

# André Belchior Mourão

Researcher with experience in Multimedia Information Retrieval, Computer Vision and Machine Learning

<https://andremourao.com>

[andre.b.mourao@gmail.com](mailto:andre.b.mourao@gmail.com)

+351917915979

I'm a researcher working on ways to **search and interpret multimedia data** (e.g., images, text, video) effectively at **large scales**. I've applied this knowledge to effectively **search** within **billions of image features**, find **relevant medical articles** for a patient case, discover **social media trends**, and help an insurance company **improve its document scanning process**.

This work resulted in **multiple publications** on top conferences and journals (e.g. **ACMMM, CIKM**), **\$40k** in **Microsoft Azure** Research Grants to test my ideas on a real cloud system and a **Fraunhofer Award** for Research of Practical Utility.

On my free time, I like to listen and play music, explore new places by running and regain my energy by cooking and enjoying craft beer. I've combined my expertise with two of hobbies (music and running) and created [powersong.run](http://powersong.run), a site that shows which songs and artists make you run faster and your heart beat harder.

## EDUCATION

2012 – 2018	<b>Ph.D. in Computer Science</b> <i>Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa (FCT/NOVA)</i>	Thesis Supervisor: Prof. Dr. João Magalhães
	Developed an architecture to improve the <b>partitioning</b> of <b>very large scale</b> (over 1 billion) image search for <b>distributed systems</b> using <b>Matrix Factorization and Clustering</b> . It was <b>benchmarked</b> on <b>Microsoft Azure cloud</b> , resulting in a system with a <b>difference in load of &lt;10% between nodes</b> .	
2010 – 2012	<b>M.Sc. degree in Computer Science</b> <i>FCT/NOVA</i>	Thesis Supervisor: Prof. Dr. João Magalhães
	Created an <b>SVM model</b> and <b>image features</b> for <b>real-time facial expression recognition</b> in C++. The model was successfully applied on a <b>competitive game</b> based solely on player's faces. Featured at multiple events: ExpoFCT (2012-17), Futurária (2014-17) and Sapo Codebits (2012)	
2007 – 2010	<b>B.Sc. degree in Computer Science</b> <i>FCT/NOVA</i>	

## EXPERIENCE

2018	<b>Post-doctoral researcher for the COGNITUS EU Horizon 2020 project</b>	Created the event plotting module, which enables journalists to find the best photos and videos for news articles and stories. It leverages on social media trends using <b>NLP</b> and <b>unsupervised ML techniques</b> with <b>Python, NumPy</b> and <b>Scikit</b> . Demo available at <a href="http://now.novasearch.org">now.novasearch.org</a> . Co-advised Master's student, resulting in top 5 TREC Precision Medicine run (thesis grade <b>19/20</b> )
2017	<b>Researcher at Die Mobiliar</b> ( <i>Collaboration with FCT/NOVA</i> )	Improved <b>Optical Character Recognition</b> quality for insurance documents by <b>10%</b> , using <b>image processing</b> (rotation detection, denoizing) and <b>NLP</b> (e.g. document language detection) with <b>Python</b> .
2015	<b>Visiting Researcher at the University of Texas, Austin</b>	Developed a <b>Java Android app</b> to for the <b>extraction of smartwatch activity and health data</b> . Created a <b>Django website</b> for the <b>analysis and visualization</b> of such data for patient assessment.

## ACADEMIC AND INDUSTRIAL PROJECTS

2017 – 2018	<b>COGNITUS – Converging broadcast and user generated content for interactive UHD services</b>	Developed an indexing architecture for large scale distributed multimedia search.
2013 – 2014	<b>Sapo Labs</b>	Developed <b>Machine Learning models</b> for <b>automatic tag generation</b> and key frame extraction for videos in the Sapo Videos platform using <b>C++</b> and <b>OpenCV</b> .

## TECHNOLOGIES, DEMOS AND SOFTWARE

Technologies	<b>C++</b> (OpenCV, Poco, Armadillo, LibSVM, FLANN), <b>Python</b> (NumPy, SciPy, sklearn, Django), <b>Java</b> (Lucene), <b>ElasticSearch</b> , <b>Microsoft Azure</b> , <b>Vowpal Wabbit</b>
Demos	<b>NovaMedSearch</b> <a href="http://medical.novasearch.org">medical.novasearch.org</a> Search engine accepts both images and text queries to find relevant cases in the biomedical literature. Responsible for image search ( <b>C++</b> ), medical text annotation and clinical trial search ( <b>Java</b> , <b>Lucene</b> )  <b>NovaVidSearch</b> <a href="http://video.novasearch.org">video.novasearch.org</a> Created a content-based video search that uses video frames as queries to find similar videos
Software	<b>DISH</b> (Distributed Indexing by Sparse Hashing) Created a framework for balanced partitioning and distribution of large-scale media indexes ( <b>C++</b> )  <b>NSS</b> (Nova Search Services) Created a <b>REST</b> framework for <b>image feature extraction, indexing, retrieval</b> , which is behind NovaMedSearch and NovaVidSearch image search ( <b>C++</b> and <b>Python</b> )
Datasets	novaemotions dataset <a href="http://novasearch.org/datasets/novaemotions">novasearch.org/datasets/novaemotions</a> Creator of the dataset, composed of <b>40.000 labelled facial expression images</b> , which is used by <b>dozens of researchers worldwide</b>

## AWARDS AND DISTINCTIONS

2017	<b>Microsoft Azure Research Award</b> – Awarded <b>\$20.000</b> in Microsoft Azure credit to apply on improving index partitioning for efficient distributed search of biomedical multimodal data.
2015	<b>Microsoft Azure Research Award</b> – Awarded <b>\$20.000</b> in Microsoft Azure credit to apply on research on learning PubMed cross-media relations.
2014	<b>TREC Clinical Decision Support</b> – Achieved the <b>2<sup>nd</sup> best Precision at 10</b> result on automatic runs.
2013	<b>Fraunhofer Portugal Challenge</b> Awarding Research of Practical Utility – <b>2<sup>o</sup> place: M.Sc. category</b> <b>ImageCLEF Medical</b> – Achieved the best overall retrieval results in the multimodal case-retrieval task (invited for <b>oral presentation</b> ). <b>TREC Federated Web Search</b> – Best nDCG@20 (normalized discounted gain) on snippet based runs (invited for <b>oral presentation</b> ).

## TEACHING EXPERIENCE

2017	<b>Web Search course (FCT/NOVA)</b> Invited lecturer on the Rank Fusion and <b>Learning to Rank</b> class for the Web Search graduate course.
2016	<b>Academia Big Data &amp; Advanced Analytics (Altran and FCT/ NOVA partnership)</b> Teaching assistant and responsible creating and grading the final student lab project.
2012 – 2014	<b>Object Oriented Programming course (FCT/ NOVA)</b> Teaching assistant for the Object Oriented Programming undergrad course.

## SELECTED PUBLICATIONS

- A. Mourão, J. Magalhães, “Low-complexity Supervised Rank Fusion Models,” ACM International Conference on Information and Knowledge Management, October 2018. **23% acceptance rate**
- A. Mourão, J. Magalhães, “Balanced Search Space Partitioning for Distributed Media Redundant Indexing,” ACM International Conference on Multimedia Retrieval, June 2017. **Oral presentation: 21% acceptance rate**
- P. Borges, A. Mourão, J. Magalhães, “Large-scale high-dimensional indexing by sparse hashing with  $l_0$  approximation” Multimedia Tools and Applications, 2016. **Journal Article**
- G. Tavares, A. Mourão, J. Magalhães, “Crowdsourcing facial expressions for affective-interaction”, Computer Vision and Image Understanding, 2016. **Journal Article**
- A. Mourão, F. Martins, J. Magalhães, “Multimodal medical information retrieval with unsupervised rank fusion,” Computerized Medical Imaging and Graphics Journal, 39, January 2015, **Journal Article**
- A. Mourão, J. Magalhães, “Competitive affective gaming: Winning with a smile”, Proceedings of the 21st ACM international conference on Multimedia (ACM MM 2013), October 2013. **Oral presentation: 20% acceptance rate**