Danger Detection in Skiing using a Helmet-mounted 360-degree Camera

Yuki Fujita*[a], Alessandro Moro[b], Sarthak Pathak[a], and Kazunori Umeda[a]

[a]Department of Precision Engineering, Chuo University, Tokyo, Japan
[b]RITECS, Tokyo, Japan

ABSTRACT

Skiing is a winter sport widely enjoyed around the world, but it carries the risk of accidents due to collisions and falls. The objective of this research is to develop a system that uses a helmet-mounted 360-degree camera to detect other skiers in the vicinity and avoid the risk of collisions. The system utilizes the object detection algorithm YOLOv8 and the skeletal estimator ViTPose to make risk decisions based on the skier's position, direction, and posture. Experiments were conducted using data collected at an indoor ski resort to verify the accuracy of the system in detecting other skiers and detecting danger.

Keywords: 360-degree camera, Skiing, Danger detection

1. INTRODUCTION

Skiing is a popular winter sport enjoyed worldwide. However, while winter sports offer excitement and thrill, they also carry the risk of injury. According to an injury survey conducted by the National Ski Safety Measures Council at ski resorts during the 2022/23 season [1], a total of 225 people have lost their lives in winter sports accidents over the past 20 years. It is crucial to be aware of safety while skiing, and maintaining awareness of the surrounding environment is essential for ensuring safety on the slopes.

To reduce the risk of fatal accidents and injuries, many skiers choose to wear helmets. The survey also indicates that the rate of helmet usage has been increasing annually, reflecting a growing awareness of safety measures. However, despite this trend, the injury rate has not shown significant changes over the past decade. Therefore, it is necessary to explore additional measures beyond helmet use to prevent accidents.

Among the primary causes of injuries, self-inflicted falls account for a significant proportion. Beginners, in particular, are prone to losing their balance until they become accustomed to skiing. Thus, as individuals improve their skiing skills, the likelihood of sustaining injuries from self-inflicted falls decreases. On the other hand, collisions with other skiers are far more dangerous than self-inflicted falls and are difficult to avoid. To prevent collisions with other skiers, it is essential to remain aware of one's surroundings and ski at an appropriate speed and on a suitable course. Additionally, some self-inflicted falls occur as a result of attempting to avoid collisions with other skiers, causing a loss of balance and subsequent falls. Therefore, reducing the likelihood of collisions with other skiers and minimizing the number of such incidents is of great importance.

Furthermore, a study by Kogo et al. [2] highlights concerns and dangers perceived by skiers with hearing impairments, including the inability to notice other skiers approaching from blind spots, sensing their presence, or hearing warning calls. Since hearing-impaired skiers have difficulty recognizing when other skiers are approaching from behind, they are at a higher risk of collisions. Thus, a method for detecting skiers approaching from blind spots or behind and assessing potential dangers is needed.

In this study, we aim to develop a method to detect surrounding skiers and identify potential hazards.

*E-mail:fujita@sensor.mech.chuo-u.ac.jp

2. RELATED WORK

As an effort to prevent collisions, the development of helmets equipped with a LiDAR sensor has been explored. Evangelos et al. [3] proposed a system that detects skiers approaching from behind using three LiDAR sensors attached to the back of a helmet and alerts the wearer. To avoid interference, the three LiDAR sensors were positioned at different angles—left, center, and right—on the back of the helmet, covering a detection range of 5 meters. Additionally, three LED lights were mounted on the front of the helmet within the wearer's field of view. When a skier was detected behind the wearer using LiDAR, the corresponding LED light illuminated to indicate the direction of the detected skier. However, this system has a limitation: if the approaching skier is outside the LiDAR's field of view, the wearer cannot be alerted. Since skiing involves rapid movement on slopes, a limited detection field of view can create blind spots, which poses a potential issue.

Additionally, Yulin et al. [4] proposed a system that detects skier falls using object detection with YOLOv5 [5] and human pose estimation with OpenPose [6], combined with a convolutional neural network (CNN). By overlaying a color skeleton directly onto the original image, their method achieved high accuracy in detecting skier falls. However, in situations where skiers are densely packed, detection becomes challenging.

These existing methods have certain limitations: LiDAR may fail to detect skiers outside its field of view, and relying solely on skeletal key points to identify fallen skiers may not be sufficient for hazard detection. Therefore, this study aims to develop a system that utilizes an omnidirectional camera capable of capturing a wide field of view. To prevent collisions with other skiers, an omnidirectional camera will be mounted on a helmet, utilizing the captured video data to detect nearby skiers and assess potential hazards across a wide area. The hazard assessment will be conducted based on three factors: the position of other skiers relative to the wearer, their movement direction, and their posture.

3. PROPOSED METHOD

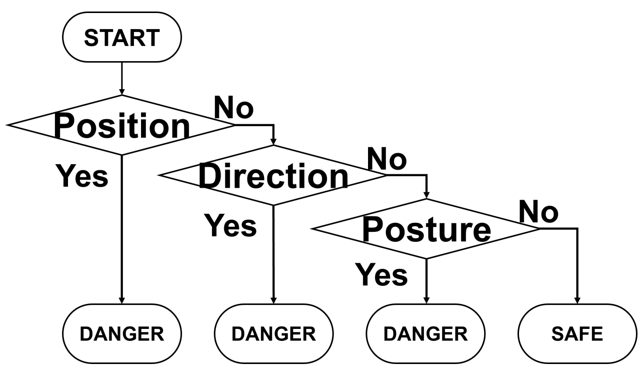

Figure 1. Danger judgment system overview

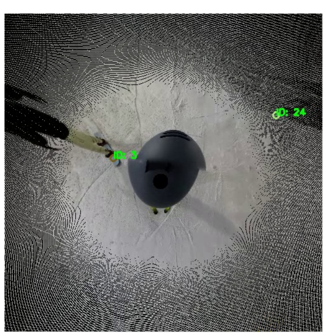

Figure 2. Bird's-eye view image with point cloud representation

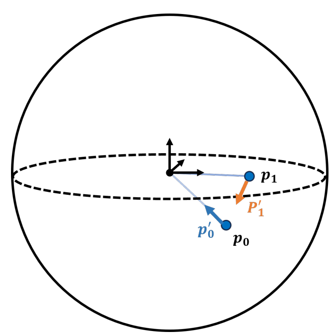

Figure 3. Determination of movement direction using the inner product

3.1 Overview of the Proposed Method

The process flow of the proposed method is shown in Fig. 1. The proposed method detects skiers using video images captured by a 360-degree camera and assesses danger based on three factors: position, direction, and posture. First, since the risk of collision is high when a skier is in close proximity to another skier, danger is assessed based on the skier's position relative to the helmet wearer. Second, if the skier is at a sufficient distance from the helmet wearer, danger is assessed based on whether the skier is approaching the helmet wearer. Finally, a skier who has lost control of their skis or has fallen may pose a collision risk and should be avoided; thus, danger is assessed based on posture. As described above, the system detects skiers in the captured video and comprehensively assesses danger by integrating three factors: the skier's position relative to the helmet wearer, their movement direction, and their posture.

3.2 Danger Judgment Based on Position

When skiing, it is essential to maintain a sufficient distance from other skiers. Since each skier may perform turns, the risk of collision increases if adequate spacing is not maintained. Therefore, this study focuses on the position of detected skiers relative to the helmet wearer.

Using an equirectangular image captured by an omnidirectional camera, where the horizontal and vertical axes represent the azimuth and elevation angles, the object detection model YOLOv8 [7] is employed to detect skiers. The system then extracts the bounding box (BBox) coordinates, specifically the top-left and bottom-right corner coordinates of each detected skier.

Additionally, the detected skiers are tracked using Norfair [8], a tool that integrates deep-learning-based object detection with object tracking across consecutive frames.

However, the equirectangular image captured by the omnidirectional camera contains distortions due to the camera's perspective, making it difficult to accurately determine the relative positions of skiers. To address this issue, the equirectangular image is transformed into a bird's-eye view image, as shown in Fig. 2. This transformation maps each pixel's coordinates from the camera's perspective to the corresponding ground coordinates.

First, each pixel's coordinates in the image are transformed into three-dimensional(3D) coordinates based on the camera's viewpoint. During this process, the camera's rotation angle and installation position are accounted for to determine the corresponding point in 3D space. Then, the intersection of these coordinates with the ground is computed and projected onto a two-dimensional(2D) plane, generating a bird's-eye view image. In this transformed image, each skier's position is represented in a ground-based coordinate system, making it easier to intuitively understand the distances between skiers.

Next, the system determines whether another skier is present within 5 m of the helmet wearer. The position of detected skiers is determined by tracking data and extracting the center coordinates of their BBoxes in the image. These center coordinates are used to compute the corresponding position on the ground.

First, the image coordinates are converted into 3D space. Then, a rotation transformation is applied based on the camera's position to determine the intersection of the camera's line of sight with the ground. This intersection is represented as (X, Y, Z). Assuming that the ground height remains constant, only the 2D coordinates (X, Y) are used to calculate the distance between skiers.

Taking the camera position as the origin, the distance d between the helmet wearer and the detected skier is computed using the following equation:

$$d = \sqrt{X^2 + Y^2} \tag{1}$$

Here, the height component Z is ignored, and only the horizontal distance is considered. If this distance is 5 m or less, the system classifies the situation as a potential collision risk due to the close proximity of skiers.

3.3 Danger Judgment Based on Direction

A skier may approach the helmet wearer at high speed from beyond 5 m. When a skier rapidly moves toward the helmet wearer, solely relying on position-based danger judgment, as described in Section 3.2, is insufficient for providing an adequate warning. Furthermore, even if a skier is moving at high speed, the risk of collision remains low unless they are heading toward the helmet wearer. Therefore, this method focuses on the movement direction of the detected skier. If the skier's movement direction is not directed toward the helmet wearer, the system classifies it as "Safe." Conversely, if the skier is heading toward the helmet wearer, it is classified as a "Potential Collision."

The procedure for danger assessment based on movement direction is described as follows. The position vector $\boldsymbol{p}_0$ of the detected skier in the current frame represents a vector pointing from the camera (origin) toward the skier. The normalized position vector $\boldsymbol{p'}$, pointing from the skier back to the camera, is calculated using the following equation:

$$\boldsymbol{p'_0} = -\frac{\boldsymbol{p}_0}{|\boldsymbol{p}_0|} \tag{2}$$

Similarly, the normalized direction vector $\boldsymbol{P'}_1$, which represents the skier's moving direction, is given by:

$$\boldsymbol{P'_1} = \frac{\boldsymbol{P_1}}{|\boldsymbol{P_1}|} \tag{3}$$

The inner product l of the position vector $\boldsymbol{p'}_0$ and the direction vector $\boldsymbol{P'}_1$ is computed as follows:

$$l = \boldsymbol{p'_0} \cdot \boldsymbol{P'_1} \tag{4}$$

This calculation determines if the detected skier is moving toward the camera. Fig. 3 illustrates the judgment of moving direction based on the inner product of the position and direction vectors.

If the inner product l is positive, it indicates that the skier is moving toward the camera. If l is negative, the skier is moving away from the camera. This concludes the danger assessment method based on the movement direction of the detected skier.

3.4 Danger Judgment Based on Posture

Ski resorts accommodate not only skiers who glide down the slopes but also those performing tricks and beginners unfamiliar with skiing. These skiers may lose balance, struggle to control their skis, or even continue sliding after falling. To prevent collisions, it is crucial to detect potential dangers early and maintain a safe distance. Furthermore, when a skier falls, their movement direction may change, making it difficult to assess risk based solely on their position and direction. Therefore, this study focuses on the posture of detected skiers.

Using video footage obtained from an omnidirectional camera, the proposed method employs ViTPose [9], a skeleton estimation model, to collect skeletal key point information of both stable skiers and those who have lost their balance. The collected skeletal key point data is labeled as either "Safe" or "Danger," and a neural network is trained to classify risk levels based on this input using deep learning.

First, ViTPose [9] is applied to a dataset to extract skeletal key point information (coordinates) of detected skiers. A file is generated to store the extracted skeletal key points along with the frame ID and tracking ID of each detected skier. Additionally, another file is generated to label each frame in the rectilinear cylindrical projection video as either "Safe" or "Danger."

Next, a neural network is constructed for training and danger detection. The proposed method adopts a one-dimensional convolutional neural network (1D CNN) model to classify detected skiers into two categories: "Safe" or "Danger." The input layer consists of the skeletal key point information extracted from the first file, while the output layer consists of two classes: "Safe" and "Danger." The intermediate layers comprise convolutional layers, pooling layers, fully connected layers, and activation functions. The first convolutional layer has an input channel size of 1, an output channel size of 64, a kernel size of 3, and a stride of 1. The pooling layer has a kernel size and stride of 2. The second convolutional layer has an input channel size of 64, an output channel size of 128, a kernel size of 3, and a stride of 1. The first fully connected layer converts the output from the convolutional layers into a 512-dimensional feature vector, and the second fully connected layer maps this 512-dimensional feature vector into a 2-dimensional output for final classification. The activation function used is ReLU (Rectified Linear Unit) [10], a nonlinear function that enables the model to learn even for problems that are not linearly separable. By training the proposed network on skeletal key point data extracted using ViTPose [9], the system detects potential dangers using a CNN.

4. EXPERIMENTS

To verify the effectiveness of the proposed method in danger detection, experiments were conducted using datasets collected for each danger detection approach.

For this experiment, a GoPro MAX equipped with an IMU (Inertial Measurement Unit) was used as the omnidirectional camera. A SMITH MAZE helmet was used, with the GoPro MAX mounted on it for dataset collection and danger detection.

First, multiple video datasets were collected, and ground truth labels for danger were assigned. Then, using the datasets collected with the omnidirectional camera, experiments were conducted to evaluate whether the proposed danger detection methods based on position, direction, and posture (as described in Sections 3.2, 3.3, and 3.4) could accurately assess danger.

4.1 Skier Detection

To generate experimental data and construct a dataset, experiments were conducted at the indoor ski resort SNOVA Shin-Yokohama [11], where skiing was performed while wearing a helmet equipped with an omnidirectional camera. A total of 72 video clips, each approximately 10 seconds long, were recorded, capturing skiers and snowboarders approaching from behind the helmet wearer.

All approaching individuals were experienced skiers or snowboarders and were instructed to perform actions that might be perceived as dangerous by the helmet wearer, such as overtaking from behind or falling. A total of 810 images were extracted from the 72 recorded video clips to construct the dataset.

Of the 810 images in the dataset, 504 were allocated for training, 208 for validation, and 98 for testing. For training, the parameters were set as follows: epochs = 100, batch size = 12. Each of the 504 training images was labeled as either "skier" or "snowboarder", and skier detection and tracking were carried out accordingly.

4.2 Determination of Ground Truth

The collected video data was presented to 12 participants, including men and women in their 20s to 40s, to investigate which frames were perceived as dangerous.

Three types of videos were used in this study:

- **Video1**: A skier approaches from behind one at a time.
- **Video2**: A skier falls in the background.
- **Video3**: Multiple skiers continuously approach from behind.

Among the 12 participants, 6 were experienced skiers, while the remaining six were inexperienced. In this experiment, frames perceived as dangerous by at least five participants were labeled as such.

Table 1 presents the frames labeled as dangerous based on the survey results for each video.

Table 1. Survey Results of Dangerous Frames [frame]

Video1	210 - 270, 360 - 420, 600 - 690, 750 - 810
Video2	150 - 210, 240 - 330, 540 - 690
Video3	210 - 240, 300 - 390

4.3 Validation Experiment for Danger Detection Based on Position and Direction

Danger detection was conducted by combining the methods described in Section 3.2 and Section 3.3.

In this experiment, the threshold for the inner product in (4) was set at 0.7. If the inner product value was 0.7 or higher, the skier was considered to be moving toward the helmet wearer, posing a potential collision risk. Fig. 4 and Fig. 5 show the experimental results.

In Fig. 4, the detected skier is positioned more than 5 m away, with the direction vector oriented downslope. Additionally, the inner product of the position vector and direction vector is -0.59, indicating that the skier is moving away from the helmet wearer. In this case, both distance and movement direction were classified as safe, confirming the accuracy of the detection.

In Fig. 5, the detected skier is within 5 m, and the trajectory forms a curve toward the helmet wearer. The inner product of the position vector and direction vector is 0.76, indicating that the skier is approaching the helmet wearer. In this case, the distance was classified as dangerous, and the movement direction was classified as a potential collision risk, confirming the accuracy of the detection.

Fig. 6 compares the dangerous frames in Video1, based on ground truth and the danger detection results using both distance and movement direction.

The accuracy of distinguishing between safe and dangerous frames was 66.1% for Video1, 64.7% for Video2, and 66.7% for Video3 respectively.

For Video1 and Video2, the relatively high accuracy can be attributed to the fact that dangerous scenes, such as skiers approaching from behind or falling, occurred as isolated events. However, in Video3, errors occurred in the frames between the first skier passing and the second skier approaching, leading to a decrease in accuracy.

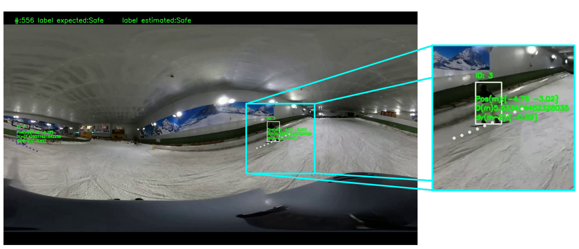

Figure 4. Danger detection based on direction: "Safe"

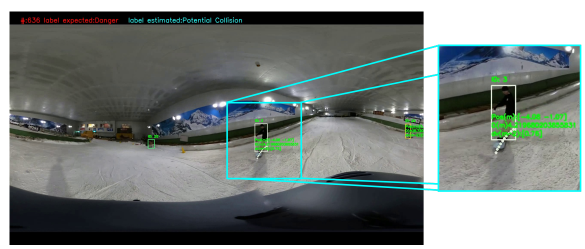

Figure 5. Danger detection based on direction: "Potential Collision"

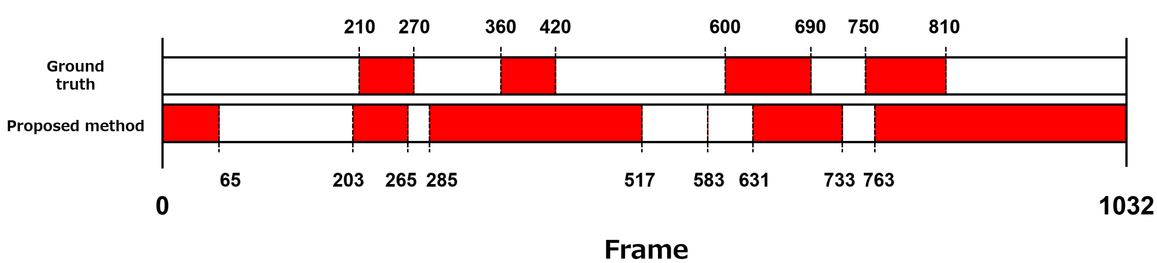

Figure 6. Comparison of Danger Detection Based on Distance and Direction in Video1

4.4 Validation Experiment for Danger Detection Based on Posture

Skeletal key point information of detected skiers was collected using the method described in Section 3.4. Fig. 7 shows an example of skeletal key points obtained using ViTPose [9].

As a result, ViTPose [9] successfully extracted skeletal key point coordinates from all frames. However, at this stage, the construction of a neural network and the validation experiment for posture-based danger detection using ViTPose's skeletal key point data have not yet been completed. In the future, a Recurrent Neural Network (RNN) [12], a type of neural network specialized for processing sequential and time-series data, will be employed. This will allow the model to retain past information and capture temporal dependencies between frames, enabling it to learn pose transitions and movements. Specifically, an RNN [12] will be used to process sequences of skeletal key point data and learn the transitions between poses in each frame. This will enable the system to recognize movement and posture patterns of detected skiers, predict future poses and movements in subsequent frames, and perform danger detection accordingly.

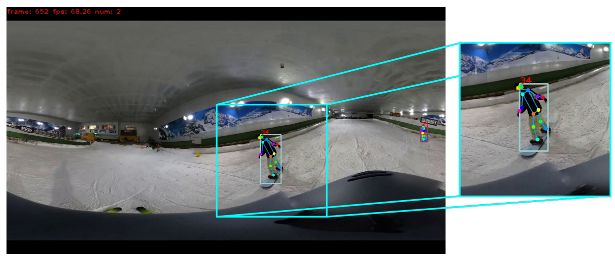

Figure 7. Extraction of skeletal key points using ViTPose

5. CONCLUSION

In this study, we proposed a skier danger detection method based on position, direction, and posture, utilizing video footage from an omnidirectional camera.

In the experiments, position-based and direction-based danger detection achieved an accuracy ranging from 64.7% to 66.7%.

For future work, we aim to integrate posture-based danger detection and enhance the accuracy of each detection method. For position-based danger detection, we plan to introduce adaptive thresholds that consider the slope inclination of ski resorts. For posture-based danger detection, we plan to develop a neural network that leverages skeletal key point information from skiers, enabling learning-based posture danger detection and accuracy assessment. Furthermore, we plan to implement a comprehensive danger detection system utilizing deep learning.

REFERENCES

[1] "Ski Resort Injury Report — Japan Cable Transportation Association," URL: https://www.nikokyo.or.jp/pages/38/. Accessed: Mar. 31, 2025.

[2] Y. Kogo and Y. Nakajima, "A Study on the Anxiety Experienced by Hearing-Impaired Individuals During Sports Activities," Adapted Physical Education and Sports Science Research, Vol. 4, No. 1, pp. 32–35, 2018.

[3] E. Niforatos *et al.*, "Augmenting Skiers' Peripheral Perception," in Proceedings of the 2017 ACM International Symposium on Wearable Computers, pp. 114–121, 2017.

[4] Y. Zhu and W. Q. Yan, "Ski Fall Detection from Digital Images Using Deep Learning," The 5th International Conference on Control and Computer Vision, pp. 70–78.

[5] J. Redmon, "YOLOv5," GitHub repository, URL: https://github.com/ultralytics/yolov5. Accessed: Mar. 31, 2025.

[6] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields," in Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1302–1310.

[7] Ultralytics, "YOLOv8: Real-Time Object Detection," GitHub, https://github.com/ultralytics/ultralytics. Accessed: Mar. 31, 2025.
[8] Norfair Developers, "Norfair: Lightweight Python Library for Real-Time 2D Object Tracking with Norfair," GitHub, https://github.com/tryolabs/norfair. Accessed: Mar. 31, 2025.
[9] Y. Xu *et al.*, "ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation," arXiv, 2022.
[10] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," in Proceedings of the 27th International Conference on Machine Learning (ICML'10).
[11] "SNOVA SHINYOKOHAMA," SNOVA Shin-Yokohama, URL: https://snovashinyoko.co.jp/page. Accessed: Accessed: Mar. 31, 2025.
[12] J. L. Elman, "Finding Structure in Time," Cognitive Science, vol. 14, no. 2, pp. 179–211, Mar. 1990.