ITUEvents

India Round– Part 1: 5G + AI + immersive + assistive services in telecommunications, ITU-ML5G-PS-19, 20, 23

Prof. Brejesh Lall, IIT/D

24 July 2020

ITU

AI/ML in 5G Challenge

Applying machine learning in communication networks

ai5gchallenge@itu.int

Register here

Join us on Slack

Sponsors

CISCO ZTE

Organizer

ITU

ITU AI/ML in 5G Challenge

Dr. Brejesh Lall
Prof., EED, IITD

brejesh@ee.iitd.ac.in
Indian Institute of Technology, Delhi, India

Website:
https://sites.google.com/view/iitd5g/

IEEE ComSoc
IEEE Communications Society
Delhi Chapter

Agenda

- Background
- Challenge on unstructured road conditions (ITU-ML5G-PS-019)
  - Background and Objective
  - Discussion
  - Datasets
- Challenge on shared experience (ITU-ML5G-PS-023)
  - Introduction
  - Scope
  - Tools
- Challenge on Immersive Video Conferencing (ITU-ML5G-PS-020)
  - Problem definition
  - Scope and Approach
  - Data Sources

Challenge 1 (ITU-ML5G-PS-019):

Smart Transportation in unstructured road conditions (Indian context)

preranam.jnu@gmail.com

Background and Objective

Background:

Indian Roads are very different than western roads

- **Unmarked, or incomplete lanes or disoriented road signs**
- **Heterogeneous nature of traffic (Two-wheelers, 3-wheelers, 4-wheelers, pedestrians)**
- **Varying traffic densities**
- **Unstructured road conditions (Potholes, rugged terrains etc.)**
- **Unpredictable behavior of traffic agents (pedestrians, cyclists, vehicles, etc.**

- Nearly 1,50,000 deaths every year in India due to road accidents

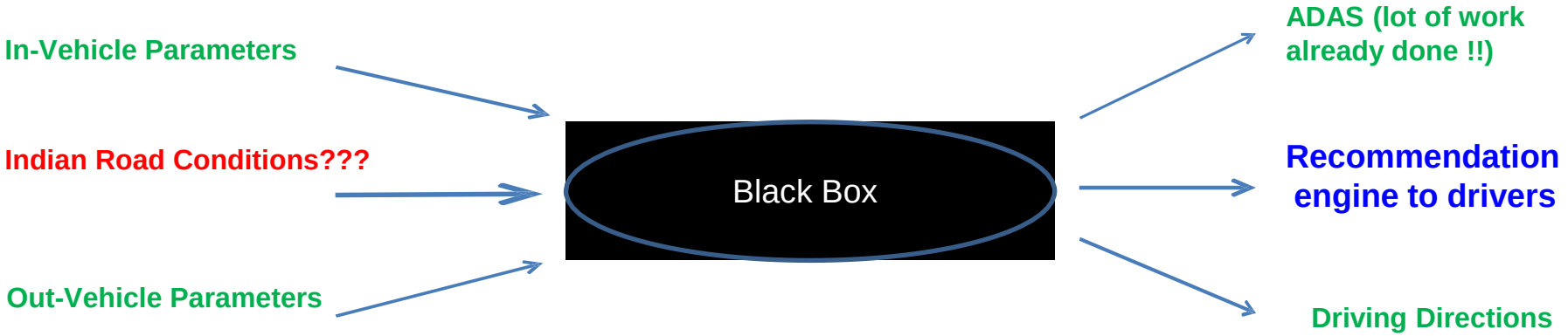
- >60 % of accidents occur in night diving conditions

Objective:

Scene understanding & Provide real-time recommendation engine to drivers' based on road and environment

Typical challenges

Improving Indian Road Safety

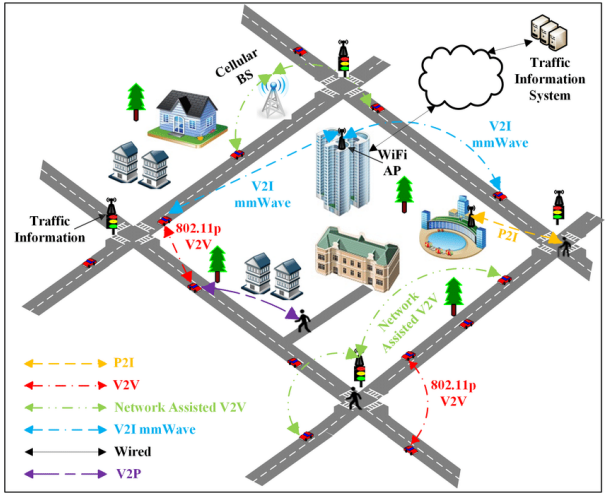


- **Stimuli Reception:** Info on other vehicles arrive 2-10s before event
- **Response Time:** Controlling vehicle takes min. 2s to 10 seconds

Real-Time Recommendation Engine within 0.5 – 1s for analysis

Change to Right Lane	Move Left in same Lane	Move Right in Same Lane
Maintain distance _ meters with vehicle ahead	Too close to pedestrian ahead	Traffic/Block ahead
+/- Acceleration	Apply Brakes	+/- speed by ____

1. V2X Communication



Picture Credits: 5GinITS

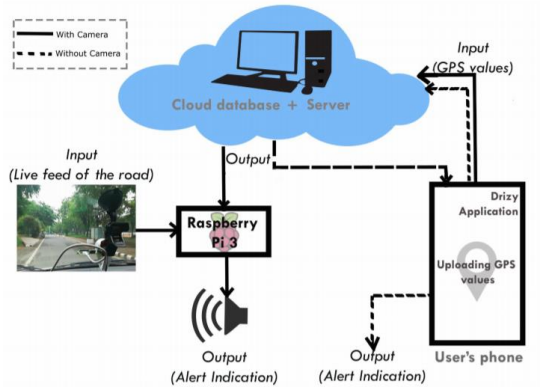
Challenge: Every vehicle should have V2X capability.

Simply NOT possible before 2030 !!!

1. Heterogenous information of vehicles
   - i. **Type of vehicle**
   - ii. **Varying traffic densities**
   - iii. **Driver response**
2. Potential time-based changes in the topology
3. Multicast routing mechanism

Our work on Collaborative Driver Assistance System

Challenging traffic settings on Indian roads (a) Two way road (b) Multiple pedestrians crossing simultaneously (c) Bus-stand (d) Medium density traffic (High heterogeneity) (e) Round turn (f) Market place



Click on Image Source

2. Camera-based Solutions

Indispensable and volume leader for ADAS and autonomous systems.

- Pixel based neighborhood assessment is traditional way of classification.
- Camera sensors that can capture only texture, color and contrast information.
- High level of detail captured by LiDARs, StereoRig, 360 degree cameras etc.
- Ever-increasing pixel resolution and reduced prices due to advancement in sensor technology.

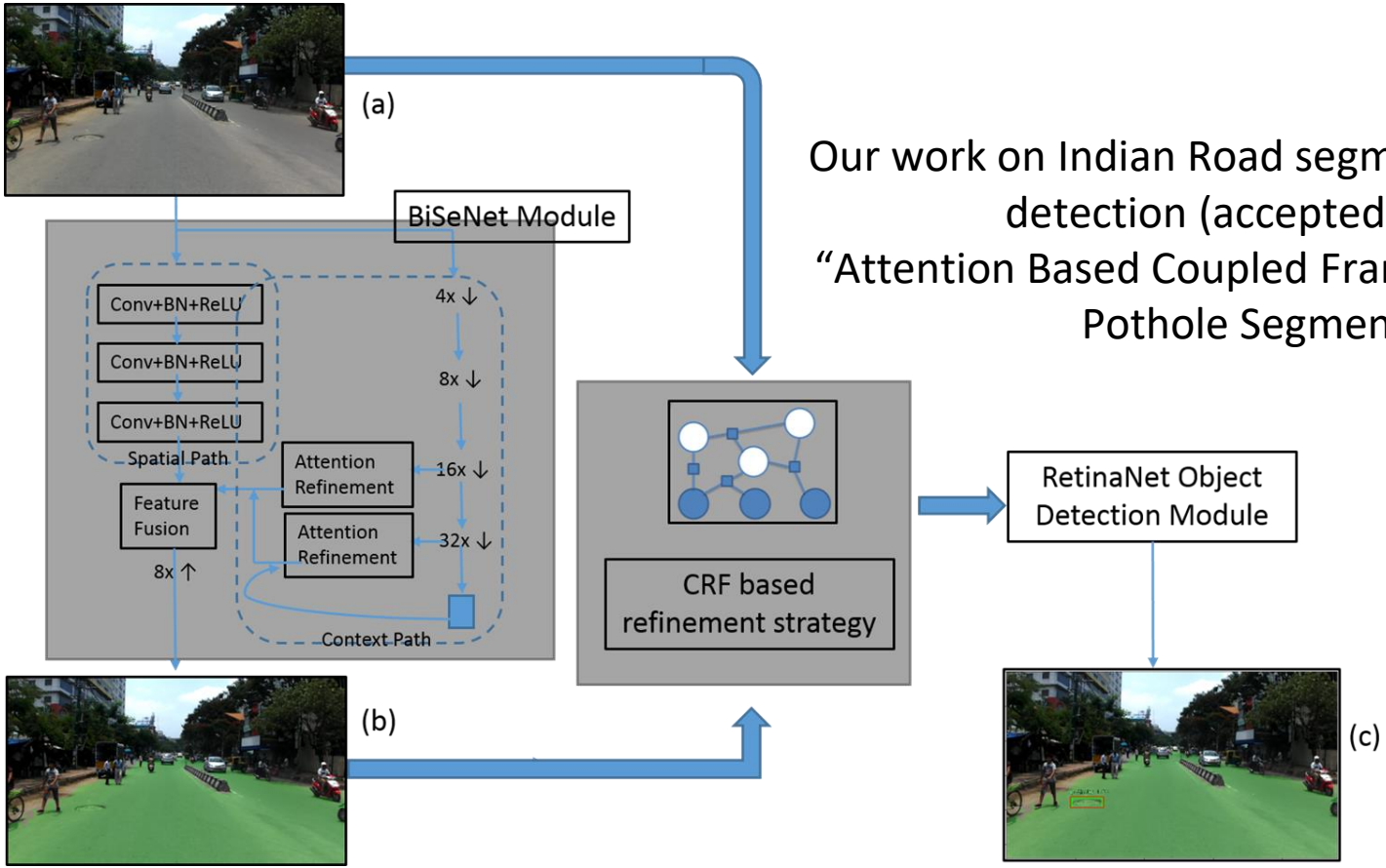
Drawbacks:

- Doesn’t work for adverse weather conditions and variations in lighting.
- Extensive processing which in turn increases the overall delay in real-time recommendations.
- Occlusion Issue: Only “see” the vehicles ahead,

  i.e., if there is a big vehicle, cannot provide data statistics beyond this vehicle Driver cannot decipher based on use of camera alone in autonomous driving

Example of how handcrafted features based detection and tracking works in Indian scenarios

Our work on Indian Road segmentation and Pothole detection (accepted in ICPR'20)
"Attention Based Coupled Framework for Road and Pothole Segmentation"



3. Improve Driver's Driveability

Driver's Driving Behavior:

- Typically looks at front of vehicle, within 60 deg – 120 deg view.

- Has different attention span based on various factors:
  - ✔ **Age of drivers**
  - ✔ **Talking to co-passenger, looking outside, no. of driving hour shift**
  - ✔ **Road condition and duration of the day**
  - ✔ **Traffic scenarios (quite varied in Indian scenario)**

  Can you measure or improve it?

Data Resources

Data source

A participant may be required to take permission to use this data-set (in case permission is required) and give due credits to the community hosting it. For train and test split use the standard split given in respective data sources else keep it as 7:3.

→ **https://insaan.iiit.ac.in/datasets/**

→ **http://www.cvlibs.net/datasets/kitti//**

→ **https://bdd-data.berkeley.edu/**

→ **https://www.kaggle.com/c/state-farm-distracted-driver-detection**

Resources: Use Collaboration or Google Credit Points or any other available free cloud resources.

Any controls or restrictions

This problem statement is open to all participants.

Specification/Paper reference (in Indian context)

1.Girish Varma, Anbumani Subramanian, Anoop Namboodiri, Manmohan Chandraker & C V Jawahar - IDD: A Dataset for Exploring Problems of Autonomous Navigation in Unconstrained Environments - IEEE Winter Conf. on Applications of Computer Vision (WACV 2019)

2.Sudhir Kumar Reddy, Girish Varma & C V Jawahar - Cityscale Road Audit System using Deep Learning - International Conference on Intelligent Robots (IROS'18)

3.Garg, N., Janveja, I., Malhotra, D., Chawla, C., Gupta, P., Bansal, H., ... & Lall, B. (2017, October). Poster: DRIZY: Collaborative Driver Assistance Over Wireless Networks. In Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking (pp. 546-548).

And many other papers in IEEE ITSS, IVS, CVPR, ICCV, ECCV, Ubicomp, Mobicom, Mobihoc etc.

Challenge 2 (ITU-ML5G-PS-023):

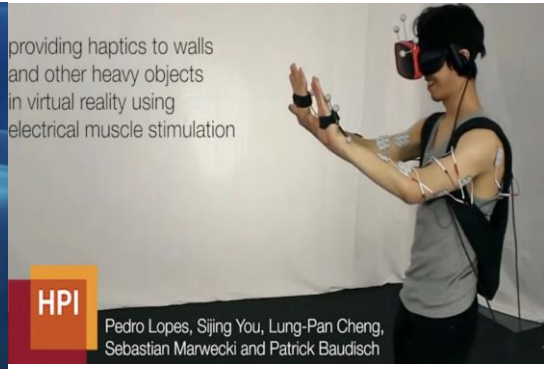
Shared Experiences (3D Augmented and Virtual Reality)

neerajku@hike.in, ankur@hike.in

Today

- World has evolved and social networking has become the norm
- Advancement of VR/AR gadgets, internet , high end mobiles
- Deep Learning models, softwares - Unity, Unreal

VR/AR Gadgets and technology



VR/AR Industry Use-Cases

1. Handheld AR - marketing , gaming, education
2. Location Based experiences - point of sale , advertising , museums
3. Outdoor Mobile AR - viewing points of interests.

Introduction- Shared experiences

- Notion of ‘being’ and ‘space’ in virtual environment
- True Immersive experience
- Interactive and multiple participants
- True representation of environment

Challenges- Shared experiences

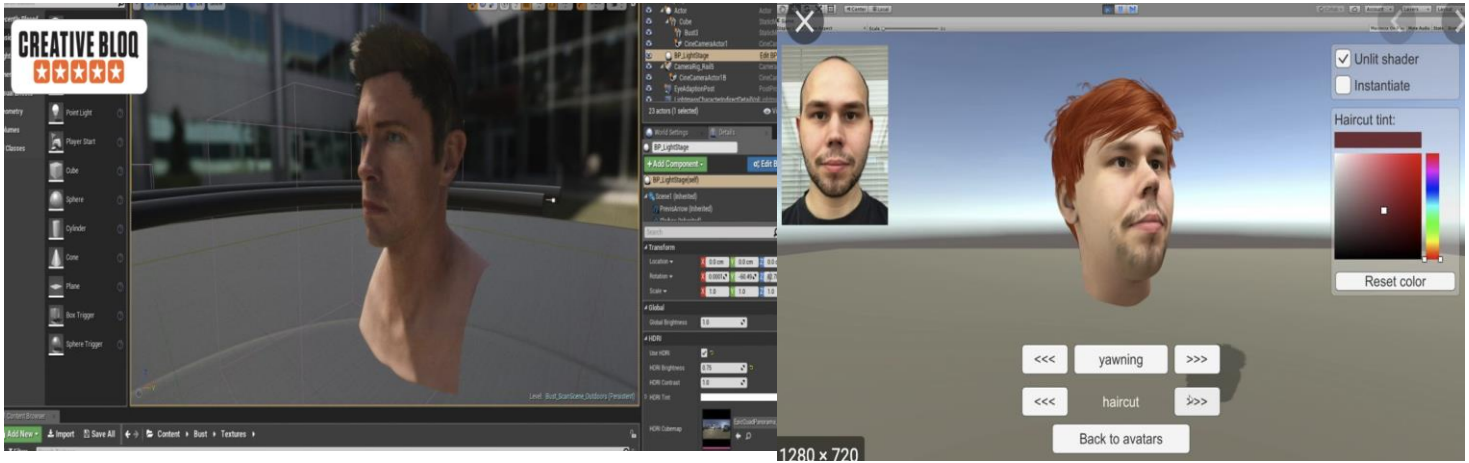
- Limitation in creating immersive experiences - togetherness, copresence, shared virtual space
- Latency and fidelity issue while creating the embodiment of ourselves
- Limitation in diversity of activities in virtual environment

Scope- Shared experiences

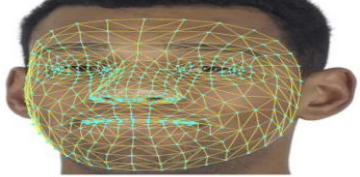
- Spatial Connection
- Identity, Co-presence, togetherness
- Activity diversity, 360 degree video/view sharing
- Real Time Interaction

Softwares

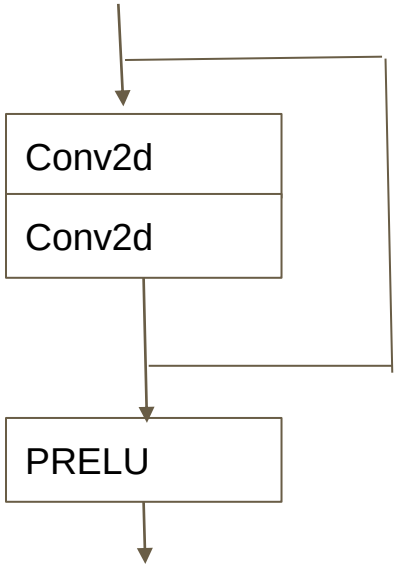
- 360 view of various places in polygoogle
- Unity , Unreal, Maya, Cinema-4D



3D face mesh Model



- FaceMeshBlock is main building block of the architecture.



https://github.com/tensorflow/tfjs-models

18 layers of FaceMeshBlock are used to generate 468 landmarks points needed to create 3D mesh

Model (input)	IOD MAD	Time, ms (iPhone XS)	Time, ms (Pixel 3)
Full (256×256)	3.96%	2.5	7.4
Light (128×128)	5.15%	1	3.4
Lightest (128×128)	5.29%	0.7	2.6

Table 1: Model performance characteristics

Objective-Shared Experiences

- Create a Immersive shared experience which contains following properties:

1. Interactive
2. Content curation : personalized and dynamic
3. Use case can be anything , birthday party, watching cricket/games, tourist site visit - Rohtang Pass , activity such as Paragliding

Data Resource

- Participants need to submit the documents showing flow charts and high level model/pipelines architecture
- Source Code of the implementation containing VR/3D based models to create shared experiences.
- Data Source -
    1. https://www.airpano.com/360photo/Taj-Mahal-India/
    2. https://www.youtube.com/playlist?list=PL-KnbPtxpFMwOTINzu45aLJxHl8Fmhoh4
    3. https://www.youtube.com/watch?v=OR_Y7vj66PU
    4. https://www.youtube.com/watch?v=7bD6xR8pAO4
    5. https://3d.bk.tudelft.nl/opendata/

- Related papers and Conferences - CVPR Workshop on Computer Vision for Augmented and Virtual Reality
- Deepali Aneja, Daniel McDuff, Shital Shah,A High-Fidelity Open Embodied Avatar with Lip Syncing and Expression Capabilities.
- Softwares - Animate Model such as UMA-2 in Unity
- Videos (360 degrees) of the experience / tourist site

Challenge 3 (ITU-ML5G-PS-020):

Improving experience and enhancing immersiveness of Video conferencing and collaboration

amitg@dview.ai, fauziyafarheen@gmail.com

Today

- Video Conferencing is literally a similar experience as it is used to be since last many years

Future communication networks are expected to support novel communication technologies such as multiple modalities based conferencing (an enhancement of video conferencing which is currently 3 modalities – speech, video and digital screen information). Multiple communication programs will co-exist with varying level of compression and representations. AI technologies are crucial to maintaining effectiveness of communication technologies involving multiple modalities and compression technologies over 5G networks. This challenge calls for methods and algorithms to improve perceptual effectiveness of enhanced communication techniques such as future video conferencing methods. Some possible communication approaches to assume are 3D holographic communication, real-time immersive communication (one to one and many to many video conferencing – e.g. attendees attending conferences virtually in an immersive environment).

Problem Definition

- To improve perceptual immersiveness, algorithms need to be designed to communicate a person's peripheral characteristics and interaction with the surroundings to communicate important information related to the event.
- An example is in a teaching scenario, the interaction of the teacher with the whiteboard is captured.
  The challenge is to capture and represent people's visual appearance and interaction to communicate and present the information in its entirety.

TO DO

- The full range of visual cues, e.g. posture, eye gaze, facial expressions, things like head nodding and hand gesturing; they all convey crucial information
- Surrounding interactions are equally important to convey the full contextual information in a video conferencing
- Create collaboration experience using multiple modalities (video, audio, shared contents, interactions with the environment)

SPECIFIC CHALLENGES

1. In a(n) (informal) video conferencing scenario, the video we see is in animated format, however there can a tool on the speaker side to find out local context (dressing style, facial grooming, ambient lighting etc). and incorporate it in the rendering at the other end.

2. In a teaching platform, again where we have animated rendering of the teacher and the board. here, the tool should mine the relative position (white board and teacher) and gestures (for emphasis e.g.) etc. and incorporate in the bit stream and render faithfully on the other side.

Inspirations

An immersive 3D video-conferencing system using shared virtual team user environments. (https://www.researchgate.net/publication/221274271_An_immersive_3D_video-conferencing_system_using_shared_virtual_team_user_environments)

Inspirations

https://www.youtube.com/watch?v=aO3TAke7_MI

Volumetric videos offer six degree-of-freedom (DoF) as well as 3D rendering, making them highly immersive, interactive, and expressive. https://dl.acm.org/doi/abs/10.1145/3301293.3302358

Data Sources and References

Data Sources

- Open data sources, e.g. http://live.ece.utexas.edu/research/LIVE_NFLX_II/live_nflx_plus.html
- The participants can portion the entire data set into testing and training for evaluation.
- We will generate our own data to ensure fairness which would not be shared with participants.

References

- https://www.sciencedirect.com/science/article/abs/pii/037872069490015 9
- DeepWarp: Photorealistic Image Resynthesis for Gaze Manipulation (https://link.springer.com/chapter/10.1007/978-3-319-46475-6_20)
- **A pipeline for multiparty volumetric video conferencing: transmission of point clouds over low latency DASH (**https://dl.acm.org/doi/abs/10.1145/3339825.3393578)

Questions

Thank you