Sindhu B Hegde

linkedin.com/in/sindhu-b-hegde Google Scholar
github.com/Sindhu-Hegde sindhu.hegde@verisk.com
+91 720 438 2416 Hubli, Karnataka, India

Research Interests

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

Education

2019 - 2021 **International Institute of Information Technology (IIIT), Hyderabad**
M.S. by Research, Computer Science and Engineering, CGPA : 9.67/10
Advisors : Dr. C V Jawahar and Dr. Vinay Namboodiri

2015 - 2019 **KLE Technological University, Hubli**
BE, Computer Science and Engineering, CGPA : 9.8/10 (**Gold medalist**)

Notable Publications

- Sindhu Hegde*, Rudrabha Mukhopadhyay*, Vinay P. Namboodiri, C.V. Jawahar, **Extreme-scale Talking-Face Video Upsampling with Audio-Visual Priors**, Under review.
- Sindhu Hegde*, Prajwal K R*, Rudrabha Mukhopadhyay*, Vinay P. Namboodiri, C.V. Jawahar, **Lip-to-Speech Synthesis for Arbitrary Identities in the Wild**, Under review.
- Rudrabha Mukhopadhyay*, Sindhu Hegde*, Vinay P. Namboodiri, C.V. Jawahar, **Audio-Visual Speech Super-Resolution**, British Machine Vision Conference (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*, Prajwal K R*, Rudrabha Mukhopadhyay*, Vinay P. Namboodiri, C.V. Jawahar, **Visual Speech Enhancement Without A Real Visual Stream**, WACV 2021.
- Sindhu Hegde, Shankar Gangisetty, **PIG-Net : Inception based deep learning architecture for 3D point cloud segmentation**, Computers & Graphics, 2021.
- Sindhu Hegde, Shankar Gangisetty, **An Evaluation of Feature Encoding Techniques for Non-Rigid and Rigid 3D Point Cloud Retrieval**, BMVC 2019.
- 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

Present
October 2021

AI/ML Engineer, Verisk, Hyderabad, India

- **Automatic Speech Recognition (ASR)** for Telephonic Speech data
  - Working on adapting ASR systems to telephonic speech. Rigorously analysed different components involved in the ASR pipeline and achieved an absolute boost of 3 WER (word error rate) with our new lightweight model.
  - Leveraging self-supervised pre-training methods using additional loss functions that enforce focusing more on content rather than features like voice and style of the speech.

- **Vision & Language Understanding**
  - Given a detailed passage weakly related to an image (such as review text for a product), the goal is to extract the visually indicated keywords in the review text.
  - This would allow to make more abstract decisions for risk analysis, customer needs understanding and product recommendations.

PyTorch | S3PRL | ESPNet | Kaldi

Projects

Extreme-scale Talking-Face Video Upsampling (Multi-modal Learning) — **Dr. C V Jawahar, Dr. Vinay Namboodiri**

CVIT, IIIT Hyderabad — March 2021 - Present

- 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. Paper under review.

PyTorch | Librosa | OpenCV | DLib

Multi-speaker Lip to Speech Synthesis (Multi-modal Learning) — **Dr. C V Jawahar, Dr. Vinay Namboodiri**

CVIT, IIIT Hyderabad — Sept 2020 - Feb 2021

- Given a sequence of lip movements, the aim is to generate the speech for any speaker, in any voice.
- To explicitly model the inherent ambiguities of multi-speaker lip to speech, proposed a novel VAE-GAN approach, in which the lip and speech sequences are mapped to a probability distribution instead of a single point.
- Tackled the problem of the sentence-level, multi-speaker lip to speech synthesis with no explicit constraints on the number of speakers and vocabulary. Paper under review.

PyTorch | Librosa | OpenCV | DLib

Audio-Visual Speech Enhancement (Multi-modal Learning) — **Dr. C V Jawahar, Dr. Vinay Namboodiri**

CVIT, IIIT 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.

PyTorch | Librosa | OpenCV | DLib

3D Object Analysis (3D Computer Vision) — **Dr. Shankar Gangisetty, 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.
- Extracted handcrafted features like improved wave kernel signature (IWKS) and metric tensor and Christoffel symbols (MTCS) and demonstrated the effectiveness of these features on 3d object categorization. Accepted as a full paper at ICVGIP 2018.

PointCloudLibrary (PCL) | Keras | VlFeat

Restaurant review analysis (NLP) — **Dr. Shankar Gangisetty**

KLE Tech, Hubli — Jan 2017 - Feb 2018

- Proposed a sentiment based food classification framework which categorizes the restaurant food reviews into different meal courses based on the review sentiments. Found out the food items liked by most of the customers in each of these courses which helps in the growth of the restaurant business. Accepted as a full paper at ICACCI 2018.
- Also worked on fake review detection where we proposed an ensemble learning framework to filter out the fake reviews.

NLTK | Keras | Scikit-learn

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 AI Summer School	2022	Google Research India	Shortlisted	Selected among the top 150 candidates to participate in the Google Research India AI Summer School, both in 2020 and 2022.
Open House REU Exhibition	2019	R & D Cell, KLE Tech	1^{st} pos.	Project presentation and demo, built as part of Research Experience for Undergraduates (REU).
24hrs Hackathon	2018	INSZoom Pvt. Ltd.	1^{st} pos.	Built a *DocScanner* which automatically scans the documents such as passports, extracts the relevant information and auto-fills the forms.
Paper Presentation	2018	Nat'l Tech Fest Pleiades	1^{st} pos.	Presented our paper based on restaurant reviews which helps to boost the growth of the restaurant business.
36hrs Nat'l level Hackathon	2017	Sandbox Startups	1^{st} pos.	Built an app which sends emergency alerts and calls the respective emergency services with just a click of the button when the user faces any kind of emergency.
Nat'l level IT Olympiad	2017	COEP, Pune	4^{th} pos.	Collaborated with students of various streams across the nation and built a GUI interface to control the operations of evaporator pilot plant.
Ideathon : Smart Cities	2017	MSRIT, Bangalore	Finalist	Proposed an idea to reduce delay in response time during emergency. Built an app which sends real-time location and a short clip of voice recording and also enabled automatic calling through app.

Co-Curricular Activities

Best Outgoing Student	Awarded as the *Best Outgoing Student* 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 *NanoSorter Hamsa Plus*. 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 *Best Student Orator of Karnataka* award from the renowned poet Mr. Channaveer Kanavi, organized by Children's Academy, Dharwad (2010).
Pratibha Puraskar	Awarded as an *All Round Performer* by Women and Children Welfare Department, Belgaum (2010).
Yogasan	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 *Mandakini* with renowned actress Sudharani (Role : Student).

Voluntary Services

Reviewer	WACV 2022, AAAI 2022
Moderator	Moderated several keynote sessions from eminent researchers during 5^{th} $Summer\ School\ on\ AI$ - 2021 @ CVIT, IIIT Hyderabad.