Towards Safety in Open-field Agricultural Robotic Applications: A Method for Human Risk Assessment using Classifiers.

José C. Mayoral [1] Lars Grimstad [1] Pål J. From [1] Grzegorz Cielniak [2]

Abstract—**Tractors and heavy machinery have been used for decades to improve the quality and overall agriculture production. Moreover, agriculture is becoming a trend domain for robotics, and as a consequence, the efforts towards automatizing agricultural task increases year by year. However, for autonomous applications, accident prevention is of prior importance for warrantying human safety during operation in any scenario. This paper rephrases human safety as a classification problem using a custom distance criterion where each detected human gets a risk level classification. We propose the use of a neural network trained to detect and classify humans in the scene according to these criteria. The proposed approach learns from real-world data corresponding to an open-field scenario and is assessed with a custom risk assessment method.**

Index Terms—**Risk Assessment, Mobile Robotics, Agricultural Robots**

Fig. 1: Human Risk Assessment in Open-field scenario.

I. Introduction

The safety aspect for autonomous mobile robots with their integration into our daily lives points to life quality improvement in various industries and drives a new industrial revolution in several applications such as agriculture. Robotic systems are relatively new in agriculture, and the number of tasks increases year by year while researchers are finding new paradigms and situations on these novel applications. This specific agricultural utilization requires safety functionality and particularly ability to avoid trespassing human personal spaces. Solutions for enhancing safety in unstructured environmental tasks must consider situation awareness, especially when an autonomous robot is fitted with a hazardous tool and navigates across an environment shared with human beings.

Collisions of humans with machines represent over 70% of the total causalities in different agricultural tasks, and the agricultural machinery represents the second source of incidents in such environments [1]. Enhancing safety capabilities of robotic systems is a key point for real-time applications, especially in the decision-making process which plans the mobile robot's subsequent actions, after identifying obstacles or human in the proximity of the platform.

There is a high interest in safety for autonomous navigation systems attempting to perform a given task without causing any harm. Typical perception systems use cameras and other sensors to provide information about mobile robot's surroundings which is used by collision avoidance algorithms. Whilst the robot detects the objects and people, it manoeuvres around any obstacles and avoids collisions, mimicking the understanding of common sense and safety.

[1] Norwegian University of Life Sciences (NMBU) {jmayoral, lars.grimstad, pal.johan.from} @nmbu.no
[2] University of Lincoln (UoL) gcielniak@lincoln.ac.uk

Safety concerns eliminating or reducing harm to any person in every situation, being the autonomous device the greatest risk source for humans in agricultural applications. Traditionally, once the person is detected, its 3D pose coordinates are computed, followed by a tracking or motion prediction process, or in best cases, model checking or safety rules procedures. Then, a safety assessment component computes people and robot trajectories, exposing safety rules violations. In contrast, our method uses a one-step classification for people's risk states based on the distance to camera criteria, and it classifies human-hazard as a safety assessment modality.

This work focuses on human risk assessment in autonomous agricultural navigation proposing a methodology based on a risk evaluation process. Specifically, this paper demonstrates a method to classify a human risk according to its appearance from the camera perspective, matching human into risk classifications using classifiers, an example can be seen on Figure 1. The presented classification method uses distance as a fundamental criterion to define safety, classifying every person's safety into four categories: lethal, dangerous, warning, and safe, from the riskiest to safest state, respectively. A neural network is trained using this criteria high-passing the need of a pose calculation. Our contributions are:

- The integration of risk assessment into a robotic agricultural domain.
- A distance-based classification system developed to measure human risk in the context of autonomous mobile robots.
- A method to calculate the risk index over the human risk classification.

II. Related Work

In terms of autonomous navigation, a motion planning algorithm must avoid any action that can potentially harm a human to grant a safe motion [2], denoting safety as the state of the world where there is no unreasonable risk [3]. Typically, society defines the acceptable risk threshold according to the potential consequences, generally involving all relevant objects and intentions of their motions [4].

Safety should imply risk probability and task-related knowledge [5]. Even more, for agricultural tasks, any autonomous mobile robot must match the requirements declared by the standards for highly automated agricultural machines to execute safely, including safe functionalities as hazard-free zone, state transition, remote emergency stop, and unintended excursions [6].

When an autonomous agricultural mobile robot executes a task in open fields, it must be aware of humans who might share the environment with the robotic system, and computer vision systems highly contribute to optimizing the different processes related to agriculture, including fruit, crop, people and obstacle detection [7].

Indeed, a deep learning model can estimate safety indexes for risk assessment, by utilising historical data and expert knowledge indicators [8], proving a method that integrates risk assessment in a data-based solution and have started a bridge between artificial intelligence and risk assessment.

Even more, risk evaluation is an on-going research for mobile robot application since it becomes a challenging approach due to the high variance of environmental features found on real-life. Thriving this issue, Mayoral et.al. [9] defines an automatic risk assessment method using a LiDAR-3D in the context of agricultural application based on a depth limited sampling trajectory prediction.

In agricultural tasks, safety restriction must concentrate on the assurance and reliability of operator and the agricultural robot, and it must satisfy some requirements [10]. Likely any risk analysis must include performance's unexpected behaviors and damage avoidance either for people, infrastructure and crops. Robustness at software and hardware levels is essential for fault detection and diagnosis that together with physical safety devices like emergency buttons or switches integrates the system.

Safety implies environmental awareness and perception, being RGB cameras are the most common sensor used as data acquisition and sensory systems. In robotic applications, they are able to detect and handle safety aspects in the presence of humans activating one or several safety strategies which return the robot to a safe state. These tactics include safety-rated monitored stop, robot speed, and position control, and near-field vision system through the set of static or dynamic safety zones [11].

Although computer vision systems provide robust solutions for different perception tasks, collisions with non-detected environmental elements still represent a common risk for autonomous navigation tasks [12]. Safety monitored stop, hand guiding, speed monitoring, or power and force suppression strategies can address those situations [13] by implementing run-time collision avoidance methods for assuring assure safety in autonomous vehicles. Besides, behavioral strategies can adapt to local changes in the environment, enhancing the system to plan future events with updated information to minimize the risk of dangerous situations and preventing robots from causing damages [14, 15].

In addition, the research community proposes to use learning methods for teaching safety concepts to robots specific to each application enhancing the robot's capability to perform particular tasks in environments such as cities or open-fields [16]. Machine learning techniques provide a robust solution mainly because of their capability to attain complex situation awareness, and these methods boost robot abilities by learning from camera inputs about the real world [17]. Furthermore, these learning approaches can enhance a safety system by detecting the relevant objects in a scene by calculating relative speed and distance to the camera [18].

However, machine learning techniques often require manual and resource-demanding annotations for any new application. As a response to this problematic, the community develop automatic labeling processes that leads to faster and more accurate ground truth data generation, and their use show more reliable results than relying on pure real-world data only [19]. A review of the attempts for improving the neural networks using simulated data shows that experiments to train with a mixture between real-world and synthetic data achieve generally better results than if trained on real data alone [20]. An alternative solution comes when a ground truth automatic process is pushed into the system leading to fast and accurate ground truths, and in some cases its use shows more reliable results than pure real-world data [19].

Simulation are a natural tool for generating new data, and recently, there are some advances in the generation of realistic data. A large-scale human synthetic dataset is present on [21], together with a novel approach to generate geometric images taking as an input a single shot of a person. Further, Lambert et. al. [22] merge the labels of a different datasets, creating a more general datasets concealing the different objects properties using a customized process getting high performance in multiple domains.

III. Methodology

The idea of the proposed methodology is to combine a human perception system with the risk assessment, and further, classify individual human hazards according to a risk criteria. It squeezes the safety assessment pipeline using the distance to the camera as a recognizable feature, instead of the 3D poses, and maps detected humans into distance-based zones, which assesses risk for each person.

Ignoring the tool-related risk requirements implied in a robust robotic operation, we detect the following risks related to human presence, summarized on Table I. It provides a general analysis for this kind of scenarios emphasizing the need of assess risk of people which interacts in the proximity of the autonomous device.

As a remark, we consider for this that the exposure of a person rolling over is lower than a person exposure to be inside ring 0 zone since it is dependable of this meaning that before

TABLE I: Human Detection Risk Analysis. Correlation between a Human (H) and rings

ID	Definition	Sev.	Exp.	Task	Avoid.	Risk Level
01	Roll over	S3	E1	Slow	A3	High
				Fast	A3	High
02	H. Ring 0	S3	E2	Slow	A2	High
				Fast	A3	Very High
03	H. Ring 1	S2	E2	Slow	A1	Medium
				Fast	A2	High
04	H. Ring 2	S1	E2	Slow	A1	Low
				Fast	A1	Low
05	H. Ring 3	S1	E2	Slow	A1	Low
				Fast	A1	Low

the person is rolled over it must be on the located in the ring 0. In the scope of this project, and taking into account the previous risk analysis, this approach focuses in the high ad very high risks of the given analysis to reduce risk during agricultural applications.

Our criteria sorts detected people into four categories which stand for qualitative risk related to the robot, based on the distance to the camera, implying the closer a person gets to the robot, the higher the risk is. Up to the application, a range of interest is selected and divided equally into four equidistant zones, labeled from R0 to R3, raising a distance-based classification method used for labeling.

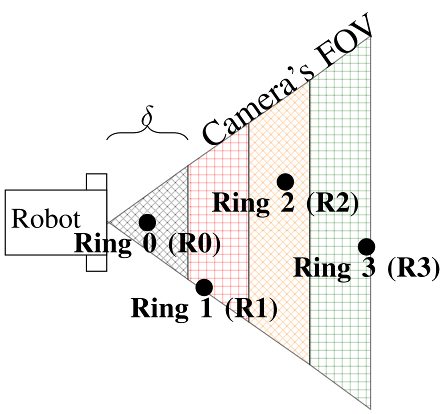

Fig. 2: The equidistant risk zones in the robot's sensory FOV used as human-hazard classification criteria.

The selected 1-dimensional criteria could demonstrate if classification can sort appearance-related features affected by the proximity using RGB cameras. Rather than the traditional classification approach, our method requires additional information for labeling, not just the human detection but the distance to the camera as well, limiting availability of the existing public datasets, particularly in open-field scenarios.

Moreover, a risk analysis is a fundamental requirement for ensuring risk evaluation and reduction in any task where autonomous devices are involved on. For some agricultural applications like seeding, harvesting, collection are executed on open-field spaces where the absence of environmental features may be constant along different fields, the complexity for isolating possible events that could affect the robot performance might not be high during daylight operations.

As qualitative evaluation criteria, a four-step misclassification severity level associates the highest risk to the persons laying on the lethal zone who are classified as in the safe region, which potentially causes harm or injuries, while sets low misclassification error to the safe persons classified as lethal considering this classification error does not generate any harm to any person. Figure 3 demonstrates the concept of the proposed misclassification rank.

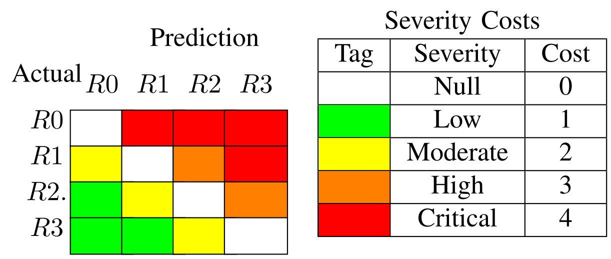

Fig. 3: Misclassification Risk Proposed Method.

The mentioned method is included as an additional component of the loss function, to minimize the critical errors, training our data prior within risk awareness knowledge. The calculations takes into consideration the confidence of the detection (p(i)) times an standardized severity cost, which indexes belongs to the ground truth and the classification output of the network, in Equation 1.

$$RiskLoss = \frac{1}{n}\sum_{i=1}^{n} p(i) * S(GT(i), \hat{x}(i)) \tag{1}$$

Our proposed method is based on a risk assessment methodology. However, for this particular setup, it has some similarities with the confusion matrix commonly used to visualize misclassifications. Therefore, we propose it as the base of a risk index calculation taken as a base the confussion matrix expressed on percentages times the severity cost expressed on Figure 3.

$$RiskIndex = \sum_{i=1}^{4}\sum_{i=1}^{4} CF(i,j) * S(i,j) \tag{2}$$

IV. Approach

Our proposed pipeline consists of three stages, shown in Figure 4, starting from the dataset, followed by an automatic labeling process, and finishing with a classification stage.

Fig. 4: Human hazard data preparation pipeline from dataset generation to training step.

A. Dataset Generation

Four data collection sessions generate an open-field dataset's subset, and this consists of up to three persons moving randomly in front of a depth camera at open field environments. The dataset consists of 13481 RGB and depth registered image frame pairs varying on lighting conditions and recorded with different cameras. After passing the automatic labeling process, only the frames with detected persons or accurate pose calculations remain. Table II summarizes the dataset given the number of frames per collection session.

TABLE II: Open-field Data Collection Summary.

Session	Raw Frames	Used Frames	Annotations
1	3759	2875	4635
2	3589	2105	2101
3	2541	1617	3244
4	3592	3570	7230
	13481	10157	17210

B. Automatic Labeling Process

For labeling, a pre-trained neural network provides the bounding boxes, while the risk class comes from the 3D person coordinates provided in the camera's frame by the projection of the bounding boxes into the rectified depth image. Pairs of RGB and depth images are matched by the timestamps of the frames as a pre-labeling step, a frame pair is shown in Figure 5.

Fig. 5: Pair of RGB with its corresponding depth image.

For this paper, the d_x values for our criteria two meters for covering the depth camera range within four risk zones. The depth image plays a role in the labeling stage, but they are not used on the run-time execution, and a human supervised procedure ensures the data quality to either remove or correct false negatives and positives. Figure 6 illustrates the complete automatic labeling procedure.

V. Experiments

The principal objective of the this experiment is to evaluate the classifiers performance and calculate its risk index. However, we are aware that the shapes of a bounding boxes might be proportional to the distance to the camera. Therefore, we add benchmark solutions that classifies giving prior knowledge on where the people are.

Our approach addresses risk quantification on agricultural open-field environments particularly when leading with human avoidance and detection. As backbone, YOLOv4 [23] is used

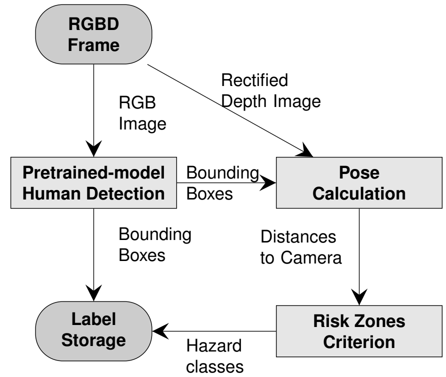

Fig. 6: Automating Labeling Process for RGBD images.

as classification network. Performance analysis between different classifier architectures is not one of our main objectives, so the training runs on a standard neural network architecture, one-stage YOLOv4, which is characterised by fast reference times, being a good choice for real-time implementations.

As benchmark, we use a selection of linear classifiers including Logistic Regression and Naive Bayes, assuming that the detected objects bounding boxes' sizes are proportional to the camera's distance, i.e., larger bounding boxes corresponding to closer obstacles. While the benchmark methods uses the bounding boxes coordinates as input, the network's input is the image itself. The objective is to evaluate how much the complexity of the network impacts the results. All the considered methods share the train and validation datasets.

TABLE III: Comparison of the selected Risk Classification methods

Method Name	Abbreviation	Input
Logistic Regression	L.R.	1D Array
Naive Bayes	N.B.	1D Array
Support Vector Classifier	SVC	1D Array
YOLO-v4-tiny	YOLOv4	Image

A. Results

Further, the automatic labeling approach for annotating the data provides the data's ground truth. Setting a pre-trained model as a bounding box generator provides a robust and fast method for splitting a person class into four risk zone classifications on detection.

Table IV displays the algorithm's overall performance compared against the three selected benchmarks, linear classifiers, and our proposal that uses a neural network. The neural network over-performs the benchmarks when talking about recall. In the other hand, L.R. and YOLOv4 get the same precision. N.B. shows the lesser results in terms of precision and L.R., the ones at recall. The metris for mAP are only availabe for the image-input based classifier.

Analysing the classification errors on the neural network approach at the open-field dataset (Table V), we observe that errors happens in consecutive risk zones. In the other

TABLE IV: Comparison of different Risk Classification methods

Method	Precision	Recall	mAP0.5	mAP.5-.95
L.R.	0.79	0.38		
N.B.	0.51	0.50		
SVC	0.78	0.69		
YOLOv4	0.79	0.76	0.78	0.61

hand, in terms of safety it can be observed that there are no misclassification problems between lethal and safe states which can produce the higher risks. This is consistent with the qualitative evaluation criteria (Figure 3) and our customized Risk Loss (Equation 1).

TABLE V: Normalized Confusion Matrix

Actual	Predictions			
	Lethal	Danger	Warning	Safe
Lethal	0.95			
Danger	0.05	0.66	0.03	0.05
Warning		0.34	0.94	0.55
Safe			0.03	0.45

The precision and recall metrics do not consider any classifications range or risk evaluation. Our method for this experiment can lead to multiple misclassification belonging to both consecutive risk zones. This phenomena can be seen in the relative high false positive range on our approach for the safety and warning classes. The risk index on Formula 2 provides a manner to assess the risk on our detector and benchmarks, giving the results in Table VI.

TABLE VI: Risk Index Calculation for the selected Risk Classification Methods

Method	Risk Index
L.R.	0.5075
N.B.	0.5900
SVC	0.515
YOLO	0.6450

As it can be observed in Table VI, our approach that uses neural network is riskier than the benchmark approaches that have prior knowledge of the location of the person. However, Table VII shows that the 90% of the risk in the approach comes from the Warning class that is consistent with the results in the related confusion matrix at Table V.

TABLE VII: Analysis of Risk Index per Class for YOLOv4 method

Class	Risk Index	Percentual
Lethal	0.0	0.0
Danger	0.0475	7.36
Warning	0.5825	90.31
Safe	0.0150	2.32

Figure 7 shows a sample of our most common classification error. The network struggles to classify persons who are further to the camera mixing the two risk zones with lesser risk, Warning and Safe.

(a) Ground truth: Danger (b) Prediction: Warning

Fig. 7: Misclassification Sample

We report true positives classifications that exemplify the classifier's output in Figure 8 where the detector can even find people lying on the grass according to the criteria.

Fig. 8: True Positive Classifications

VI. Conclusions

The presented approach uses the distance to camera as a classification feature providing a novel solution for integrating safety at autonomous application compacting human detection using as methodology the risk analysis of the applications. Overall, a neural network based approach performs better than linear classifier. Nonetheless, the input of these classifiers are the bounding boxes. In the other hand, despite these results our analysis show that the high misclassification rate between Warning and Safe increases significantly the risk index in the YOLOv4 method.

This approach can contribute to safety for other domains, extending the distance classification threshold as it is done here to match the data from the automotive dataset. Notwithstanding, we suggest that this safety functionality should be part of the mobile robots' sensor fusion vision systems designed to handle risk on agricultural applications.

A narrow safety classification system concept has proved its potential to implement functional safety action. Besides, the system must be enhanced with other unsafe situations to increase its robustness, not just pure human awareness. As a different approach, robot state information like speed could be added to the network on the final layers to generate a dynamic distance classification, mixing time-to-collision and appearance-based risk state estimation.

References

[1] Juan P. Vasconez, George A. Kantor, and Fernando A. Auat Cheein. "Human–robot interaction in agriculture: A survey and current challenges". In: *Biosystems Engineering* 179 (2019), pp. 35–48. ISSN: 1537-5110.

[2] Joseph C. Musto and L.K. Lauderbaugh. "A Safety-Based Decision Architecture for Autonomous Making Systems". In: (1991).

[3] *Road vehicles – Functional safety*. Standard. International Organization for Standardization, 2018.

[4] Andreas Reschka. "Safety Concept for Autonomous Vehicles". In: *Autonomous Driving: Technical, Legal and Social Aspects*. 2016, pp. 1–706. ISBN: 9783662488478.

[5] Terje Aven and Bodil S. Krohn. "A new perspective on how to understand, assess and manage risk and the unforeseen". In: *Reliability Engineering & System Safety* 121 (2014), pp. 1–10. ISSN: 0951-8320.

[6] *Agricultural machinery and tractors – Safety of highly automated agricultural machines – Principles for design*. Standard. Geneva, CH: International Organization for Standardization, Nov. 2018.

[7] Redmond Ramin Shamshiri et al. "Research and development in agricultural robotics: A perspective of digital farming". In: *International Journal of Agricultural and Biological Engineering* 11.4 (2018), pp. 1–14. ISSN: 19346352.

[8] Luca Gualtieri, Erwin Rauch, and Renato Vidoni. "Emerging research fields in safety and ergonomics in industrial collaborative robotics: A systematic literature review". In: *Robotics and Computer-Integrated Manufacturing* 67 (2021), p. 101998. ISSN: 0736-5845.

[9] José C Mayoral et al. *Integration of a Human-aware Risk-based Braking System into an Open-Field Mobile Robot*. 2021.

[10] Eleftheria Mitka. "Strategy for Safer Agricultural Robots". In: *Economics World* 6.6 (2018), pp. 472–481. ISSN: 23287144. DOI: 10.17265/2328-7144/2018.06.006.

[11] Roni-Jussi Halme et al. "Review of vision-based safety systems for human-robot collaboration". In: *Procedia CIRP* 72 (2018). 51st CIRP Conference on Manufacturing Systems, pp. 111–116. ISSN: 2212-8271.

[12] Lefteris Benos, Avital Bechar, and Dionysis Bochtis. "Safety and ergonomics in human-robot interactive agricultural operations". In: *Biosystems Engineering* 200 (2020), pp. 55–72. ISSN: 1537-5110.

[13] *ISO/TS 15066 Robots and robotic devices — Collaborative robots*. Standard. Geneva, CH: International Organization for Standardization, 2016.

[14] Roger Woodman et al. "Online Hazard Analysis for Autonomous Robots". In: *Towards Autonomous Robotic Systems*. Ed. by Roderich Groß et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 406–407.

[15] Mathilde Machin et al. "SMOF - A Safety MOnitoring Framework for Autonomous Systems To cite this version : HAL Id : hal-01394139 SMOF - A Safety MOnitoring Framework for Autonomous Systems". In: (2018).

[16] Eric Trautmann and Laura Ray. "Mobility characterization for autonomous mobile robots using machine learning". In: *Autonomous Robots* 30.4 (2011), pp. 369–383. ISSN: 09295593.

[17] Jahanzaib Shabbir and Tarique Anwer. "A Survey of Deep Learning Techniques for Mobile Robot Applications". In: *ArXiv* abs/1803.07608 (2018).

[18] Jagannath Aghav, Poorwa Hirve, and Mayura Nene. "Deep learning for Real Time Collision Detection and Avoidance". In: *Proceedings of International Conference on Communication Computing and Networking*. 2017.

[19] Marcelo C. Ghilardi et al. *Automatic Dataset Augmentation Using Virtual Human Simulation*. 2019. arXiv: 1905.00261 [cs.CV].

[20] Viktor Seib, Benjamin Lange, and Stefan Wirtz. *Mixing Real and Synthetic Data to Enhance Neural Network Training – A Review of Current Approaches*. 2020. arXiv: 2007.08781 [cs.CV].

[21] Albert Pumarola et al. "3DPeople: Modeling the Geometry of Dressed Humans". In: *International Conference in Computer Vision (ICCV)*. 2019.

[22] J. Lambert et al. "MSeg: A Composite Dataset for Multi-Domain Semantic Segmentation". In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020, pp. 2876–2885.

[23] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. "YOLOv4: Optimal Speed and Accuracy of Object Detection". In: *CoRR* abs/2004.10934 (2020). arXiv: 2004.10934. URL: https://arxiv.org/abs/2004.10934.