Using emotion recognition and temporary mobile social network in on-board services for car passengers

Mario G.C.A. Cimino[1[0000-0002-1031-1959]], Pierfrancesco Foglia[1[0000-0001-6432-4504]], Cosimo A. Prete[1[0000-0002-8467-8198]], and Antonio Di Tecco[1,2]

[1] Department of Information Engineering, University of Pisa, Pisa, Italy
[2]University of Florence, Florence, Italy
{mario.cimino, pierfrancesco.foglia, antonio.prete}@unipi.it,
antonio.ditecco@phd.unipi.it

Abstract.
In next-generation cars, passengers will have more time for fun and relaxation, as well as the number of unknown passengers traveling together will increase. Thanks to the progress in Artificial Intelligence and Machine Learning techniques, new interaction models could be exploited to develop specialized applications that will be informed of the passengers' experience. The mood and the emotional state of driver and passengers can be detected, and utilized to improve safety and comfort by taking actions that improve driver and passengers' emotional state. Temporary Mobile Social Networking (TMSN) is a key functionality that can enhance passengers' user experience by allowing passengers to form a mobile social group with shared interests and activities for a time-limited period by utilizing their already existing social networking accounts. By minimizing isolation and promoting sociability, TMSN aims to redesign user profiles and interfaces automatically into a group-wise passengers' profile and a common interface. This work proposes and develops the generation of TMSN-inspired music selection through the Spotify music streaming service. The results obtained are promising and encourage further development toward the concept of in-car entertainment. Finally, we evaluate the performance of lightweight and heavy intelligent models that recognize the emotion of a person from its face, using Raspberry Pi 4 B devices. The results show that it is possible to realize a system with face detector and facial emotion recognition models on edge devices with sufficient performance (Frame per Second) to detect at least emotions expressed through macro-expressions.

Keywords: Autonomous Car, In-Car Entertainment, Temporary Mobile Social Networking, Music Streaming Service. Face Detection, Emotion Detection.

1 Introduction

The automotive industry is witnessing a real revolution stemming from the dramatic increase of ICT usage for improving vehicle safety, while promoting new entertainment services and studying solutions for autonomous driving [1], [2].

Nonetheless, the unstoppable and incremental advancement of autonomous cars, both owned or as shared commodities, will drive this innovation and consequently entertainment and personal assistance services will become essential for car passengers [3].

Thanks to the progress in Artificial Intelligence and Machine Learning techniques [4], new interaction models could be exploited to develop specialized applications that will be informed of the passengers' experience and situations, including personal interests and attractions, interactions within and outside the car environment, behavioral reactions to stimuli that change along the journey [3], [5]. Also, we will have the creating new model of temporary mobile social network, due to the increased time available to passengers as the level of car automation increases, and the presence of users sharing the same environment as in car sharing services [6].

In particular, the mood and the emotional state of driver and passengers can be detected [4], thanks to the enriched set of sensor available on-board. They can been utilized to improve safety and comfort by taking actions that improve the driver and passengers emotional state [4], and it can be exploited also to adapt the services offered by the applications to the user state via affective computing techniques [7], [8]. Preliminary examples are the BMW Emotional Browser [9] or the more recent Personal Assistant [10], that adapts the vehicle's interior to suit the drivers mood, but they utilize explicit driver inputs. The further innovation step is to realize a feedback tool, based on automatic user mood detection , that can be used to adapt the services offered to the needs of passengers [11], [12].

This paper is focused on the use of both Mobile Social Network (MSN) and Emotion Detection in the cart.

For what concern MSN, nowadays mobile technology ensures that car passengers have constant network connectivity and application functionality. As a result, they can easily obtain positive traveling experience by using mobile social networks, which offer a variety of entertainment options such as music and video streaming, feeds, stories, and so on [13], [2].

MSNs have already been introduced in cars through Android Auto or Apple CarPlay, the two major platforms for interoperability between smart phone and the car's dashboard information and entertainment unit. The number of MSN products can sensibly increase with the increasing level of car automation, in which the car controls a significant number of driving operations. Car sharing is also expected to gain popularity in this trend. As a result, next-generation cars will give passengers more time to have fun and relax, as well as increase the number of unknown passengers traveling together.

An important paradigm made possible by next generation cars is group-to-many interaction, which can further enrich an MSN user's interaction with the physical world by reducing his isolation. The user is alone in his physical world and is connected to the others via MSN in the traditional one-to-many interaction. With group-to-many interaction, a group of users temporarily lives in the same physical space (a car), interacts in person, and shares their MSN experience as a whole. The concept of Temporary MSN (TMSN) has received attention in the literature as a conceptual framework to be used at hotels, concerts, theme parks, and sports arenas, where people form a mobile social group for a limited time through common physical interaction [6]. People who are confined in a specific location (that can be a car) are allowed to join the TMSN and interact with others in a group-wise manner, improving mobile users' experiences with such temporal friends.

LobbyFriend was the first TMSN in the hotel industry, allowing establishments to maintain contact with guests throughout their stay, whether they were just staying at the same hotel or several in the area. All interactions in the TMSN are deleted when a guest leaves the establishment [6].

TMSN is more than just sharing a common space and a collaborative playlist [14], [15], it is about algorithms that exploit the user's profile and the current environmental context, allowing for augmented interaction through proactive services like automatic recommendation technology[16]. Today, powerful analytics and algorithms based on key techniques of modern sociology are successfully used to manage social media platforms. As such, in this study, TMSN is an additional intelligent layer on top of an existing ecosystem of services available on next-generation vehicles.

In this paper, a group-wise TMSN is proposed as a design paradigm for in-car entertainment. In particular, a functional design of an audio streaming automatic recommender protocol based on TMSN and Spotify analytics is illustrated in the context of social music. A prototype based on the Spotify API [17] has been developed and tested.

For what concern emotion detection, facial affect analysis is one of the less intrusive techniques that aims at estimating the emotions of a person [18], so its use can be exploited in the automotive environment [12]. According to the recent trend [18], a system for facial affect analysis is based on a Face Detector (FD), which find the location of a face (one or more faces) and extract it from an input image. The sub-image containing the face is then provided as input to a neural network (we call it FER – Face Emotion Recognition in the remaining of the paper) that finds facial landmarks estimates either categorical (happy, sad, etc.) or continuous dimensional measures of affective display, the most noteworthy of which are valence (how negative or positive the emotional display is) and arousal (how calming or exciting the emotional display looks like) [19].

FD and FER are generally realized by machine learning tools [18], [12], and their computational demand can be met by a cloud service, but this solution has limitations related to privacy, connectivity and the costs associated with building and maintaining the connected infrastructure [20]. An Edge solution, in which processing is performed locally, would greatly simplify the system architecture and solve the problems related to privacy an connectivity [20]. In this work, we identify also the level of performance achievable by running state-of-the-art Face Detection and Facial Emotion Recognition algorithms on devices typically used in the edge domain, and whether the achievable performance is compatible with real-time emotion detection, so that they can be executed on low cost in-vehicle hardware.

The paper extends [21], by performing new user experiments with a prototype TMSN system, and adds the evaluation of facial affect analysis on edge devices.

The paper is structured as follows. Section 2 illustrates the core concepts and functional design of a recommender protocol based on TMSN and Spotify analytics. Experimental results are given section 3. Section 4 is devoted to the performance analysis of different FED and FER algorithms running on an edge device. Section 5 concludes the paper.

2 FUNCTIONAL DESIGN of A TSMN PLAYLIST RECOMMENDER

The TMSN-based playlist recommender uses the ambient (car) and MSN data, when dealing with audio streaming applications. In this context, it is better to base the design on a standard music ontology [22] to improve interoperability (an important issue when dealing with distributed systems [23]), as the ontology provides a common vocabulary for exchanging music-related data across various applications [24].

Figure 1 depicts an ontology diagram showing the fundamental concepts and static relationships for audio streaming recommendation. Each concept is enclosed in a rectangular shape in the figure. Relationships connect concepts, and are represented by labelled oriented edges. Properties, shown in lower case, may also characterize some concepts. A track recommendation based on the passengers' social profiles is the main outcome of the ontology. The passengers' social profiles include listened tracks, artists, and genres. From the top-middle, a Passenger is in a Mobile Social Net, is in a Car, listens a Track, which is made by an Artist, and belongs to a Genre. A Car plays a Track, and manages a Temporary Mobile Social Net. The Temporary Mobile Social Net generates a Passengers Profile. On the other side, a Mobile Social Net builds a Social Profile, which is merged in a Passengers Profile. As a final point, the Passengers Profile recommends a Track.

The protocol of an audio streaming recommender based on TMSN and Spotify analytics is described via the Business Process Model and Notation (BPMN). It is a graphical representation built on a solid mathematical foundation to enable consistency checking execution, simulation, and automation [25]. It is also appropriate for standardizing and facilitating communication among all parties involved. A rectangular area in BPMN represents a participant who, via exchange of messages, gets involved in a protocol. The protocol is handled in each rectangular area by events (circles), activities (rounded boxes), and decision/merge nodes (diamonds). Solid and dotted arrows represent sequence flows and data flows; cylindrical shapes are used to represent data storage.

Figure 2 depict an audio streaming recommender protocol based on TMSN and Spotify analytics, built on the ontology in described in Figure 1 and represented via BPMN. For readability reasons, only the fundamental aspects of the proposed approach are covered. The protocol generates a playlist based on passengers' shared music interests (tracks, artists and genres), using Spotify Analytics.

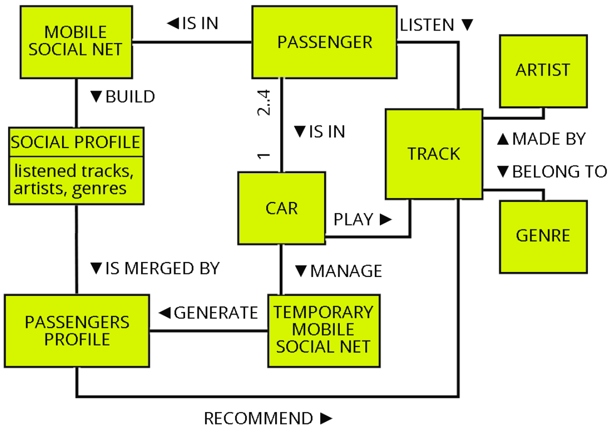


Fig. 1. TMSN Ontology in the context of audio streaming recommendation [21].

The starting of the protocol is represented by a white envelope in a thin circle, while its ending by black envelopes in thick circles. When new passengers are detected by the TMSN, the protocol begins (on top left in figure 5). It ends when a playlist is determined. As a first step, the TMSN requests all passengers' social profiles from Spotify Analytics. The gear icon for the task indicates that it is a service task, which is supported by Spotify's web services. The collected social profiles include each passenger's listened tracks, artists, and genres (according to the defined ontology). Social profiles are stored for protocol’s subsequent steps. In BPMN notation, script tasks represent internally developed task denoted by a sheet icon. The set of common tracks shared by all passengers is then identified by a script task. If any tracks are discovered, the recommended playlist is created. A script task, on the other hand, identifies the common artists. If common artists are found, the recommended play list is generated if some tracks are found by a service task that queries Spotify Analytics for the top tracks for them, or for their recommended tracks. If the last query produces no results, or there no common artist in the passengers’ social profiles, the common genres in social profiles are identified and, if any are found, the recommended playlist is generated by asking Spotify Analytics for the recommended tracks for common genres. Finally, if the previous task generates no results, the playlist is generated by the recommended tracks for top artist asked to Spotify Analytics by a service task.

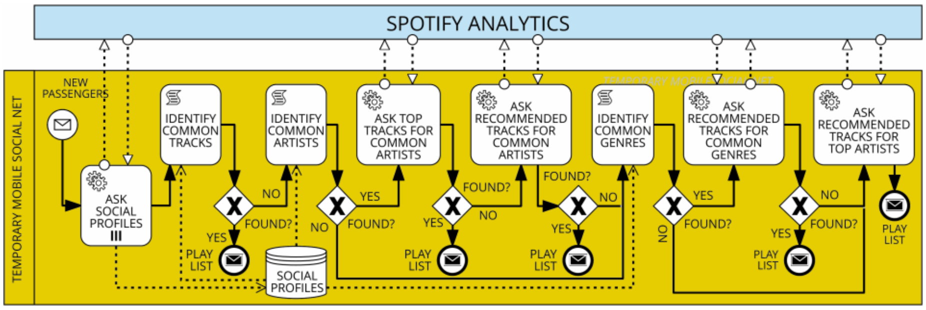


Fig. 2. BPMN protocol of an audio streaming recommender based on TMSN and Spotify analytics [21].

3 EXPERIMENTAL ANALYSIS

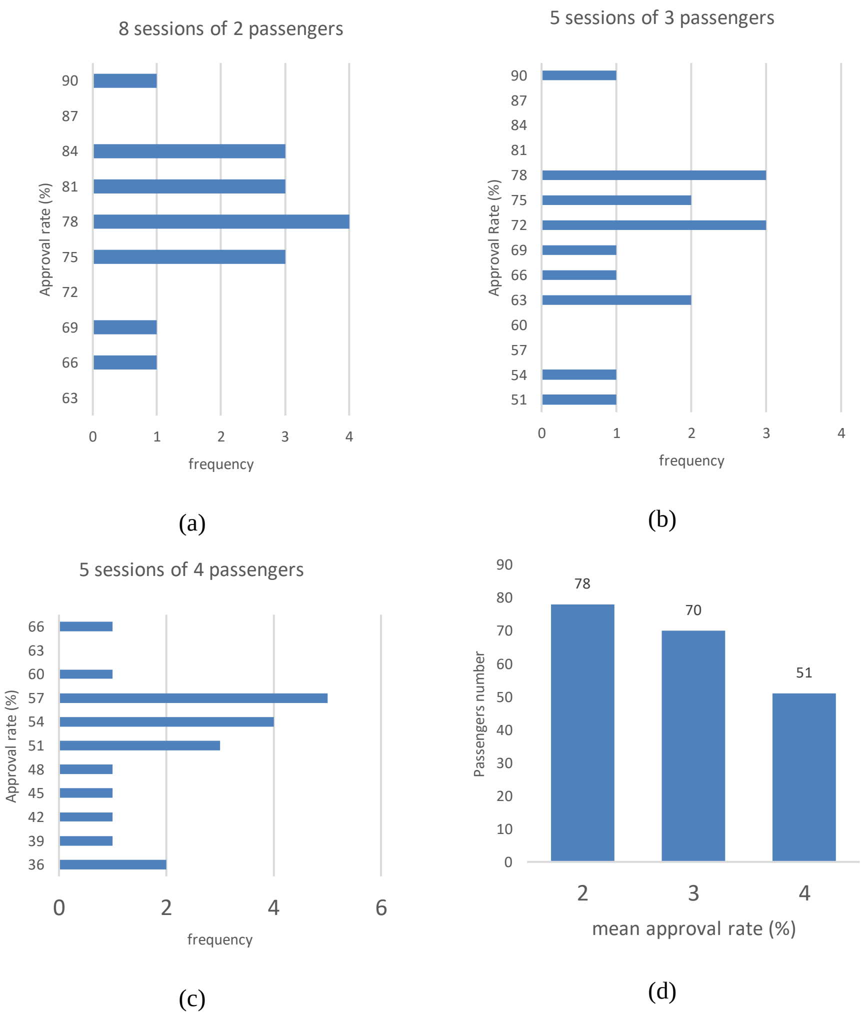
The protocol is purposely designed to heavily exploit Spotify Analytics services. Indeed, it has been implemented and experimented on both a desktop computer and a Raspberry PI4b, a small CMP [26] single-board computer, equipped with WIFI. Detail on the implementation may be found in [21].

We performed user test on the liking of the generated playlist to assess the validity of the protocol described in the previous session. Six people took part in carrying out the following sessions: 8 two-passenger sessions, 5 three-passenger sessions and 5 four-passenger sessions.

Each passenger has given a rating of liking/disliking at the end of listening to a recommended track. Each passenger's approval rate was calculated at the end of listening to the recommended playlist as follows:

$$approval\ rate\ = \frac{\text{number of } liked\ tracks}{total\ number\ of\ tracks} \times 100 \tag{1}$$

Results of the user test are given in Figure 3. In particular, Figure 3-a, -b, -c show the approval rate histograms for the 2, 3, and 4 passenger sessions, respectively. They show that the more the number of passengers increases, the more the approval rating decreases. This trend is due to the greater difficulty in finding common tracks, common artists and common genres as the number of passengers increases. Figure 3-d illustrates this trend based on the mean approval versus the number of passengers.



(a) (b) (c) (d)

Fig. 3. Approval rate histograms for: 2 passengers (a), 3 passengers (b), 4 passengers (c) and mean approval rate vs passengers' number (d).

Finally, we evaluated the generalization ability of the recommended playlist generated by the protocol (the generalization ability of the protocol). The generalization is evaluated as follows:

$$generalization\ rate = \text{No. of } known\ tracks \,/\, \text{No. of } liked\ tracks \times 100 \qquad (2)$$

If all of the tracks in its own playlist are liked, the maximum value is 1. (i.e., no additional liked tracks). The value decreases as the number of unknown tracks that are liked grows (i.e., better generalization).

Figure 4 shows the mean generalization rate versus the number of passengers attending the session. In general, as also observed in [21], the ability of the protocol to generalize increases, due to the greater variety of tracks, authors and genres available in the passenger playlists, that represents the starting point for the recommender protocol.

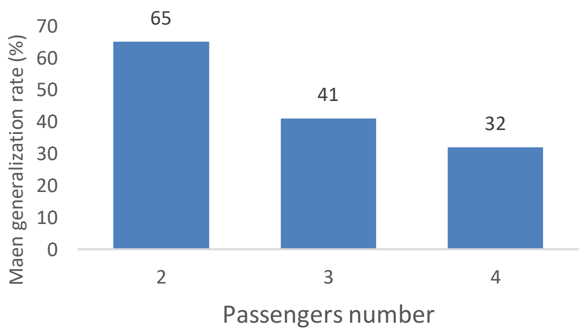


Fig. 4. Mean generalization rate

4 PERFORMANCE ANALYSIS OF FACE DETECTION AND EMOTION RECOGNITION ALGORITHMS

In this section, we characterize the performance of different FER and FD algorithms as run on Edge (low cost) hardware platforms.
As a hardware platform, we choose the Raspberry PI4-b computer board, which is a low-cost computing platform that is used for embedded system and general-purpose computing applications [27], equipped with WIFI interface. Thanks to its computing power that can be extended via accelerator, has been successfully utilized for realizing systems that run machine learning and deep neural network algorithm [28], [29].

4.1 Hardware systems

We utilized a Raspberry Pi 4 B as a base system to run the FD and FER algorithms by using the accelerator Neural compute stick 2 (NCS2).

The Raspberry Pi 4 B has 2 GB RAM, quad core Cortex-A72 64-bit 1.5 GHz CPU [30], 128 GB class 10 micro sd card, and we installed Raspbian Bullseye 11 64-bit OS thought Raspberry Pi Imager on Raspberry [31].

We used the accelerator Neural Compute Stick 2 (NCS2) to improve computing performance on Raspberries [32]. It has Intel Movidius Myriad X Vision Processing Unit processor, and an USB 3.0 Type-A to communicate with host devices. OpenVINO toolkit must be installed on host device for optimizing and deploying AI inference models in NCS2 device [33].

4.2 Selected Face Detection Algorithms

FD finds the location of a face (one or more faces) and extract it from an input image.

We studied and tested performance (mean execution time) by using algorithms present in the frameworks OpenCV and Darknet.

OpenCV is an open-source real-time optimized Computer Vision library used for different scopes [34], [35], [36]. OpenCV has many face detections algorithms, but after an explorative analysis based on the execution time, we selected Haar cascade [37] and Improved Local Binary Patterns (ILBP) [38] for further analysis.

Darknet is an open-source neural network framework available for different hardware architectures [39]. In Darknet, we considered the Yoloface-500k v2 lightweight machine learning algorithms for face detection [40]. This model is seen as an improvement based on YOLOv3 [41].

4.3 Selected Face Emotion Recognition Algorithms

FER algorithms find facial landmarks and estimates either categorical emotion or continuous dimensional measures of affective display, the most noteworthy of which are valence and arousal [18].

We studied and tested performance in term of accuracy by using two algorithms available in scientific literature. One algorithm is into Deepface framework, a lightweight deep face recognition library for Python [42], and the other one is Emonet [18].

Deepface FER neural network classifier that can distinguish among seven kinds of emotions as angry, disgust, fear, happy, sad, surprise, neutral. Author reports an accuracy of 59% on the FER dataset. Emonet is a more complex neural classifier that distinguish among eight kinds of emotions like contempt, etc., and generates as output also valence and arousal values, and other data. It has been specifically designed for the analysis of images of facial display recorded in naturalistic conditions. Authors report an accuracy of 75% on the AffectNet dataset.

We chose these two algorithms because they have different complexity, and therefore different computational requirements.

4.4 Analysis of Face Detection algorithm

Methodology. We recorded an experimental video to study the FD algorithms and compute the performance indexes to find the optimal FD.

The video was gathered through a webcam at 17 frame per second (FPS) in HD resolution by a volunteer that watched a movie of 01:15 min. The video contained 1275 frames with the participant's face. Then, the video has been converted in different resolutions by using the OpenCV bicubic interpolation method: 256x144, 426x240, 640x360, 854x480, 1280x720, and 1920x1080, to evaluate the effects of resolution on performance. The FD elaborated all frames for each resolution 30 times to compute the performances indexes.

We defined two main performance indexes: the execution time and the number of faces. The former is related to performance, the latter to detection quality.

Performance Analysis.

We compared the performance of the 3 chosen FD algorithms (HAAR, ILBP and Yoloface-500k v2) to select an FD for use in the subsequent analysis with RES systems. Our aims is to find a fast enough algorithm, with an acceptable quality of detection.

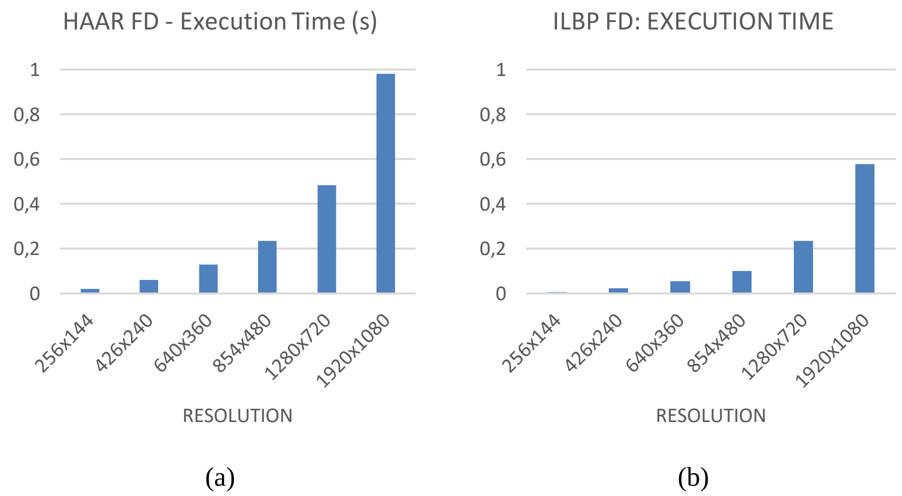


Fig. 5. Execution Time of: (a) HAAR FD and (b) ILBP FD.

Figure 5 (a) and (b) and Figure 6 show the execution time as the resolution varies for the HAAR, ILPB and Yoloface-500k v2, respectively. Execution time increases as resolution increases for both HAAR and BLP, and is less in BLP (about half). It is almost unchanged for Yoloface-500k v2 as the resolution varies, and is significantly lower than the other FDs (except for the two lowest resolutions). This is because Yoloface-500k v2 works on a constant frame size (352x288), and acquired frames are resized with a time on the order of a few msec. From the perspective of the number of faces not detected, Yoloface-500k v2 detects all faces in our experiments (figure not shown), HAAR (Figure 7-a) has a small number of undetected faces, while ILPB (Figure 7-b) has a large number of undetected faces at the lowest resolution. This value decreases significantly and then settles down to zero. Given the performance measured in our experiments, FD Yoloface-500k v2 will be used in subsequent analyses.

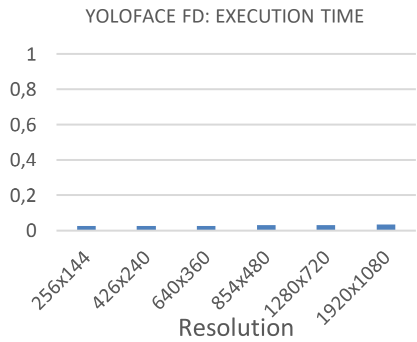
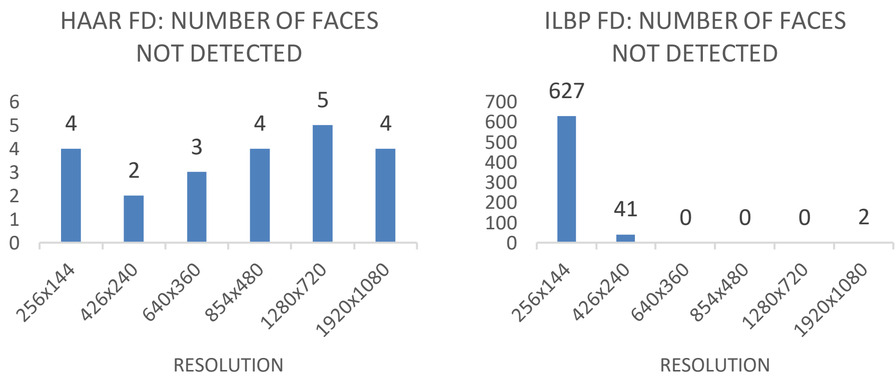


Fig. 6. Execution Time of the YOLOFACE FD.



(a) (b)

Fig. 7. Number of Faces not detected for: (a) HAAR FD and (b) ILBP FD.

4.5 Overall System performance

After the selection of the Face Detector, two versions of the overall systems were built utilizing the DeepFace and Emonet FER models, and their performance were compared. How did in the previous cases, to collect performance data, we run the experiments 30 times for each of the different resolution videos-considered .

The overall system with Deepface was deployed with both the FD and the FER running on the Raspberry Pi 4 B. The system implementing the Emonet FER was deployed with the FD running on the Raspberry and the FER running on the NCS2 accelerator. We used the accelerator because Emonet takes more than 10 seconds to process a frame when run on the raspberry, and this value is not compatible with a real-time detection. Moreover, we improved the execution time of the system deploying Emonet by configuring the accelerator to run in asynchronous (pipelined) mode

[33], so that the face detection of a frame was overlapped with the Emotion recognition of the previous frame.

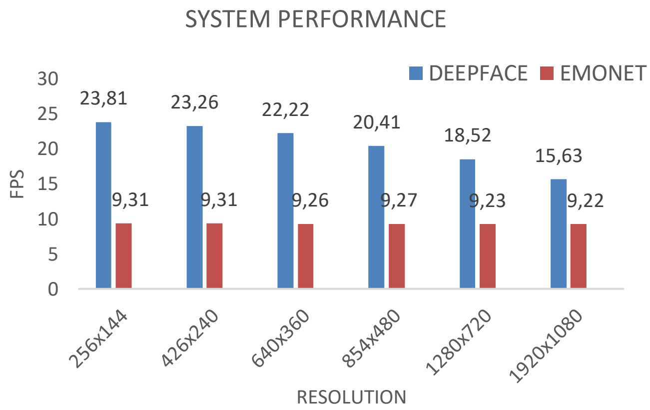


Fig. 8. Overall System Performance (expressed in Frame per Seconds, FPS) for the system implementing DeepFace and the one implementing Emonet. .

Figure 8 shows the performance (expressed as throughput, in frame per seconds - FPS) of the two systems. DeepFace has the highest throughput, always exceeding 15 FPS. The dependence by the resolution for DeepFace is mainly due to the image capture time, which increases as the resolution changes. For Emonet, the throughput is just over 9 FPS, and little dependent on resolution. This is because the throughput is dominated by the execution time of the FER, which is executed on the accelerator. In conclusion, both systems seem suitable for emotion detection, at least for emotions expressed through macroexpressions, which can be detected with the throughput achievable by both systems [43], [44], [45]. Future work involve both the evaluation of the FER systems with other devices and the exploitation of the hints furnished by the FER system within recommender systems.

5 CONCLUSIONS

In next-generation cars, passengers will have more time for fun and relaxation, as well as will increase the number of unknown passengers traveling together. Thanks to technological advances, new class of services can be introduced to improve drive and passenger's user experience and state.

In this work, we propose group-wise Temporary Mobile Social Networking recommender as a design paradigm for in-car entertainment. A functional design is illustrated in the context of social music and a prototype has been implemented, based on Spotify Analytics and running on Raspberry device. User test have been conducted by involving six people in various sessions attended by a different number of participants.

The approval and generalization rates obtained from the experiments show that as the number of passengers increases, the approval rate decreases for the lower number of common tracks, artists, and genres among passengers. Nevertheless, as the number of passengers increases, the generalization capacity of the system increases, providing a growing number of liked tracks that are not already known. The results of user tests are encouraging, and in particular the system's ability to generalize demonstrates the potential of the proposed approach in improving user experience. We plan to investigate the method further, enhancing the analytics of the system that creates the playlists and focusing more on the user experience while expanding our experimentation.

Finally, we evaluate the performance of lightweight and heavy intelligent models that recognize the emotion of a person from its face, using Raspberry Pi 4 B devices, to investigate the feasibility of implementing basic emotion detection services on board on low performance device without the support of external services. The results show that is possible to realize a system with Face detector and Facial emotion recognition models on edge devices with sufficient performance (Frame per Second) to detect at least emotion expressed through macro-expressions.

Future work involves both the evaluation of the intelligent model on other devices and the exploitation of the hints furnished by the FER system within recommender systems.

Acknowledgements

Work partially supported by the Italian Ministry of Education and Research (MIUR) in the framework of: (i) the CrossLab project (Departments of Excellence); (ii) the FoReLab project (Departments of Excellence); (iii) the National Recovery and Resilience Plan in the National Center for Sustainable Mobility MOST/Spoke10. Work partially carried out by the University of Pisa in the framework of the PRA_2022_101 project “Decision Support Systems for territorial networks for managing ecosystem services”. Research partially funded by PNRR - M4C2 - Investimento 1.3, Partenariato Esteso PE00000013 - "FAIR - Future Artificial Intelligence Research" - Spoke 1 "Human-centered AI", funded by the European Commission under the NextGeneration EU programme.

References

1. Alexia Athanasopoulou, Mark de Reuver, Shahrokh Nikou, Harry Bouwman, What technology enabled services impact business models in the automotive industry? An exploratory study, Futures, Volume 109, 2019, Pages 73-83, ISSN 0016-3287, https://doi.org/10.1016/j.futures.2019.04.001.

2. Bilius, L. B., & Vatavu, R. D. (2020). A multistudy investigation of drivers and passengers' gesture and voice input preferences for in-vehicle interactions. Journal of Intelligent Transportation Systems, 25(2), 197-220.
3. Connected car report 2016: Opportunities, risk, and turmoil on the road to autonomous vehicles. Strategy (2016). https://www.strategyand.pwc.com/reports/connected-car-2016-study.
4. Rong, Y., Han, C., Hellert, C., Loyal, A., & Kasneci, E. (2021). Artificial Intelligence Methods in In-Cabin Use Cases: A Survey. IEEE Intelligent Transportation Systems Magazine.
5. Arena, Fabio & Pau, Giovanni & Severino, Alessandro. (2020). An Overview on the Current Status and Future Perspectives of Smart Cars. Infrastructures. 5. 53. 10.3390/infrastructures5070053.
6. Yin, Y., Xia, J., Li, Y., Xu, W., & Yu, L. (2019). Group-wise itinerary planning in temporary mobile social network. IEEE Access, 7, 83682-83693.
7. Aranha, R. V., Corrêa, C. G., & Nunes, F. L. (2019). Adapting software with affective computing: a systematic review. IEEE Transactions on Affective Computing, 12(4), 883-899.
8. Foglia, P., Zanda, M., & Prete, C.A:, I. (2014). Towards relating physiological signals to usability metrics: a case study with a web avatar. WSEAS Transactions on Computers, 13, 624.
9. Meixner, G., Häcker, C., Decker, B., Gerlach, S., Hess, A., Holl, K., & Zhang, R. (2017). Retrospective and future automotive infotainment systems—100 years of user interface evolution. In Automotive user interfaces (pp. 3-53). Springer, Cham.
10. Liane Yvkoff, BMW Rolls-Out Its Intelligent Personal Assistant Feature Via Over-The-Air Update, URL: https://www.forbes.com/sites/lianeyvkoff/2019/05/30/bmw-rolls-out-its-intelligent-personal-assistant-feature-via-over-the-air-update/. Access date: Dec.2022
11. 12 Trends that Will Shape the Future of the Car Industry by 2030. Available online: https://www.hyundai.news/eu/stories/12-trends-that-will-shape-the-future-of-the-car-industry-by-2030/. Access date Dec. 2020.
12. Zepf, Sebastian & Hernandez, Javier & Schmitt, Alexander & Minker, Wolfgang & Picard, Rosalind. (2020). Driver Emotion Recognition for Intelligent Vehicles: A Survey. ACM Computing Surveys. 53. 1-30. 10.1145/3388790.
13. Coppola, R., & Morisio, M. (2016). Connected car: technologies, issues, future trends. ACM Computing Surveys (CSUR), 49(3), 1-36.
14. Spotify CL, Collaborative Playlist, support.spotify.com/us/ article/collaborative-playlists/, accessed Dec 2022.
15. Spotify FM, Family Mix, support.spotify.com/us/article/ family-mix/, accessed Dec 2022.
16. Cimino M.G.C.A., Lazzerini B., Marcelloni F., Castellano G., Fanelli A.M., Torsello M.A. (2011). A Collaborative Situation-Aware Scheme for Mobile Service Recommendation, Proceedings of the 11th International Conference on Intelligent Systems Design and Applications, 130-135 (2011), doi: 10.1109/ISDA.2011.6121643
17. Spotify API, support.spotify.com/us/article/spotify-in-the-car/, accessed Dec 2022.
18. Toisoul, A., Kossaifi, J., Bulat, A., Tzimiropoulos, G., & Pantic, M. (2021). Estimation of continuous valence and arousal levels from faces in naturalistic conditions. *Nature Machine Intelligence*, *3*(1), 42-50.
19. Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. Psychological bulletin, 139(4), 917.

20. Y. -L. Lee, P. -K. Tsung and M. Wu, "Techology trend of edge AI," 2018 International Symposium on VLSI Design, Automation and Test (VLSI-DAT), 2018, pp. 1-2, doi: 10.1109/VLSI-DAT.2018.8373244.
21. Cimino M.G.C.A., Di Tecco A., Foglia P., et al. In-Car Entertainment via Group-wise Temporary Mobile Social Networking (2022) Internl. Conf. on Vehicle Technology and Intelligent Transport Systems, VEHITS - Proceedings, pp. 432 – 437. DOI: 10.5220/0011096000003191
22. MO, Music Ontology, musicontology.com, accessed Dec. 2022.
23. Campanelli S., Foglia P., Prete C.A. An architecture to integrate IEC 61131-3 systems in an IEC 61499 distributed solution (2015) Computers in Industry, 72, pp. 47 – 67.
24. Ciaramella, A., Cimino, M.G.C.A., Marcelloni F., Straccia, U. (2010). Combining Fuzzy Logic and Semantic Web to Enable Situation-Awareness in Service Recommendation", Lecture Notes in Computer Science, 6261:31-45.
25. Cimino, M.G.C.A., Palumbo, F., Vaglini, G. *et al*. (2017) Evaluating the impact of smart technologies on harbor’s logistics via BPMN modeling and simulation. Information Technology Management, 18, 223–239.
26. Foglia, P., & Solinas, M. (2014). Exploiting replication to improve performances of NUCA-based CMP systems. ACM transactions on embedded computing systems (TECS), 13(3s), 1-23.
27. Daher, A.W., Rizik, A., Muselli, M., Chible, H., Caviglia, D.D. (2021). Porting Rulex Machine Learning Software to the Raspberry Pi as an Edge Computing Device. In: Applications in Electronics Pervading Industry, Environment and Society. ApplePies 2020. Lecture Notes in Electrical Engineering, vol 738. Springer, Cham.
28. Zamir, M.; Ali, N.; Naseem, A.; Ahmed Frasteen, A.; Zafar, B.; Assam, M.; Othman, M.; Attia, E.-A. Face Detection & Recognition from Images & Videos Based on CNN & Raspberry Pi. *Computation* 2022, *10*, 148.
29. A. A. Süzen, B. Duman and B. Şen, "Benchmark Analysis of Jetson TX2, Jetson Nano and Raspberry PI using Deep-CNN," *2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 2020, pp. 1-5.
30. Raspberry Pi 4 B, https://www.raspberrypi.com/products/raspberry-pi-4-model-b/, Access date: Dec 2022
31. Raspberry Pi Imager, https://www.raspberrypi.com/software/, Access date: Dec. 2022
32. Neural Compute Stick 2, www.intel.com/content/www/us/en/developer/articles/tool/neuralcomputestick.html, Access date: Dec 2022
33. Intel Distribution of OpenVINO Toolkit, www.intel.com/content/www/us/en/developer/tools/openvinotoolkit/overview.html, Access date: Dec 2022
34. OpenCV, https://opencv.org/, Access date: 04/12/2022
35. Bradski, G.; Kaehler, A. Learning OpenCV: Computer Vision with the OpenCV Library; O’Reilly Media, Inc.: Sebastopol, CA,USA, 2008.
36. De Vitis, G.A., Foglia, P. and Prete, C.A. (2020), Row-level algorithm to improve real-time performance of glass tube defect detection in the production phase. IET Image Process., 14: 2911-2921. https://doi.org/10.1049/iet-ipr.2019.1506.
37. Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International journal of computer vision*, *57*(2), 137-154.
38. Jin, H., Liu, Q., Lu, H., & Tong, X. (2004, December). Face detection using improved LBP under Bayesian framework. In Third International Conference on Image and Graphics (ICIG'04) (pp. 306-309). IEEE.

39. Joseph Redmon, Darknet: Open-Source Neural Networks in C, Darknet, https://pjreddie.com/darknet/, Access date: Dec 2022
40. Ma, X. https://github.com/dog-qiuqiu/MobileNet-Yolo, Access date: Dec 2022
41. Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.
42. Serengil, S. I., & Ozpinar, A. (2021, October). Hyperextended lightface: A facial attribute analysis framework. In *2021 International Conference on Engineering and Emerging Technologies (ICEET)* (pp. 1-4). IEEE.
43. Bhatti, Y. K., Jamil, A., Nida, N., Yousaf, M. H., Viriri, S., & Velastin, S. A. (2021). Facial expression recognition of instructor using deep features and extreme learning machine. Computational Intelligence and Neuroscience, 2021.
44. D. Matsumoto and H. S. Hwang, “Reading facial expressionsof emotion,” Psychological Science Agenda, vol. 25, 2011.
45. Ekman, P. (2003). Emotions revealed (2nd ed.). New York: Times Books