Thomas Breuel

Professional Objectives

- leading world-class research in machine learning and deep learning
- solving real-world problems through machine learning

Research Interests

- very large scale machine learning
- distributed and GPU computing
- sequence modeling (e.g. time series, natural language)
- image and video analysis
- cognitive science, neuroscience

Skills

Primary Technologies: deep learning, machine learning, pattern recognition, signal/image processing

Primary Languages: Python, C, C++

Publications

Up-to-date list on Google Scholar: https://goo.gl/OnnFnf

Other Information

Citizenship: dual US/German citizen

Languages: bilingual English/German

Professional Activities: program committee for many conferences in pattern recognition, document analysis, and related areas; organization of the CBDAR series of workshops; summer schools and tutorials in deep learning and document analysis

Work Experience

2016-current
Distinguished Research Scientist
NVIDIA, Inc.

- key driver of very large (petascale) deep learning technologies at NVIDIA, combining systems research and development, deep learning technologies, hardware infrastructure
- development of novel deep-learning based document analysis and information extraction technologies for large commercial customers based on GPUs
- collaboration with researchers on image generation, tracking, motion analysis, and segmentation problems, with applications to self-driving cars and video analytics
- advisor to management senior management on novel computing technologies (quantum computing, containerization, cloud computing)

2014-2016
Researcher
Google, Inc.

- created new methods for automating and simplifying training of large scale neural networks
- technical supervision/contributions for Google projects: internationalization tools, OCR, Torch/Brain integration, Slim Tensorflow tools, distributed training tools
- research in mid-level vision for image segmentation, depth and motion estimation, 3D reconstruction, in particular using recurrent neural networks

2004-2014
Professor of Computer Science, U. Kaiserslautern
Director of the Image Understanding and Pattern Recognition Laboratory
(Joint appointment as a research director at the DFKI until 2009.)

- built a research group of 25 post-docs, graduate students, and undergraduate researchers
- obtained and managed research grants to finance research group (see list below)
- developed a new curriculum in pattern recognition, image processing, and human-computer interaction (see course contents below)
- supervised Ph.D. theses in computer vision, neural networks, machine learning, and document analysis
- acquired annual research grant funding from public, corporate, and not-for-profit sources, to grow to about 20 staff members
- successfully delivered both commercial and open source software to clients (see project list)
- regularly published in conferences in document analysis, image processing, computer vision, and pattern recognition

2004-present
Consulting and Contracting

- 2006-2008 Visiting Researcher (Contractor), Google, Mountain View, CA
- 2004 Visiting Researcher, Microsoft Research, Redmond, WA
- (additional engagements)

1998-2004
Xerox Palo Alto Research Center (PARC), Palo Alto, CA, USA
Senior Member of Research Staff

- Initiated and conducted research in computer vision, pattern recognition, and document analysis.
- Developed layout analysis technology behind UbiText, a software system for making documents accessible on handheld devices (See publications.)
- Initiated the GroupFire project, a project for collaborative and personalized Internet search methods, until its spin-out from PARC.
- Supervised student research in pattern recognition, guest lecturer at Stanford.
- Contributed to PARC infrastructure, planning, and project evaluation committees.

1995-1998
IBM Almaden Research Center (Computer Science), San Jose, CA, USA
Member of Research Staff

- Provided technology and support for IBM's DCS 2000 team on IBM's bid for automating the Year 2000 US Census.
- Delivered a system for browsing and accessing video image databases based on standard Internet client technologies (HTML, Java, HTTP) and IBM's database and storage backends in collaboration with other researchers and business groups. The system received the highest possible rating as an IBM 'first of a kind' project and has been used to showcase IBM's video technologies at major trade shows (including NAB).
- Developed novel methods and algorithms for content-based image retrieval from multimedia databases.
- Developed techniques and software for web-based access to image databases; the results of this work were used in web-based systems for access to museum image and art collections, a system for shape-based retrieval of trademark images, and other applications.

1994-1996
Consultant to the US Bureau of the Census

- Deployed handwriting recognition system after being selected as the single source for handwriting recognition technologies for the 1995 Census.
- Designed the handwriting recognition related workflow formats and standards for the system.
- Supported the execution and evaluation of handwriting technologies in the 1995 Census.

1992-1995
Institut Dalle Molle d'Intelligence Artificielle Perceptive (IDIAP), Martigny, Switzerland Directeur de Recherche

- Developed a research program in computer vision and handwriting recognition (see publications).
- Developed a handwriting recognition system combining neural networks and probabilistic finite state transducers for language modeling.
- Responsible for obtaining research funding, supervising student research, and collaborating with industry.

Education

1986-1992 Ph.D., Massachusetts Institute of Technology

MIT Artificial Intelligence Laboratory and Department of Brain and Cognitive Sciences, Cambridge, Massachusetts
Thesis: Geometric Aspects of Visual Object Recognition
Thesis Supervisor: Tomaso Poggio

- Developed the first practical, globally optimal algorithms for geometric matching.
- Demonstrated the feasibility of view-based recognition for 3D object recognition.
- Developed a technique for implementing lexical closures in C/C++ compilers that is binary compatible with existing C/C++ ABIs. The technique is in wide use today in the GNU C/C++ compilers.

1982-1986 B.A./M.A. Harvard University

Department of Biochemistry and Molecular Biology

Thesis: Analysis of the Promoter Region of the *hsp-82* Heat Shock Gene in D. melanogaster
Thesis Supervisor: Matthew Meselson.

1982-1992 German National Scholarship, Studienstiftung des Deutschen Volkes

1982 Winner, German national science competition "Jugend Forscht" (Operating System and Runtime for Microprocessors)

1978-1982 Georg Büchner Gymnasium, Letter, Germany

Abitur, Areas of concentration: Physics, Mathematics
Final Grade: 1.0 (top grade)

Patents

Systems and methods for unforgeable document tagging
US Pat. 6321648 - Filed Feb 14, 2000 - Xerox Corporation

System and method for searching and recommending documents in a collection ...
US Pat. 6493702 - Filed May 5, 1999 - Xerox Corporation

Systems and methods for forgery detection and deterrence of printed documents
US Pat. 6970259 - Filed Nov 28, 2000 - Xerox Corporation

Method for effecting actions over vertical surfaces
US Pat. 6457792 - Filed Nov 29, 1999 - Xerox Corporation

System for effecting actions over vertical surfaces
US Pat. 6367901 - Filed Nov 29, 1999 - Xerox Corporation

[APPLICATION] Method and system for document image layout deconstruction and redisplay system
US Pat. App 10/064,892 - Filed Aug 27, 2002

[APPLICATION] INFORMATION EXTRACTION FROM HTML DOCUMENTS BY STRUCTURAL MATCHING
US Pat. App 10/248,681 - Filed Feb 7, 2003

Projects and Research Grants

The following is a list of funded projects that Thomas Breuel has been a principal investigator or primary contributor for. Funding agencies and major collaborators are in parentheses. This page is only intended as an overview; if you like more information about a particular project, including the individual contributors, names of collaborators, amounts of funding, and durations, please contact me. Project descriptions are partially taken from the official project sites (i.e., project summaries are not completely written by me).

SICURA (BMBF, Smiths Detection, TU Darmstadt)

The SICURA project aims to develop object recognition and image database retrieval techniques for automated X-ray image analysis. Applications include airline carry-on and luggage screening. Challenges of the project include the presence of transparency, non-standard sensor geometries, and the absence of texture and fine shape detail in X-ray images, making the application of currently popular object recognition techniques difficult. The project concluded successfully in 2013 and delivered and demonstrated an integrated system consisting of contributions from Smiths Heimann, TU Darmstadt, and IUPR, capable of detecting weapons and liquid-filled bottles in X-rays with high reliability on a live X-ray machine. IUPR developed multiple subsystems capable of detecting both guns and bottles in cluttered luggage items. The public SICURA project web site is at http://goo.gl/No08LH (many of the results are not public).

DECAPOD (Mellon Foundation, U. Toronto, OCAD Toronto)

Decapod is a project that built a low-cost digitization solution that allows for rare materials, materials held in collections without large budgets, and other scholarly content to be digitized into a high-quality PDF format. This project incorporates off-the-shelf hardware and open source software to accomplish this goal. Technologies developed by IUPR include stereo-based book capture and dewarping, novel font generation methods from scanned text using clustering and linguistic constraints, high quality PDF generation. www.decapodproject.org

OCRopus (Google, DFKI)

OCRopus™ is a state-of-the-art document analysis and OCR system, featuring pluggable layout analysis, pluggable character recognition, statistical natural language modeling, and multi-lingual capabilities. The system has been initialy developed in our group with generous sponsoring by Google. As part of OCRopus, we have developed a number of new image analysis, document analysis, and character recognition technologies, including filter-based layout analysis, snakelet-based layout analysis, fast adaptive binarization using integral images, and highly scalable logistic local regression classifiers. The original system was written in C++. In order to speed up development, since 2009, the project has been converted to Python with some C/C++ modules. Since 2013, the primary recognizer is based on recurrent neural networks and yields state-of-the-art performance. We are currently working on replacing all parts of the system (binarization, layout analysis, text recognition, and language modeling) with recurrent neural networks. Research-quality command line tools and source code can be found at www.ocropus.org

TextGrid (BMBF, multiple partners)

TextGrid has, since its start in 2006, established the infrastructure for a virtual research environment for the humanities. In continuous exchange with the scientific community, TextGrid has developed a variety of tools and services available for free download in a stable version. Together with the TextGrid Repository, the Virtual Research environment TextGrid offers humanist researcher in the humanities sustainable editing, storing and publishing of their data in a thoroughly tested and safe environment. IUPR has contributed tools and software for OCR within the TextGrid environment, as well as developed a high performance Fraktur recognizer based on OCRopus. Official site: www.textgrid.de

NetSec (Deutsche Telekom, DFKI)

A project applying statistical, decision theoretic, pattern recognition, and machine learning techniques to the automated and adaptive analysis of network traffic. Goals include: Identification and remediation of DDoS attacks and intrusion attempts (zero-day exploits); behavioral analysis and anomaly detection; traffic modeling and forecasting in networks; early warning in critical infrastructures. The project released a number of open source packages, including BoNeSi (botnet simulator), repache (Apache traffic replay generator), nf-HiShape (high-performance traffic shaping for Linux).

InViRe / Moonvid (DFG, DFKI)

The aim of the InViRe project is the development of an usable prototypical system for content-based search in video databases. The project has explored and pioneered a number of ideas and technologies for enhancing the quality of

image and video retrieval, such as the use of large amounts of web data for enhancing retrieval results and the use of motion segmentation and statistical context to improve retrieval results (see publications). Here is the [project homepage.](#)

PaREn (BMBF, DFKI)

The PaREn Project is creating tools necessary for allowing non-experts to use, train, test, and deploy pattern recognition and machine learning modules in real-world software systems. The expected benefits are a far wider usage of pattern recognition and machine learning methods, leading to both better quality of the decisions and behaviors of software systems, as well as lower development costs. In addition to publications, resulted in a number of open source plugins to RapidMiner.

IPeT - Image-Based Personal Computing (BMBF, DFKI)

A project exploring the use of imaging technologies as part of personal computing environments, including camera-based interaction, camera-based document capture, visual similarity search, document image dewarping, round-trip visual HTML verification, and a number of other techniques. Here is a [list of demonstrators](#).

Courses Developed and Taught

Courses are usually 24 lecture hours per semester. Many of these courses are based on an extensive collection of iPython Notebooks implementing all the major algorithms and demonstrating the major techniques and results taught in class visually and computationally. Course contents vary somewhat from year to year (the URLs point to the last time the course was taught.) Courses are offered at the undergraduate (Bachelor) level and at the graduate (Master) level.

Learning and Perception lw.iupr.com (Bachelor)

Array programming, Bayesian methods, ML methods, nearest neighbor classification, feature extraction, kernel methods, perceptrons, clustering, VQ, k-means, linear and non-linear filtering, Fourier Transform, FFT, compression, template matching, DTW, HMMs. Applications in digital photography, multimedia image retrieval, speech recognition, and text analysis.

Neural Computation and Self-Organization ncso.iupr.com (Master)

Review of array languages and data parallel computation. Bayesian methods, decision theory, ML and Bayesian parameter estimation. Gradient descent methods, functions approximations, posteriors, and classification. Unsupervised learning: k-means, SOM, ICA, PCA. McCulloch Pitts neurons, perceptrons and MLPs. Neurons and action potentials, mathematical neuron models. Computational capabilities of linear threshold units and spiking neural networks. Turing-equivalence. Neural systems as nonlinear dynamical systems. Structure and properties of the human visual system. Development and self-organization in the visual system. Feature hierarchies, HMAX, convolutional neural networks. Psychophysical results on object recognition, detection theory. Attention, salience, grouping. Additional topics of current research interest.

Privacy, Identity and Computational Forensics compfor.iupr.com (Master)

The course covers methods for computational analysis of digital data and physical evidence. We will be examining how personal information and identity can be leaked and what techniques there are for protecting personal information. Possible topics include: interests groups: businesses, police, intelligence agencies, criminals, kinds of privacy and identity data, writer identification from text, handwriting identification from writing samples, printer identification from printouts, digital camera identification from images, social network analysis, pornography identification, forensic applications of image data, CAPTCHAs and reverse Turing tests, keystroke analysis for identity, biometric identification, cryptography, privacy, and identity, psychology and technology of phishing, multi-factor authentication, hard disk forensics, chain of custody and standards of evidence in forensics.

Foundations and Frontiers of Artificial Intelligence ffai.iupr.com (Master)

Topics include: the possibility of AI, strong/weak AI, Turing test; philosophical objections, Chinese rooms, symbol grounding, physicality; scientific method, automated scientific discovery; evolution of autonomous behavior, evolutionary psychology; synthetic biology, organic computing; foundations from neurobiology; Turing-Church thesis, physical limits of computation, hypercomputation and AI; theories of consciousness; free will, dualism; intelligent agents and game theory; direct brain machine interfaces, augmented cognition, mind reading; nanotechnology, transhumanism, the singularity; simulation hypothesis, virtual worlds; quantum mechanics and observation; social and intelligent agents, interaction, "laws of robotics"; history and literature of AI and artificial agents.

Natural Language Processing and Applications nlpa.iupr.com (Master)

Topics include: text processing and encoding; string algorithms, edit distance; statistical language models; spell correction; n-gram models; word sense disambiguation; Markov models, parts-of-speech tagging; probabilistic grammars and parsing; text alignment, clustering, text categorization; statistical machine translation; applications in speech recognition, handwriting recognition, and OCR; language acquisition; machine learning for NLP; cognitive and psychological aspects of NLP

Multimedia Information Retrieval mmir.iupr.com (Master)

Topics include: models of information retrieval, nearest neighbors, range queries, hash algorithms PLSA and topic models vector space models, text categorization, analysis, tagging, parsing, content based image and video retrieval, image retrieval based on color, texture, and shape, visual bag of words model, grouping, static and motion segmentation, scene cuts, geometric indexing, verification, and object recognition, automatic annotation and categorization, HMM-based methods (audio, video, music, document retrieval), generative and discriminative methods, selected special purpose applications, such as face detection, x-ray image analysis, performance evaluation and competitions, applications in consumer imaging, security, forensics, and copyright and plagiarism detection.

Document and Content Analysis dca.iupr.com (Master)

Most of the data we interact with day-to-day does not come in the form of data structures or databases, but instead in the form of documents and document images. This course introduces students to the formats, techniques, and algorithms used for representing, compressing, analyzing, processing, and displaying documents. Topics covered include: document formats and standards (TIFF, JPEG, PDF, PostScript, SVG), document image compression (G4, MRC, token based compression, JPEG2000), logical markup (HTML, XML, word processing formats, DocBook), writings systems of the world character sets and character encodings (ASCII, Unicode, special coding systems), text rendering, layout, ligatures, and hyphenation (Pango), typesetting and page layout systems (text flow, Word, LaTeX, etc.), OCR (character recognition, page segmentation), spelling and orthographic variation, statistical language modeling, document capture, page image dewarping and handheld document capture, named entity recognition, information extraction, table recognition, document search and retrieval, text mining, document databases, reading, psychophysics, and human-document interaction, document security and forensics.

Introduction to Human Computer Interaction (Bachelor), Intelligent Human Computer Interaction (Bachelor)

Courses generally introducing Bachelor students to the psychology and practice of HCI. Covers a variety of topics, for example, goals of HCI, test methods, Fitts law and related laws, HCI and security.

Additional Teaching

Each semester, I teach a seminar and a project class. The seminar introduces students to reading the literature, developing hypotheses, discovering opportunities for research in the literature, and summarizing and presenting their results. Past topics for seminars have included: recent topics in statistical NLP, AI and gaming, cortical models for visual object recognition, stereo vision. Projects get students familiarized with doing research, in collaboration with more experienced researchers in the group. Projects last 3-6 months and are part time. Project topics have included developing retrieval modules for multimedia databases, testing and evaluating different classifiers, developing and testing new document analysis software, and building 3D printers from scratch.