Regional Trajectory Analysis through Multi-Person Tracking with mmWave Radar

Andre Pearce, J. Andrew Zhang and Richard Xu
School of Electrical and Data Engineering
University of Technology Sydney
Sydney, Australia
andre.pearce@student.uts.edu.au, {andrew.zhang, yida.xu}@uts.edu.au

***Abstract*—Understanding characteristics of an environment can benefit the ability to perform more reliable multiple object tracking (MOT). This paper proposes an approach to extract entry and exit points for an environment using MOT trajectories. The observed environment is divided into a grid, in which regions are defined for. The trajectories obtained from the MOT data are organised into their respective regions. For each region an activity heatmap is formed and classified using a convolutional neural network (CNN) to determine if the region consists of either an entry or exit point. The entry and exit points classified are projected onto the MOT plane to illustrate the entry and exit points of the observed environment. The approach proposed in this paper provides a foundation in which future work can build on to enhance the MOT capability in real-time through a greater understanding of the observed environment.**

***Index Terms*—multiple object tracking, movement pattern, trajectory pattern, mmWave, convolutional neural network**

I. Introduction

Multiple object tracking (MOT) with mmWave has become quite achievable and researched due to recent innovations in mmWave radars within the last 5 years. The ability to perform multiple object tracking with mmWave is typically a standalone architecture purely involving signal analysis techniques. Through this approach alone, limitations can be seen when situated in complex environments involving frequent environment interactions from the tracked targets.

The authors of [1] demonstrate a thorough implementation to performing MOT with mmWave. Another approach to performing MOT with mmWave can be seen in [2]. The novel advancements and differences, in the context of mmWave MOT, presented in [1] and [2] reside in the tracking prediction algorithm. Although the implementation techniques are different in [1] and [2], the fundamental architecture to MOT with mmWave remains consistent. Due to the inability of mmWave reflections to penetrate solid objects [3], multiple object tracking in environments where many solid objects are presented can be unreliable and inconsistent. This problem is notable in indoor environments where it is common to have many fixed objects obstructing the field of view of the radar, such as furniture, columns, etc. The implementations demonstrated in [1] and [2], as well as others following a similar mmWave MOT architecture, are subject to this tracking problem.

The authors are with the Global Big Data Technologies Center and the School of Electrical and Data Engineering at the University of Technology Sydney, Sydney, Australia.

This paper ultimately proposes an additional step to the general architecture of MOT with mmWave radar. The primary aim of the additional step is to extract generalised tracking patterns to ultimately provide insight into characteristics of the environment that is being observed. A movement pattern in the context of this paper can ultimately be explained as a commonality of transitions between different start and end states. Regional dominant movement patterns can be defined as the most frequently occurring transitional pattern for a given start and end state.

The study of regional dominant movement patterns in trajectory data has been well explored. A study performed by [4] demonstrates the potential to determine regional dominant movement patterns in trajectory data using a pre-defined taxonomy of trajectory patterns. The authors in [4] define different types of movement patterns, which are then used as a criteria for a trained convolutional neural network (CNN) to classify trajectory clusters. Another study performed in [5] demonstrates a novel approach for mining trajectory patterns, as opposed to utilising a taxonomy of pre-defined patterns. Mining trajectory patterns is a necessary approach to take in situations where the types of trajectory patterns that might occur are not known. In situations where particular trajectory patterns are known or searched for, maintaining a taxonomy of those patterns can yield more performant.

Applying the concept of regional dominant movement patterns to trajectory data, termed as regional dominant trajectory patterns (RDTP), can fundamentally expose characteristics of the environment that is being sampled in the field of view. This paper aims to demonstrate a novel approach to extract environmental characteristics through movement patterns in trajectory data.

The presented novel advancements and philosophy applied to mmWave MOT make the following contributes:

- More reliable tracking persistence with mmWave compared to existing solutions
- A combined tracking and sensing architecture model
- A model for classifying key environmental points through mmWave tracking data
- A foundational step towards a general approach to incorporating environmental characteristics into mmWave

Fig. 1. Block diagram of overall stages involved in the RDTP MOT architecture.

MOT

II. RDTP MOT Architecture

The architecture adopted for the RDTP MOT environmental characterization explored in this paper can be broken into 5 stages. The sequencing of the 5 stages can be seen in figure 1. The methodology and implementation of these stages will be explored in section III of this paper. The remainder of this section of the paper will describe the purpose of each of the stages depicted in figure 1 in relation to the problem statement.

A. Problem Statement

Whilst performing MOT, environmental characteristics are not known or cumulatively derived. As a result, MOT cannot reliably be maintained in environments where non-penetrable objects are causing disturbances in persisted tracks. Prior to enhancing MOT with derived environmental characteristics, an approach to determine environmental characteristics is required.

B. Proposed Framework

The block diagram in figure 1 illustrates the proposed architecture to establish an approach to derive environmental characteristics from MOT trajectory data. The *Trajectory Collection* stage illustrated in figure 1 encapsulates the process of performing MOT with the mmWave radar. The output of this stage is vector of tracked objects in the form of trajectories. The *Trajectory Collection* stage does not involve any specific alterations to perform MOT due to the environmental characterization that is performed.

The stages following *Trajectory Collection* all specifically attribute towards deriving environmental characterization of the MOT field of view. The *Pre-processing and Normalization* stage is required to massage the data in preparation before parsing it for trajectory patterns. This stage involves transforming the vector of tracked objects returned from the *Trajectory Collection* stage into an vector of regional trajectories. In addition, as a CNN is used for *Trajectory Analysis*, the input shape must be a fixed length vector. Therefore, a normalization process is undertaken as part of the *Pre-processing and Normalization* stage to re-shape the regional trajectory vector.

The *Trajectory Analysis* stage consumes the normalized regional trajectory vector and applies a CNN against the dataset to classify RDTP. The classification criterion is based on a taxonomy of generalised trajectory patterns that have been pre-defined. The taxonomy defined has been chosen to extract entry and exit points in the MOT field of view. The output of the *Trajectory Analysis* stage is an associated dominant trajectory pattern for each region in the field of view.

The *Entry & Exit Association* stage involves the correlation and grouping of regions that have been tagged as entry and exit points within the field of view. The significance of this correlation is to identify bounding boxes around non-penetrable environment features that the tracked object can pass behind/through, such as columns. In addition to identifying non-penetrable objects, the *Entry & Exit Association* stage is used to identify true entrances and exits to the MOT field of view, such as doorways. The associations identified are passed through the *Projection* stage to present the environment characteristics on the same plot as the MOT data.

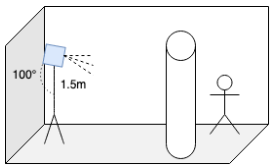

Fig. 2. Radar installation position.

III. Methodology and Implementation

To implement the RDTP MOT architecture, a Texas Instrument (TI) IWR6843 mmWave sensor was used. The mmWave sensor was installed on a TI MMWAVEICBOOST evaluation board, in which raw ADC data from the sensor was streamed via a TI DCA1000EVM. The sensor was mounted on a tripod and positioned approximately 1.5m above ground. The sensor was also tilted downward at a 10°angle to achieve the best field of view. This position is in accordance with the recommendation provided by Texas Instruments. Figure 2 illustrates the positioning of the sensor in relation to the external environment. The remainder of this section describes the methodology adopted for each of the RDTP MOT stages, seen in figure 1.

A. MOT Trajectory Collection

The 'chirp' is transmitted from the mmWave radar, in which an intermediate frequency (IF) signal is calculated from. The IF signal is calculated by determining the difference in the reflected signal frequency between the transmitter and receiver pair.

Using the IF and the frequency slope S, the range from the radar in which an object was detected can be computed with:

$$d = \frac{f_{IF}c}{2S} \tag{1}$$

where f_{IF} is the frequency of the IF signal and c is the speed of light $3 \times 10^8 m/s$. The FFT of the IF signal is calculated to detect objects residing at different ranges.

The velocity of detected objects is then calculated by analysing the phase shift in the IF signal. Two chirps are transmitted with T_c time between transmission. The velocity of the object can be calculated as:

$$v = \frac{\lambda w}{4\pi T_c} \tag{2}$$

where λ is the wave length and w is the phase difference.

The angle of arrival of the signal can be estimated by averaging the phase difference in each of the transmitter

receiver pairs. The phase difference θ for a single pair can be calculated as:

$$\theta = \sin^{-}1(\frac{\lambda w}{2\pi d}) \tag{3}$$

where d is the range of a detected object, calculated using (1).

Using (1), (2) and (3) a point cloud graph is constructed, representing the reflected signals that correlate to detected objects. Reflected signals that are present in the previous frame are deemed as static object and are removed at this stage.

Following the construction of the point cloud graph, point cloud clusters are determined using a modified DBScan optimized for cluttered environments, as presented in [6]. The clusters formed with DBScan at F_n are associated with clusters formed at F_{n-1} using the Hungarian Algorithm, where F_n is the current frame and F_{n-1} is the previous frame. Finally, a Kalman filter is used to predict and correct the associated object track.

A single tracked object i is represented as the tuple:

$$TO_i = (p, v) \tag{4}$$

where p is the two dimensional vector $[x, y]$ representing the current coordinates of the tracked object and v is the current velocity of the tracked object.

For a single given frame j, the set of tracked objects present in the frame can be illustrated as:

$$FTO_j = \{TO_1, TO_2, ..., TO_n\} \tag{5}$$

where TO is a tuple in the form of (4) and n is the total number of tracked objects in the given frame j.

For each sampled frame, a set of tracked object is persisted. This is expressed as:

$$TTO = \{FTO_1, FTO_2, ..., FTO_n\} \tag{6}$$

where FTO represents a set of tracked objects that are present in a single frame and n is the total number of frames.

B. Pre-processing and Normalization

To prepare for *RDTP Analysis*, a grid $G = (i \times j)$ must be constructed for a field of view of size $(l \times w)$. The terms l and w are expressed in meters, whilst i and j are the number of cells that l and w should be split into. Regions are defined as a collection of adjacent cells from G that form $R_x = (m \times n)$, where x is the region number. Region dimension m and n must both be factors of i and j respectively. Regions must not overlap with each other and must fit into G perfectly.

For each region R_x, an activity heatmap ofR_x is required to perform $RDTPAnalysis$. In order to construct a heatmap for each R_x, the vector represented in (6) will need to be transformed so that the tracked objects TO are grouped into their respective region. A given TO is correlated with an R_x through the coordinates it occupies, present in (4). The related FTO that the tracked object took place at must not be lost when transforming the vector, this information will be required when performing *Environmental Association and Bounding*.

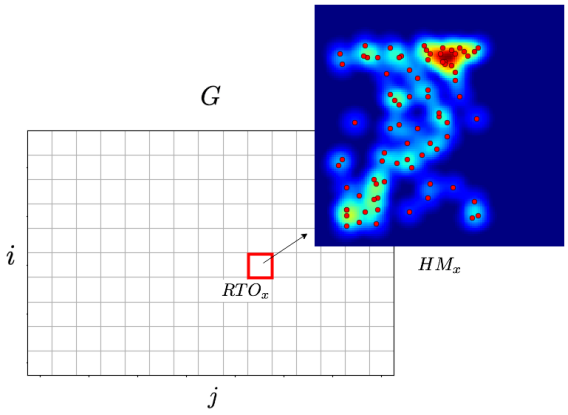

Fig. 3. Visual representation of preprocessing and normalization of the tracking data.

The set of regional tracked objects is expressed as a collection of tuples:

$$RTO_x = \{(F, TO)_1, (F, TO)_2, ..., (F, TO)_n\} \tag{7}$$

where F the frame the coupled TO occurred at, n is the total number of tracked objects and x is the region index.

The activity heatmap image can then be constructed for each RTO_x expressed by (7). Each TO equates to activity in a region, activity is positioned at the coordinates of the respective TO. The activity heatmap is illustrated as a fixed sized image, stored in a set expressed as:

$$RHM = \{HM_1, HM_2, ..., HM_n\} \tag{8}$$

where HM is the activity heatmap constructed for the respective RTO_x expressed in (7) and n is the total number of regions in G. The tracked regional tracked objects RTO_x are ultimately normalised through their illustration as an activity heatmap.

The transformations and stages in which the data undertakes throughout the pre-processing and normalization phase has been illustrated in figure 3. Figure 3 aims to illustrate a field of view of size $l \times w$ divided into a grid G of size $i \times j$. The tracked objects are respectively arranged in G as per the transformations described previously in this section, so that for each RTO_x a respective heatmap HM_x can be defined and collated as per (8).

1) Heatmap Generation: In order to generate the heatmaps expressed in (8), the region cell R_x, must first be constructed. After constructing the region cell, the tracked object coordinates can be superimposed onto the region so that intensity estimation can be performed. To calculate the intensity of the region, a Kernel Density Estimation algorithm is implemented [7]. For a given tracked object coordinate (x, y), expressed as TO_x, within the region R_x, the density of the coordinate is expressed as:

$$D(TO_x) = \sum_{i=1}^{N} K(TO_x - RP_i; h) \tag{9}$$

where K is the kernel function adopted for the group of points $RP_i; i = 1 \cdots N$ within the region R_x and h is the bandwidth for K. A number of kernel shapes were experimented with, in which it was discovered that a Quartic kernel, based on the one first presented by [8], provided the best resolution of local structures in the tracked object coordinates. Hence, the Quartic kernel function that we applied to 9 is expressed as:

$$K(u;h) = \begin{cases} \frac{15}{16}(1-(\frac{u}{h})^2)^2, & u \leq h, \\ 0, & u > h, \end{cases} \tag{10}$$

The following algorithm illustrates the functions that were derived to yield the heatmaps in accordance to the density estimation described in 9 and 10.

```
def quartic_kernel(d, h):
  # Determine bandwidth constraints
  if d <= h:
      # Evaluate Quartic kernel equation
      return (15/16) * (1 - (d/h)**2)**2
  else:
      # Distance between points is beyond
      # the bandwidth constraint
      return 0

def distance(a, b):
  # Calculate distance between two points
  d2 = ((a[0] - b[0])**2) + \
       ((a[1] - b[1])**2)
  return math.sqrt(d2)

def eval_region_density(grid_centers,
                        points, h):
  density = []
  # Loop through all cells in the region
  for c in range(len(grid_centers)):
      col_density = []
      for r in range(
        len(grid_centers[c])):
          den_sum = 0
          for p in points:
              # Determine point distance
              # from center
              gc = grid_centers[c][r]
              d = distance(gc, p)
              # Estimate density
              den_sum += quartic_kernel(
                           d, h
                         )
          col_density.append(density_sum)
      density.append(col_density)
  return density
```

C. RDTP Analysis

The purpose of the *RDTP Analysis* is to extract entry and exit points from the regional trajectories formed by the tracked objects. In order to do so, an idealistic taxonomy of regional patterns that expose entry and exits should be pre-defined.

Fig. 4. Trajectory taxonomy.

Figure 4 defines the taxonomy utilised to classify entry and exit trajectories. The trajectory used to classify an entry movement pattern is illustrated in figure 4 (a), whilst figure 4 (b) demonstrates the trajectory used to classify an exit movement pattern.

A single classifier is used to determine the probability that a given activity heatmap, expressed in the set illustrated in (8), is one of the movement patterns expressed in figure 4. The classifier is ultimately based on the handwritten digit recognition network presented in [9].

Figure 5 illustrates the design of the network that was implemented to classify entry and exit trajectories of the mmWave MOT field of view. The input shape of the network is a 32×32 pixel activity heatmap. A convolutional and maximum pooling layer is added to extract features from the activity heatmap. The convolutional layer consists of 16 filters of size 5×5. Another convolutional and maximum pooling layer is added to the network with 32 filters of size 3×3. Finally, a softmax layer is added of size 3×1 is added. The 3 softmax classes (C_1, C_2 and C_3) are the two movement patterns defined in figure 4, respectfully C_1 and C_2, and the probability that neither C_1 or C_2 are accurate, respectfully C_3.

1) Loss Function: The loss function utilised to correct weights is a categorical cross-entropy loss function. As such, the loss function implemented is as follows:

$$Error = -\sum_{i=1}^{3} y_i \cdot \log \hat{y}_i \tag{11}$$

where y_i is the target value and $\hat{y}_i$ is the respective scalar in the output of the softmax layer of the proposed model. The limit of the summation in 11 is set to 3 due to the fact that there are 3 possible classifications the model output may be.

A categorical cross-entropy loss function was selected for optimization of this model primarily due to the fact that the model is a classifier by nature [10]. Furthermore, the categorical target features of this model are encoded as a one-hot vector. Table I illustrates the one-hot encoding that was used for the target features of the model output. Due to the model output target features being encoded as a one-hot vector, a cross-entropy loss function, specifically categorical, is most appropriate for optimization [11].

D. Environmental Association and Bounding

The classified regional activity heatmaps are organised into those that are deemed either an entry or exit trajectory pattern and those that are not. The regional activity heatmaps that

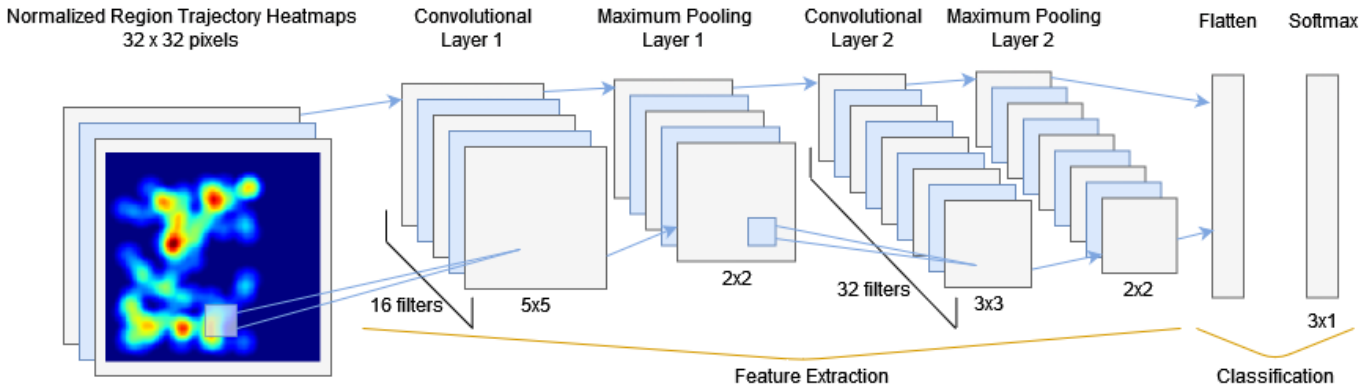

Fig. 5. RDTP network layer design.

Target Feature	Encoding		
Entry	1	0	0
Exit	0	1	0
Neither	0	0	1

TABLE I
TARGET FEATURE ONE-HOT ENCODING MODEL OUTPUT VECTOR.

Scene	# 'Entry' Collected	# 'Exit' Collected
Bedroom	96	127
Kitchen	102	110
Hallway	134	109
Living Room	104	95
Office	126	97

TABLE II
SUMMARY OF DATA POINTS COLLECTED ACROSS 5 DIFFERENT ENVIRONMENT SETTINGS.

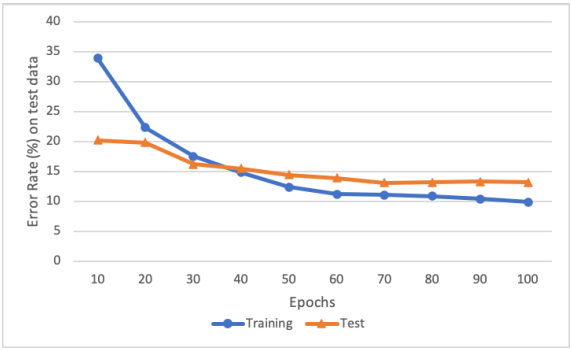

Fig. 6. Training and test error rate percentage across epochs.

are not an entry or exit trajectory pattern are negated going forward. The remaining regional activity maps are correlated back to the individual tracked objects that constitute the region, illustrated in (7). In this state the regions that have been classified as entry and exit points of the field of view are projected onto the MOT plane and rectangularly bounded.

IV. Experimental Results and Analysis

The methodology and implementation discussed in this paper was trained and tested using real data. A data-set of 1100 entry and exit events were collected using the MOT trajectory collection process described in the previous section of this paper. These events were collected across 5 different environments. Table II summarises the data points collected across the various scenes.

The data-set was then pre-processed and transformed into a collection of activity heatmaps, as described in the previous section of this paper. The activity heatmaps were then manually tagged for identification of entry and exit events. The training set of data occupied 80% of the tagged dataset. The remaining 30% was reserved for testing. After performing 65 epochs, taking 102.39 seconds to complete training, the average accuracy of the network was 87.18%. Figure 6 demonstrates the change in error over the number of iterations performed. A value of 65 epochs was deemed appropriate due to the notable convergence evident in figure 6.

The classified entry and exit points are projected onto the MOT plane, demonstrated in figure 7 (a) and (b). The projected entry and exit points in figure 7 (b) can be compared to the image of the observed environment illustrated in figure 7 (c). During MOT in this environment, at most two individuals were present in the field of view. The individuals independently walked around in the field of view, as well as leaving and re-entering through the door on the left side of the room and ducking behind the couch in the center of the room. It is evident that the trained network successfully classified these 4 entry and exit events through the projects presented in figure 7 (b).

V. Conclusion and Future Work

This paper demonstrates an approach to extract entry and exit environment characteristics using trajectories obtained from mmWave MOT. Through the classification of regional activity heatmaps, the probability that a region is either an entry or exit point in the environment can be determined. The research conducted in this only operates in an offline context of MOT data. Additionally, the system presented operates under a basic assumption that no entry or exit point will co-exist in the the one region. This assumption of the system is indirectly

(a)

(b)

(c)

Fig. 7. MOT Plot with projections and scene.

present due to the process undertaken to normalize trajectory vectors for a given region.

Future research would include enhancing the model to perform environment characterization and projection in real-time with MOT. Furthermore, it would be valuable to reevaluate the approach taken to illustrate the vector of a heatmap, so that the system is capable of operating on entry and exit points that overlap regions. It is also of interest to assess the potential to feedback the environment characterizations as a parameter that is considered when performing object tracking prediction and correction. The inclusion of this research in a real-time context opens opportunities for many potential enhancements while performing MOT.

REFERENCES

[1] P. Zhao, C. X. Lu, J. Wang, C. Chen, W. Wang, N. Trigoni, and A. Markham, "mid: Tracking and identifying people with millimeter wave radar," in *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, 2019, pp. 33–40.

[2] M. Z. Ikram and M. Ali, "3-d object tracking in millimeter-wave radar for advanced driver assistance systems," in *2013 IEEE Global Conference on Signal and Information Processing*, 2013, pp. 723–726.

[3] Z. Zhong, J. Zhao, and C. Li, "Outdoor-to-indoor channel measurement and coverage analysis for 5g typical spectrums," *International Journal of Antennas and Propagation*, vol. 2019, pp. 1–10, 09 2019.

[4] C. Yang and G. Gidófalvi, "Detecting regional dominant movement patterns in trajectory data with a convolutional neural network," *International Journal of Geographical Information Science*, vol. 34, no. 5, pp. 996–1021, 2020. [Online]. Available: https://doi.org/10.1080/13658816.2019.1700510

[5] Z. Zhang, X. Zhao, Y. Zhang, J. Zhang, H. Nie, and Y. Lou, "Efficient mining of hotspot regional patterns with multi-semantic trajectories," *Big Data Research*, vol. 22, p. 100157, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2214579620300253

[6] T. Wagner, R. Feger, and A. Stelzer, "Modification of dbscan and application to range/doppler/doa measurements for pedestrian recognition with an automotive radar system," in *2015 European Radar Conference (EuRAD)*, 2015, pp. 269–272.

[7] C. M. Bishop, *Pattern Recognition and Machine Learning*, ser. Information Science and Statistics, M. Jordan, B. Schölkopf, and J. Kleinberg, Eds. Springer, 2006.

[8] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall, 1986.

[9] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[10] Z. Akata, F. Perronnin, Z. Harchaoui, and C. Schmid, "Good practice in large-scale learning for image classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 507–520, 2014.

[11] A. Demirkaya, J. Chen, and S. Oymak, "Exploring the role of loss functions in multiclass classification," in *2020 54th Annual Conference on Information Sciences and Systems (CISS)*, 2020, pp. 1–5.