

Sindhu B Hegde



[sindhu-hegde.github.io](https://github.com/Sindhu-Hegde) [in linkedin.com/in/sindhu-b-hegde](https://www.linkedin.com/in/sindhu-b-hegde) [Google Scholar](https://scholar.google.com/citations?user=...)
github.com/Sindhu-Hegde [@ sindhu@robots.ox.ac.uk](mailto:sindhu@robots.ox.ac.uk)
+44 788 312 9860 Oxford, United Kingdom

RESEARCH INTERESTS

Computer Vision, Machine Learning, Deep Learning, Video Understanding, Multi-modal learning : Vision + Speech/Language

EDUCATION

- 2022 - Present **University of Oxford, UK**
PhD, *Advisor* : Prof. Andrew Zisserman
2025 Google PhD Fellowship recipient
- 2019 - 2021 **International Institute of Information Technology (IIIT), Hyderabad**
M.S. by Research, Computer Science and Engineering, CGPA : 9.67/10
Advisors : Prof. C V Jawahar and Prof. Vinay Namboodiri
- 2015 - 2019 **KLE Technological University, Hubli**
BE, Computer Science and Engineering, CGPA : 9.8/10 (**Gold medalist**)

NOTABLE PUBLICATIONS

- > [Sindhu Hegde](#), K R Prajwal, Taein Kwon, Andrew Zisserman, **Understanding Co-speech Gestures in-the-wild**, [International Conference in Computer Vision \(ICCV\)](#) 2025 (**Oral**).
- > K R Prajwal*, [Sindhu Hegde*](#), Andrew Zisserman, **Scaling Multilingual Visual Speech Recognition**, [ICASSP](#) 2025 (**Oral**).
- > [Sindhu Hegde](#), Andrew Zisserman, **GestSync : Determining who is speaking without a talking head**, [BMVC](#) 2023 (**Oral**).
- > [Sindhu Hegde*](#), Rudrabha Mukhopadhyay*, C.V. Jawahar, Vinay P. Namboodiri, **Towards accurate lip-to-speech synthesis in-the-wild**, [ACM-MM](#) 2023.
- > [Sindhu Hegde*](#), Rudrabha Mukhopadhyay*, Vinay P. Namboodiri, C.V. Jawahar, **Extreme-scale Talking-Face Video Upsampling with Audio-Visual Priors**, [ACM-MM](#) 2022.
- > [Sindhu Hegde*](#), Prajwal K R*, Rudrabha Mukhopadhyay*, Vinay P. Namboodiri, C.V. Jawahar, **Visual Speech Enhancement Without A Real Visual Stream**, [WACV](#) 2021.
- > Rudrabha Mukhopadhyay*, [Sindhu Hegde*](#), Vinay P. Namboodiri, C.V. Jawahar, **Audio-Visual Speech Super-Resolution**, [BMVC](#) 2021 (**Oral**).
- > Parul Kapoor, Rudrabha Mukhopadhyay, [Sindhu Hegde](#), Vinay P. Namboodiri, C.V. Jawahar, **Towards Automatic Speech to Sign Language Generation**, [INTERSPEECH](#) 2021.
- > [Sindhu Hegde](#), Shankar Gangisetty, **PIG-Net : Inception based deep learning architecture for 3D point cloud segmentation**, [Computers & Graphics](#), 2021.
- > [Sindhu Hegde](#), Shankar Gangisetty and Uma Mudenagudi, **3DOC : A Framework for Categorization of Rigid and Non-Rigid 3D Objects**, Women in Computer Vision (WiCV) Workshop @ CVPR, 2019.

Full publications list is available on [Google Scholar page](#).

EXPERIENCE

- Sept 2022 - April 2022 **Lead Data Scientist, VERISK, Hyderabad, India**
 - > **Image Forensics** : Worked on advancing image forgery detection on insurance claims data. Designed texture and exposure detection filters to make the algorithm more robust in real-world applications.

TensorFlow OpenCV ImageMagic
- March 2022 - October 2021 **AI/ML Engineer, VERISK, Hyderabad, India**
 - > **Automatic Speech Recognition (ASR)** : Worked on adapting ASR systems to telephonic speech. Achieved an absolute boost of 3 WER (word error rate) by designing a new lightweight ASR model.

PyTorch S3PRL ESPNet Kaldi

MULTILINGUAL LIP-READING (Multi-modal Learning)

DR. ANDREW ZISSERMAN

VGG, University of Oxford

May 2024 - Present

- Introduced a new large-scale public multi-lingual lip-reading dataset : MultiVSR, comprising a total of $\approx 12,000$ hours of video data covering 13 languages.
- Proposed a multi-task architecture capable of jointly performing visual language identification and lip-reading.
- Achieved strong performance across all languages on both tasks. Accepted as a full paper at ICASSP 2025 (oral).

[Project page](#) [MultiVSR dataset](#) [Code](#) [Video](#) [Poster](#)

GESTURE REPRESENTATION LEARNING (Semantic Gesture Understanding)

DR. ANDREW ZISSERMAN

VGG, University of Oxford

Jan 2024 - Present

- Given a gesture sequence, the goal is to learn a semantic-aware gesture representation by associating the gesture movements with the spoken content.
- Proposed a tri-model architecture that aligns gestures with both speech and text, capturing associations based on semantic content of text as well as style of speech (intonation, stress, prosody).
- Developed evaluation datasets and benchmarks for three tasks : (i) cross-modal gesture retrieval, (ii) gestured-word spotting, and (iii) active speaker detection using gestures. Evaluated model performance on these tasks and observed strong results. Accepted as a full paper at ICCV 2025 (oral).

[Project page](#) [AVS-Spot dataset](#) [Code](#)

GESTURE SYNCHRONIZATION (Temporal Gesture Understanding)

DR. ANDREW ZISSERMAN

VGG, University of Oxford

Jan 2023 - Dec 2023

- Introduced a new task “Gesture-Sync” : determining if a person’s gestures are correlated with their speech or not.
- Proposed a self-supervised model for this task and investigated the impacts of different video representations like RGB, keypoint vectors and keypoint images.
- Showed that the representations learned from the gesture synchronization task can be effectively transferred to the active speaker detection task, achieving superior performance in identifying “who is speaking” in multi-speaker scenarios. Accepted as a full paper at BMVC 2023 (oral).

[Project page](#) [Code](#) [Video](#) [Hugging Face Demo](#)

EXTREME-SCALE TALKING-FACE VIDEO UPSAMPLING (Multi-modal Learning)

DR. C V JAWAHAR, DR. VINAY NAMBOODIRI

CVIT, IIT Hyderabad

March 2021 - Oct 2022

- Given very low-resolution videos (8x8 pixels), the goal is to reconstruct realistic, high-quality talking-face videos at extreme scale-factors like 32x.
- Proposed a novel audio-visual talking-face video upsampling framework and demonstrated that exploiting the adequate priors in terms of the audio signal and a single target identity image is quintessential for the challenging task.
- Demonstrated the applicability of the proposed network in low-bandwidth video conferencing. Achieved superior quality of results with a great reduction in the compression ratio. Accepted as a full paper at ACM-MM 2022.

[Project page](#) [arXiv](#) [Video](#)

AUDIO-VISUAL SPEECH ENHANCEMENT (Multi-modal Learning)

DR. C V JAWAHAR, DR. VINAY NAMBOODIRI

CVIT, IIT Hyderabad

March 2020 - Aug 2020

- Tackled the problem of speech enhancement in unconstrained settings by suppressing the background noise and enhancing the voice of the target speaker.
- Proposed a novel pseudo-visual speech enhancement architecture which enhances the speech in natural and high noise conditions, even in cases where the visual information is unavailable or is corrupted.
- Introduced a hybrid paradigm which brings together the best of audio-only and audio-visual approaches by incorporating a lip synthesis model to generate accurate lip movements (pseudo-visual stream). Accepted as a full paper at WACV 2021.

[Project page](#) [arXiv](#) [Code](#) [Video](#)

3D OBJECT ANALYSIS (3D Computer Vision)

DR. SHANKAR GANGISSETTY, DR. UMA MUDENAGUDI

KLE Tech, Hubli

Feb 2018 - Aug 2019

- Proposed a framework for the analysis of 3D Point Cloud data and experimented on different applications such as 3D object classification, retrieval and segmentation.
- Explored the different encoding techniques to create a compact shape signature for representation of 3d objects. Provided an extensive evaluation using various combinations of local and global features for non-rigid and rigid point cloud retrieval task. Accepted as a full paper at BMVC 2019.

</> SKILLS

Python	● ● ● ● ●
PyTorch, Keras	● ● ● ● ●
OpenCV, Librosa, NLTK, PCL	● ● ● ● ○
C, Java, C#	● ● ● ● ○
Oracle, SQL Server	● ● ● ● ○
Shell scripting, HTML, MeanStack	● ● ● ○ ○

📖 COURSE WORK

Computer Vision	A
Introduction to NLP	A
NLP Applications	A
Big Data and Analysis	A
Statistical Methods in AI	A-
Digital Image Processing	A-

🏆 TECHNICAL AWARDS AND ACCOMPLISHMENTS

Google PhD Fellowship	2025	Google	Selected	One of 255 PhD students worldwide selected for the Google PhD Fellowship in Machine Perception.
ICASSP Travel Grant	2025	Hyderabad, India	Selected	Awarded a travel grant to attend the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2025.
ICVSS	2023	Sicily, Italy	Selected	Selected among the top 150 candidates to attend the International Computer Vision Summer School (ICVSS) in Italy.
Google AI Summer School	2022	Google Research	Selected	Selected among the top 150 candidates to participate in the Google Research India AI Summer School in both 2020 and 2022.
24hrs Hackathon	2018	INSZoom Pvt. Ltd.	1 st pos.	Developed <i>DocScanner</i> , an app that scans documents (e.g. passports), extracts key details, and auto-fills forms.
Paper Presentation	2018	Nat'l Tech Fest	1 st pos.	Presented our paper on analyzing restaurant reviews to support business growth.
36hrs Nat'l level Hackathon	2017	Sandbox Startups	1 st pos.	Developed a one-click emergency alert app that notifies and calls relevant services instantly.
Nat'l level IT Olympiad	2017	COEP, Pune	4 th pos.	Collaborated with students from diverse disciplines across the country to develop a GUI for controlling an evaporator pilot plant.
Ideathon : Smart Cities	2017	MSRIT, Bangalore	Finalist	Designed an emergency-response app to minimize delays by instantly sharing location, a brief voice clip, and auto-initiating calls.

👥 CO-CURRICULAR ACTIVITIES

Best Outgoing Student	Awarded as the <i>Best Outgoing Student</i> by KLE Technological University for the overall performance (technical and non-technical) in the course of 4 years (2019).
Voice-over artist	Selected as the voice-over artist for the promotion video of <i>NanoSorter Hamsa Plus</i> . 🔗 nanoPix
Host	Hosted various events including Bollywood singer Shaan's music concert during centenary celebrations of KLE Tech, Benny Dayal's music concert and DaanUtsav.
Best Student Orator	Received the <i>Best Student Orator of Karnataka</i> award from the renowned poet Mr. Channaveer Kanavi, organized by Children's Academy, Dharwad (2010).
Pratibha Puraskar Yogasan	Awarded as an <i>All Round Performer</i> by Women and Children Welfare Department, Belgaum (2010). Achieved 4 th position in international level yogasan competition, organized by Brahma Shree Narayana Guru Yoga Mandir, Mysore (2008).
Acting	Acted in a Kannada movie <i>Mandakini</i> with renowned actress Sudharani (Role : Student).

🤝 VOLUNTARY SERVICES

Reviewer	CVPR, ECCV, ICCV, IJCV, CVIU, WACV, AAAI
Moderator	Moderated several keynote sessions from eminent researchers during <i>Summer School on AI - 2021</i> and 2022 @ CVIT, IIIT Hyderabad