

# Behjat Siddiquie

☎ 301-646-9620 ✉ [behjats@gmail.com](mailto:behjats@gmail.com) [linkedin.com/in/behjatsiddiquie](https://www.linkedin.com/in/behjatsiddiquie) [behjat.github.io](https://github.com/behjat) 📍 Seattle, WA

## Summary

---

I am an applied researcher with over twelve years of industry experience. I have a track record of transforming state-of-the-art research into innovative consumer products and features. I led the research effort to launch Amazon Go's cashierless checkout technology in third party stores. As part of this effort, I built deep learning based computer vision systems that have been deployed in over a hundred physical stores worldwide, and serve over fifty million customers annually. My research in Generative AI has been published in top-tier conferences.

## Experience

---

### Amazon

Oct. 2018 – current

*Senior Applied Scientist/Tech Lead Manager on the Amazon Go research team*

*Seattle, WA*

- **Role:** I am a Research Lead at Amazon Go, building technology for [cashierless shopping](#). I lead a number of critical projects, in which I work with the engineering and product leadership to translate business goals into concrete research problems, design the technical roadmap for 4-6 member research teams, hire and mentor team members, lead the execution, and partner with engineering teams to productionize solutions. Select projects are listed below:
- **Improved customer receipt accuracy:** Achieved a 30% error reduction in Amazon Go's core receipt generation system by incorporating product context in transformer based customer activity detection models. (2022-23).
- **Launched Amazon Go in third party stores:** Built, from the ground up, a computer vision based system to generate, maintain and update, in real-time, a complete map of product placement in a physical store, which was a pre-requisite to Amazon Go launching in 3rd party stores. This system runs in over a [hundred 3rd party stores worldwide](#) and is based on deep learning models for 3D object detection and large-scale product identification. (2018–21).
- **Fundamental Research:** Led several student intern projects, leading to CVPR/ICCV publications in Generative AI.

### Amazon

May 2016 – Sep. 2018

*Applied Scientist on the Amazon Go research team*

*Seattle, WA*

- I designed, prototyped and deployed multiple key components of Amazon Go's person tracking system, improving accuracy, reducing costs and enabling Amazon Go to scale up from a single store in early 2018 to over a hundred stores in 2023. I developed deep learning based solutions for object detection, semantic and instance segmentation, person re-identification and multi-task learning.

### SRI International

Oct. 2011 – May 2016

*Senior Computer Scientist (Mar. 2015 – May 2016); Computer Scientist (Oct. 2011 – Mar. 2015)*

*Princeton, NJ*

- **Role:** I worked on several CV/ML projects, in various roles, including as a co-Principal Investigator and as a Tech Lead. Select projects are listed below:
- Developed a software suite for performing real-time human behavior recognition based on gestures, facial expressions and voice, using multi-modal sensors; modeled human interactions with temporal deep networks.
- Developed techniques to use image analytics for enhancing social-media based security applications.
- Co-authored several research grants, which received over \$2M in research funding from DARPA, ONR and AFRL.

### University of Maryland, College Park, MD, USA

Aug. 2006 – Oct. 2011

*Graduate Research Assistant (Advisor: Prof. Larry S. Davis)*

*College Park, MD*

- **PhD Thesis:** Image Retrieval based on Complex and Descriptive Queries.
- Active, online and incremental learning for visual scene understanding and image classification.
- Activity recognition in high, medium and low-resolution video.
- Research internships at IBM Research (2010) and SRI International (2009).

## Education

---

### University of Maryland, College Park

Aug. 2006 - Oct. 2011

*PhD in Computer Science (2011); MS in Computer Science (2009)*

*College Park, MD*

### Indian Institute of Technology, Bombay

July 2002 - May 2006

*BTech. in Computer Science*

*Mumbai, India*

## Selected Publications ([google scholar](#))

---

- Koutilya PNVR, B. Singh, P. Ghosh, **B. Siddiquie** and D. Jacobs, **LD-ZNet: A Latent Diffusion Approach for Text-Based Image Segmentation**, (ICCV) 2023, (Oral).
- M. Suhail, A. Mittal, **B. Siddiquie**, C. Broaddus, J. Eledath, G. Medioni and L. Sigal, **Energy-Based Learning for Scene Graph Generation**, (CVPR) 2021, ([nominated for the best paper award](#)).
- M. Amer, T. Shields, **B. Siddiquie**, A. Tamrakar, A. Divakaran and S. Chai, **Deep Multimodal Fusion: A Hybrid Approach**, (IJCV) 2018.
- **B. Siddiquie**, B. White, A. Sharma and L. S. Davis, **Multi-Modal Image Retrieval for Complex Queries using Small Codes**, (ICMR) 2014.
- M. Amer, **B. Siddiquie**, C. Richey and A. Divakaran, **Emotion Detection in Speech using Deep Networks**, (ICASSP) 2014.
- R. Feris, **B. Siddiquie**, J. Petterson, Y. Zhai, A. Datta, L. Brown and S. Pankanti, **Large-Scale Vehicle Detection, Indexing, and Search in Urban Surveillance Videos**, IEEE Transactions on Multimedia, 2012.
- **B. Siddiquie**, R. Feris and L. S. Davis, **Image Ranking and Retrieval Based on Multi-Attribute Queries**, (CVPR) 2011, (Oral).
- **B. Siddiquie** and A. Gupta, **Beyond Active Noun Tagging: Modeling Contextual Interactions for Multi-Class Active Learning**, (CVPR) 2010, (Oral).
- A. Kembhavi, **B. Siddiquie**, R. Mieziako, S. McCloskey and L. S. Davis, **Incremental Multiple Kernel Learning for Object Recognition**, (ICCV) 2009.

## Patents

---

- Inferring facility planograms, (Amazon), [11,093,785](#).
- Image-based detection of planogram product spaces, (Amazon), [11,842,321](#); [11,810,362](#).
- Re-identification of agents using image analysis and machine learning, (Amazon), [11,386,306](#).
- Recognizing salient video events through learning-based multimodal analysis of visual features and audio-based analytics, (SRI), [10,679,063](#).
- Exploiting multi-modal affect and semantics to assess the persuasiveness of a video, (SRI), [10,303,768](#).
- Dynamic hybrid models for multimodal analysis, (SRI), [9,875,445](#).
- Multi-Modal Modeling of Temporal Interaction Sequences, (SRI), [9,734,730](#).
- Image Ranking Based on Attribute Correlation, (IBM), [8,903,198](#).
- Object Detection in Crowded Scenes, (IBM), [8,811,663](#).
- Video based Detection of Multiple Object Types under Varying Poses, (IBM), [8,620,026](#).
- Multi-View Object Detection using Appearance Model Transfer from Similar Scenes, (IBM), [9,224,046](#).

## Technical Skills

---

**Languages:** Python, PyTorch, C/C++

## Citizenship

---

US Citizen