter of the Association for Computing Machinery (ACM)**
- (\*) Achievements:**
- Organized several **programming contests** for undergrad students at USU (8 students were chosen to attend the regional/national contests and received travel funds from USU-ACM in 2010)
  - Participated in visits to the **Adams Elementary School** in Logan, Utah to introduce **children** to some important Computing concepts such as safe Internet browsing, robot programming, and game development

---

## Research Experience

A list of most relevant/recent publications is shown below. For a complete list, please visit [www.omarflores.info/index.php?id=publications](http://www.omarflores.info/index.php?id=publications).

### CONFERENCE PROCEEDINGS

- March 2012 **Omar U. Florez and Curtis Dyreson**, *Discovering Activity Interactions in a Single Pass over a Video Stream*, ACM Symposium on Applied Computing (ACM SAC 2012), Trento, Italy.
- October 2011 **Omar U. Florez**, *Similar Events do not Last the Same in Aerospace and Engineering Processes, so Neither their Rules!*, NASA Conference on Intelligent Data Understanding (CIDU 2011), Mountain View, California.
- October 2011 **Omar U. Florez and Curtis Dyreson**, *Scalable Similarity Search of Timeseries with Variable Dimensionality*, 20<sup>th</sup> ACM Conference on Information and Knowledge Management (CIKM 2011), Scotland, UK.
- October 2011 **Curtis Dyreson and Omar U. Florez**, *Building a Display of Missing Information in a Data Sieve*, ACM 14<sup>th</sup> International Workshop On Data Warehousing and OLAP (DOLAP 2011) at CIKM 2011, Scotland, UK.
- October 2010 **Omar U. Florez and Curtis Dyreson**, *Mining Rules to Explain Activities in Videos*, 19<sup>th</sup> ACM Conference on Information and Knowledge Management (CIKM 2010), Toronto, Canada.  
Acceptance rate: 17.9%
- October 2010 **Curtis Dyreson and Omar U. Florez**, *Data Aspects in a Relational Database*, 19<sup>th</sup> ACM Conference on Information and Knowledge Management (CIKM 2010), Toronto, Canada.  
Acceptance rate: 17.9%
- March 2010 **Omar U. Florez and Curtis Dyreson**, *Sublinear Similarity Search of Realistic Timeseries and its Application to Human Motion*, 11<sup>th</sup> ACM SIGMM International Conference on Multimedia Information Retrieval (MIR 2010), Pennsylvania, USA.  
Acceptance rate: 16%
- April 2009 **Omar U. Florez, Xiaojun Qi, and Alexander Ocsa**, *MOBHRG: Fast K-Nearest Neighbor Search by Overlap Reduction of Hyperspherical Regions*, 34<sup>th</sup> International ACM Conference on Acoustics, Speech, and Signal Processing (ICASSP 2009), Taipei, Taiwan.  
Acceptance rate: 25%
- March 2009 **Omar U. Florez and SeungJin Lim**, *Discovery of Interpretable Time Series in Video Data Through Distribution of Spatiotemporal Gradients*, 24<sup>th</sup> Annual ACM Symposium on Applied Computing (ACM SAC 2009), Hawaii, USA.  
Acceptance rate: 32%

September 2008 **Omar U. Florez and SeungJin Lim**, *HRG: A Graph Structure for Fast Similarity Search in Metric Spaces*, 19<sup>th</sup> International Conference on Database and Expert Systems Applications (DEXA'08), Turin, Italy.  
Acceptance rate: 36%

### JOURNALS

November 2008 **Omar U. Florez and SeungJin Lim**, *A Spoken Natural Language-Based Interface for Querying SQL Databases*, In International Journal of Information Technology and Intelligent Computing

### BOOK CHAPTERS

**Omar U. Florez and SeungJin Lim**, *Modeling Query Events in Spoken Natural Language for Human-Database Interaction*, In Human Computer Interaction: New Developments, ISBN 978-3-902613-38-7. InTech Education and Publishing. Vienna, Austria.

---

## Teaching Experience

Spring term - 2010 **Teacher Assistant for Advanced Databases (CS 6800)**, With Dr. Curtis Dyreson at USU.  
Fall term - 2009 **Teacher Assistant for Database Systems (CS 5800)**, With Dr. Curtis Dyreson at USU.  
Spring term - 2008 **Teacher Assistant for Grad Seminar (CS 7900)**, With Dr. SeungJin Lim at USU.

---

## Invited Talks

June 2011 **Content Protection for Medical Videos and Background**, IBM Almaden Research Center, San Jose, CA, USA.  
April 2011 **Concurrent Analysis of Traffic in Streaming Video Data**, Doctoral Consortium in the Richard Tapia Celebration of Diversity in Computing Conference, San Francisco, CA, USA.  
February 2011 **Advanced Similarity Search of Realistic Timeseries**, Computer Science department, University of California - Riverside, Riverside, CA, USA.  
December 2010 **Understanding Traffic in Video Data**, 4th Symposium of Computer Graphics and Image Processing (SCGI-2010), San Pablo Catholic University, Arequipa, Peru

---

## Computer Skills (experience in parenthesis)

Programming	JAVASE (good), C/C++ (good), JAVAME (familiar), Matlab (good), C MPI (for parallel computing) (familiar), R (familiar), C# (familiar)	Professional Certification	Microsoft Certified Systems Administrator (familiar), Microsoft Certified Systems Engineer (familiar)
Technical writing	L <sup>A</sup> T <sub>E</sub> X(good)	OS	Windows (good), Linux (good)

---

## Professional References

**Dr. Dulce Poncion**: Research Manager, IBM Research (Dulce@almaden.ibm.com)  
**Dr. Curtis Dyreson**: Academic Advisor, Utah State University (Curtis.Dyreson@usu.edu)  
**Dr. Mike Hinchey**: Director of Lero-the Irish Software Engineering Research Centre and Professor of Software Engineering at University of Limerick, Ireland (Mike.Hinchey@lero.ie)  
**Dr. Ernesto Cuadros**: President of the Peruvian Computing Society (ecuadros@spc.org.pe)