



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공 학 석 사 학 위 논 문

# AR Grape Thinning Support



Shun Tamura

공 학 석 사 학 위 논 문

# AR Grape Thinning Support

지도교수 장 원 두

이 논문을 공학석사 학위논문으로 제출함.

2025년 2월

부 경 대 학 교 대 학 원

인 공 지 능 융 합 학 과

Shun Tamura

Shun Tamura의 공학석사  
학위논문을 인준함.

2025년 2월 21일

위 원 장      공학박사      김 훈 희 (인)  
위      원      공학박사      유 승 호 (인)  
위      원      컴퓨터이공학박사      장 원 두 (인)

# Contents

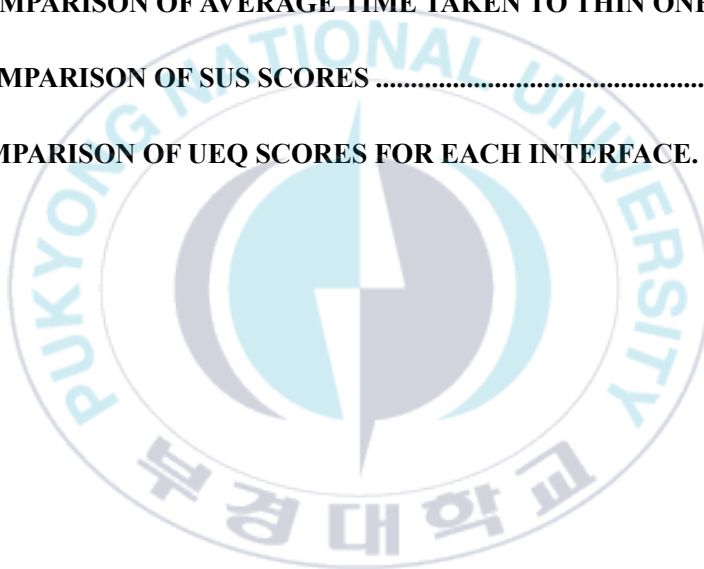
Abstract .....	vi
1. Introduction .....	- 1 -
2. Related Work .....	- 2 -
3. Proposed Method .....	- 3 -
A. Superimposing of berry contour .....	- 4 -
B. Judging the execution of thinning operation .....	- 6 -
C. Improvements of user interface .....	- 7 -
1) Displaying a frame around the entire grape cluster .....	- 7 -
2) Blinking contour lines .....	- 8 -
3) Adding blue background to the instruction image .....	- 9 -
4) Voice instructions .....	- 9 -
4. Experiments .....	- 11 -
A. Experimental purpose .....	- 11 -
B. Experimental Settings .....	- 11 -
C. Experimental procedure .....	- 12 -
D. Evaluation Indicators .....	- 13 -
E. Experimental Results .....	- 14 -
5. DISCUSSION .....	- 21 -

Average Time to Thin One Grape .....	- 21 -
Evaluation of Usability and User Experience .....	- 21 -
User Feedback.....	- 22 -
6. CONCLUSION.....	- 23 -
References.....	- 24 -



## Table

<b>Table I PARTICIPANT DEMOGRAPHICS.....</b>	<b>- 12 -</b>
<b>Table II AVERAGE TIME TAKEN TO THIN ONE BERRY. ....</b>	<b>- 14 -</b>
<b>Table III COMPARISON OF AVERAGE TIME TAKEN TO THIN ONE BERRY -</b>	<b>15 -</b>
<b>Table IV COMPARISON OF SUS SCORES .....</b>	<b>- 17 -</b>
<b>Table V COMPARISON OF UEQ SCORES FOR EACH INTERFACE. ....</b>	<b>- 19 -</b>



# Figure

Figure 1 Workflow of the AR-based grape thinning support system.....	- 4 -
Figure 2 Superimpose contour on target berry.....	- 6 -
Figure 3 Calculate 3d berry position from unit vector.....	- 6 -
Figure 4 Judged as in the process of thinning.....	- 7 -
Figure 5 Judged as thinning completed.....	- 7 -
Figure 6 Display of the frame of grape cluster.....	- 8 -
Figure 7 Blinking of the berry outline.....	- 8 -
Figure 8 Background for enhancing image visibility.....	- 9 -
Figure 9 Three interfaces compared in the experiment.....	- 11 -
Figure 10 Scores for each interface.....	- 16 -
Figure 11 UEQ results and benchmarks for each interface.....	- 18 -

AR Grape Thinning Support

Shun Tamura

부 경 대 학 교 대 학 원 인공지능융합학과



## Abstract

Recent advancements in deep neural networks (DNN) and augmented reality (AR) have improved the efficiency and automation of agriculture. This study proposes an AR grape thinning support system to assist in grape thinning operations. The proposed system uses DNN to predict grape berries that need to be thinned and uses the optical see-through Head-Mounted Display (HMD) HoloLens 2 to superimpose contour information over the real target berry, making it easy for users to identify the berry to be thinned. Additionally, the hand-tracking function of HoloLens 2 is utilized to monitor the thinning operation in real-time and provide voice instructions to improve work efficiency and user experience. The evaluation experiment compared three interfaces: "image only", "image with contour overlay", and "image with contour overlay and voice instructions", evaluating using the metric of time taken to thin one grape cluster and the usability and user experience. The results showed that the image with contour overlay and voice instructions could significantly improve usability.

# 1. Introduction

Recent advancements in deep neural networks (DNN) [1], [2], [3] and augmented reality (AR) [4], [5] have introduced new approaches for improving the efficiency and automation of agricultural tasks. Among these, the automation and optimization of grape thinning operations, which require considerable effort and skill, have attracted significant attention [6], [7], [11]. Grape thinning is a critical process that involves removing some grape berries so that the remaining ones can have sufficient space to grow larger and have less competition for nutrition to become of higher quality. Several studies have proposed methods utilizing DNN and AR to enhance thinning efficiency. For instance, Buayai et al. proposed an automatic grape counting system that combines DNN and AR [6], successfully improving the efficiency of thinning operations. Additionally, Buayai et al. developed a grape thinning support system using Microsoft HoloLens 2, capturing images of grapes and using DNN to predict the grapes to be thinned and present the information to users by displaying an instruction image on HoloLens 2 [7]. However, this system has the drawback of requiring users to compare the images with the actual grapes, making it difficult to accurately identify which berry be to thin. To address such issue, this study proposes a new grape thinning support system that directly display the contours of the berry to be removed on the berry itself. The system identifies the berries to be thinned from images captured using HoloLens 2 and superimposes the contour lines over their 3D positions, allowing users to easily identify the grapes to be thinned. Additionally, the hand-tracking function of HoloLens 2 is used to track the thinning operation in real-time and provide voice instructions to improve work efficiency and user experience. This paper introduces the algorithm for realizing the proposed grape thinning operation support system as well as the evaluation experiments for validate the effectiveness and usability of the proposed system in comparison with existing methods. Specifically, the aim is to reduce the time taken to thin one berry and enable users to perform the task more easily and accurately.

## 2. Related Work

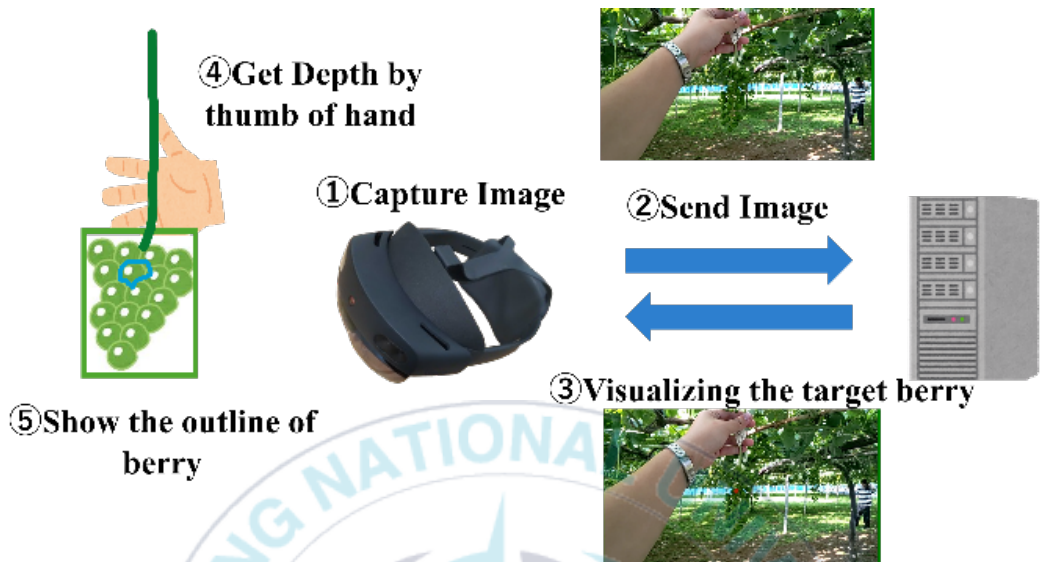
AR technology has been utilized to support various operations such as assembly of factory parts[8], construction and civil engineering tasks [9], and machine inspection [10]. In recent years, its application in agriculture has also attracted attention. Buayai's group proposed technology to support grape thinning and flower cluster pruning using AR[7], [11]. The system proposed in [11] uses the latest DNN models to detect grape clusters and identify the grape berries to be removed during operations and visualize the target berries using HMD. By using HMD, users can focus on the task without holding the device. However, their method visualizes the instruction by displaying an image of grape cluster in which the color of the berry to be removed is changed to red. It is very difficult for the user to identify the position of the target berry on the real cluster by referring to the instruction image. To address this issue, this study proposes an approach that superimposes contour lines directly on the actual grape berry, eliminating the need to compare the instruction image with the real cluster and allowing for more intuitive visualization.

### 3. Proposed Method

The grape thinning support system proposed in this study is implemented in a server-client approach as shown in Fig 1.

- ① Images are captured using the camera of HoloLens2.
- ② Sent to a server. And the server detects grape cluster and individual berries and calculates the probability of thinning for each grape berry using Buayai et al.'s method [7].
- ③ The thinning probabilities and mask information of each berry are then sent back to HoloLens2.
- ④ The depth to the berries is calculated using the position information of the thumb of the hand holding the grape cluster.
- ⑤ The parameters of the camera and the depth information to the grapes are used to calculate the 3D positions of the target berry, which is the one with the highest probability of thinning in current implementation. Then contour line of the target berry is generated use its mask information and superimposed on the berry using the 3D position information.

Additionally, the hand-tracking function of HoloLens2 is used to determine whether the thinning operation has been completed.



**Figure 1 Workflow of the AR-based grape thinning support system.**

#### **A. Superimposing of berry contour**

This study proposes to superimpose contours over the target grape berry to be thinned in 3D space, as shown in Fig 2. The steps are as follows: first, 2D mask images are obtained from the berry detection DNN model run on server. Next, the 2D coordinates on the image are converted to 3D coordinates in the world space using the parameters of camera and the depth information to the berries. The camera parameters at the moment the image was captured are obtained from HoloLens2. The depth to the berries is calculated using the position information of the thumb of the hand holding the grape cluster, which is obtained using the standard function of HoloLens 2's hand tracking. A plane P passing through the position of the thumb and is perpendicular to the forward direction of the camera is virtually placed. Then the distance from camera to P is estimated as the distance from the camera to the grape stem. This method is chosen because there is no standard way to access depth information from HoloLens2, and custom code can make HoloLens 2 unstable and prone to crashing. Additionally, the hand is larger and more stable than the grape berries for depth acquisition.

Next, we describe how to compute the coordinates of the target berry in 3D world space given its position on the captured 2D image. First, the method given in [12] is used to convert 2D coordinates on the image into the 3D coordinates in the world space to get the view direction vector  $V$ . Then  $D$ , the distance from the camera to the real bunch can be calculated as the length of line segment between the camera and the intersection of the line along  $V$  direction and plane  $P$ .

Since the hand is holding the stem on top of the grape cluster, the thumb should be a bit further from the camera than the target berry. This difference should depend on the radius of the grape cluster, which can be approximated with a small offset. In current implementation, we empirical set  $d=30\text{mm}$ . Then, by multiplying the unit vector  $V$  by the distance  $D$ , the vector  $V'$  from the camera to the target berry can be obtained, and by adding the world coordinates of the camera to  $V'$ , the world coordinates of the target berry can be obtained. This flow is shown in Fig 3.

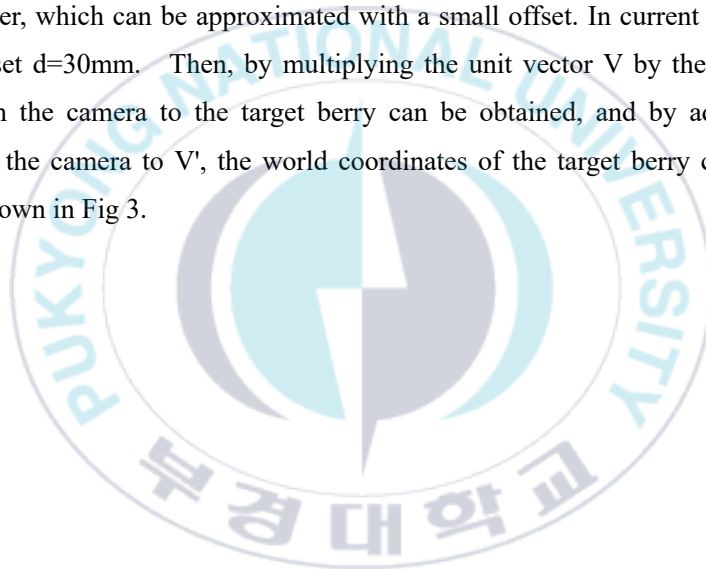




Figure 2 Superimpose contour on target berry.

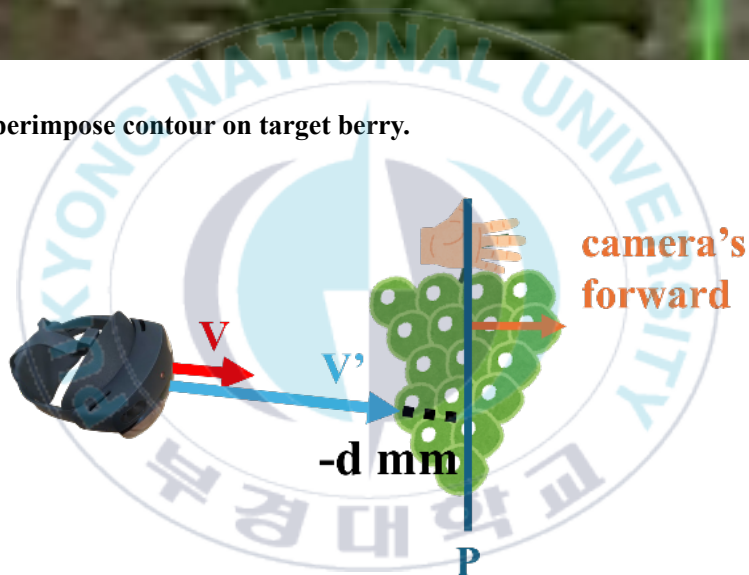


Figure 3 Calculate 3d berry position from unit vector.

### B. Judging the execution of thinning operation

This study uses the hand-tracking function of HoloLens 2 to determine whether thinning operations are in progress. Specifically, when the hand holding the scissors approaches the grape cluster, as shown in Fig 4, thinning operation is judged to be in progress. Currently, image capturing has stopped to avoid updating the image while the user is thinning. When the

hand holding the scissors moves away from the grape cluster, as shown in Fig 5, thinning is judged to be completed and the system starts detecting the next grape cluster. This approach allows the system to respond quickly to user's operations, making thinning operations more efficient.



**Figure 4 Judged as in the process of thinning**



**Figure 5 Judged as thinning completed.**

### **C. Improvements of user interface**

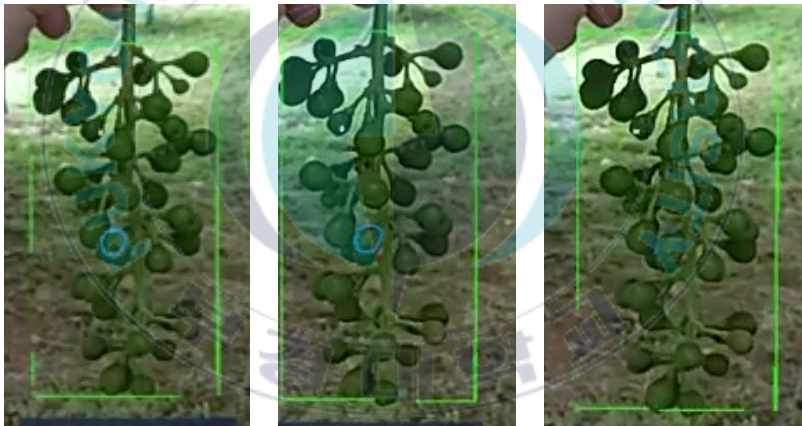
To make it easier for users to identify the target berry to be thinned, the following UI improvements were made:

#### **1) Displaying a frame around the entire grape cluster**

Instead of displaying contours on the target berry only, a frame is also displayed around the entire grape cluster, as shown in Fig 6. This helps users see the position shift clearly when they move the grape cluster and allows them to correct the position using their hands. Additionally, knowing the relative positions of the berries within the cluster helps users identify the berry to be thinned more quickly and accurately.



**Figure 6 Display of the frame of grape cluster.**



**Figure 7 Blinking of the berry outline.**

## 2) **Blinking contour lines**

Constantly displaying the contour lines can make the actual berries hard to see, making it difficult for users to identify them during thinning. Therefore, as shown in Fig 7, the contour lines blink to make the actual berries visible and help users accurately identify the berry to be thinned.

### 3) Adding blue background to the instruction image

In addition to superimpose the contour line over the target berry, an instruction image with the target berry indicated in red color is also displayed. By adding a blue background behind the image, as shown in Fig 8, the color contrast between the image and the background can be enhanced, making the image easier to see. This UI improvement allows users to quickly grasp the information in the image, improving work efficiency.



**Figure 8 Background for enhancing image visibility.**

### 4) Voice instructions

To enable users to effectively use the system, voice instructions are also provided allowing users to easily understand the next steps. The introduced voice instructions are as follows:

a) "Capture": This instruction notifying the user that the system is capturing an image so as to prompts the user to stop moving the grape cluster.

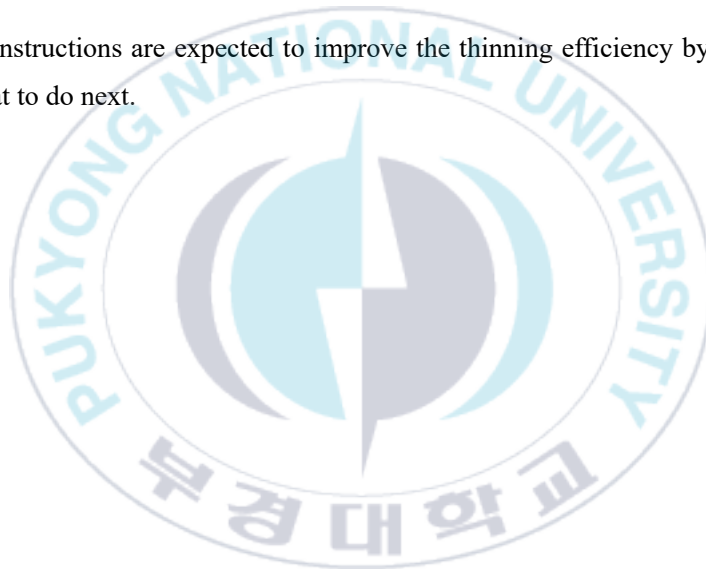
b) "Stop": This instruction tells the user to stop moving the grape cluster until the

server returns the results. It is important to keep the cluster still, as moving it can make the overlay inaccurate.

c) "Thin the berries": This instruction clearly tells the user what to do with the displayed instructions. Following the instructions allows for efficient thinning operations.

d) "Rotate": The current system determines the berries to be thinned from the 2D image, thus it doesn't consider the occluded berries. This instruction is introduced to prompt the user to rotate the grape cluster so as to thin the cluster from different view angle evenly..

These voice instructions are expected to improve the thinning efficiency by making it clear to the user what to do next.



## 4. Experiments

### A. Experimental purpose

To verify whether the proposed superimposing based target berry visualization as well as the user interface design can make thinning operations more efficient and improves user experience, experiments are conducted to compare the following three different interfaces shown in Fig 9.



Instruction image only (I).

Instruction image and  
contour overlay (I-O).

Instruction image, contour  
overlay and voice  
instructions(I-O-V).

**Figure 9 Three interfaces compared in the experiment.**

### B. Experimental Settings

The study involved six participants, each performing the thinning with all 3 interfaces. For each grape cluster, the thinning stops when the numbers of the berries in the cluster are reduced to 40. The order in which the participants performed the experiment was determined using a Latin square design to eliminate learning effects. The details of the participants is shown in Table I.

**Table I PARTICIPANT DEMOGRAPHICS.**

	sex	age	glasses	Hours of HMD Experience	Number of grape clusters thinned
1	M	21	yes	5	15
2	M	21	no	6	20
3	M	21	no	6	20
4	M	28	no	10	15
5	M	25	yes	720	100
6	F	29	yes	0	0

**C. Experimental procedure**

The experimental procedure conducted in this study is as follows:

Conduct a questionnaire: A questionnaire about the experience with HoloLens grape thinning work experience, age, and gender was conducted.

Instruction on HMD wearing and operating methods: Participants were instructed on how to wear and operate HoloLens 2. If there were no questions, they moved on to the next step.

UI experience and mock thinning operation: Participants experienced the system while the experimenter explained how to use it. Participants performed mock thinning operations using their hands. If there were no questions, they moved on to the next step.

Video recording: Thinning operations were recorded from the side using a video camera. The time taken and the number of berries thinned were recorded with the video.

Thinning operations: Participants started thinning operations upon the experimenter's signal and continued until the system judged that the number of berries in the cluster had reached 40.

Post-experiment questionnaire: After the experiment, the System Usability Scale (SUS)[13] and User Experience Questionnaire (UEQ)[14] were conducted. Subsequently, participants' free comments were recorded.

#### **D. Evaluation Indicators.**

The study uses the following indicators for evaluation:

1. Average time taken to thin one berry

The average time taken to thin one berry is computed as follows:

$$K = (T_{\text{total}} - T_{\text{c}} + T_{\text{l}}) / N_{\text{(thinned)}} \quad (1)$$

Here,  $T_{\text{total}}$  is, the total time used for thinning one cluster which was measured using the recorded video  $T_{\text{c}}$  is the total time taken for network communication.  $T_{\text{l}}$  is the time lost caused by the failure of identifying the target berry.  $N_{\text{thinned}}$  is the number of berries thinned in one cluster.

The time lost when target berry was not identified is calculated as follows:

$$T_{\text{l}} = (1 + T_{\text{c}}) \times N_{\text{fail}} \quad (2)$$

Here,  $T_{\text{c}}$  is the average time taken for network communication, and  $N_{\text{fail}}$  is the total number of target berries not identified.

2. Evaluation using System Usability Scale (SUS) and User Experience Questionnaire (UEQ)

These evaluation methods are conducted according to the guidelines given in [12] and [13], respectively.

3. Participants' comments

Free comments collected after the experiment were included in the evaluation.

**E. Experimental Results**

1. Average Time Taken to Thin One Berry

The average time taken to thin one cluster for each interface is shown in Table II. The average time  $t$  was 16.7 seconds for image only (I), 11.2 seconds for image with contour overlay (I-O), and 10.5 seconds for image with contour overlay and voice instructions (I-O-V), with I-O-V resulting in the shortest time. A paired t-test was conducted to verify whether there were significant differences in the average time taken to thin one grape between interfaces I, I-O, and I-O-V, applying Bonferroni correction to address multiple comparisons. The results of each comparison are shown in Table III. These results indicate that there were no significant differences in the thinning time between every two interfaces, and none of the interfaces showed superiority in thinning time compared to the others.

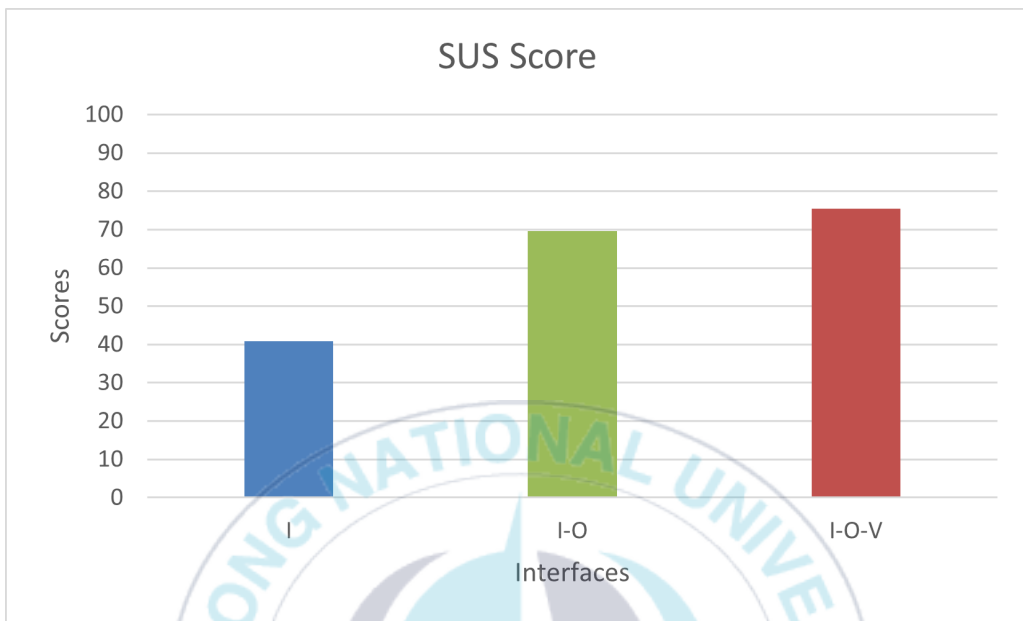
**Table II AVERAGE TIME TAKEN TO THIN ONE BERRY.**

Interface	I	I-O	I-O-V

Average Time (s)	16.7	11.2	10.5
------------------	------	------	------

**Table III COMPARISON OF AVERAGE TIME TAKEN TO THIN ONE BERRY**

Comparison	t-Statistic	p-Value	Bonferroni Corrected p-Value
I vs I-O	1.1984	0.2845	0.8534
I vs I-O-V	1.2759	0.2581	0.7742
I-O vs I-O-V	1.6061	0.1692	0.5075



**Figure 10 Scores for each interface.**

## 2. Evaluation using System Usability Scale (SUS)

The average SUS scores for each interface are shown in Fig 10.

1. "Image only (I)": The average SUS score was 40.8, corresponding to a rating between Poor and OK in usability evaluation.
2. "Image with contour overlay (I-O)": The average SUS score was 69.6, corresponding to a rating between OK and Good.
3. "Image with contour overlay and voice instructions (I-O-V)": The average SUS score was 75.4, corresponding to a rating between Good and Excellent.

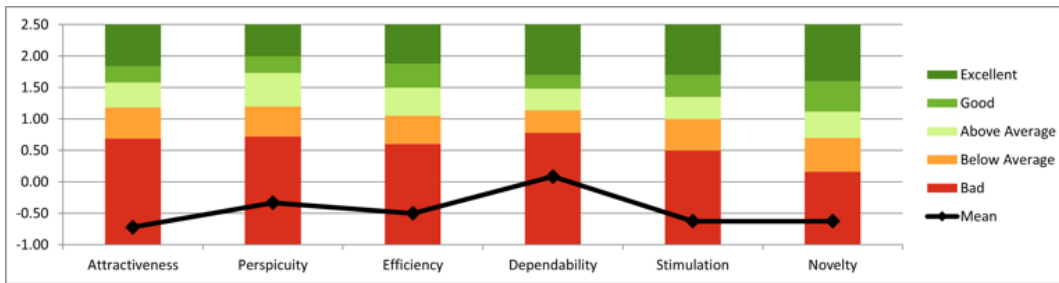
A paired t-test was conducted to verify whether there were significant differences in SUS scores between interfaces, applying Bonferroni correction to address multiple comparisons.

The results of each comparison are shown in Table IV.

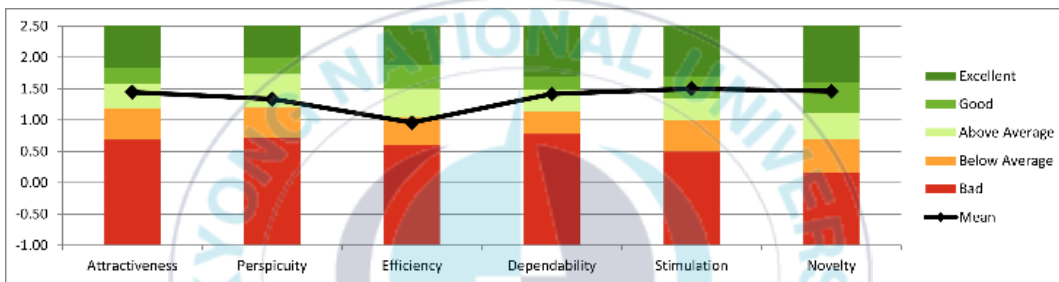
**Table IV COMPARISON OF SUS SCORES**

Comparison	t-Statistic	p-Value	Bonferroni Corrected p-Value
I vs I-O	-3.8369	0.0122	0.0365*
I vs I-O-V	-5.0308	0.004	0.012*
I-O vs I-O-V	-1.4	0.2204	0.6612

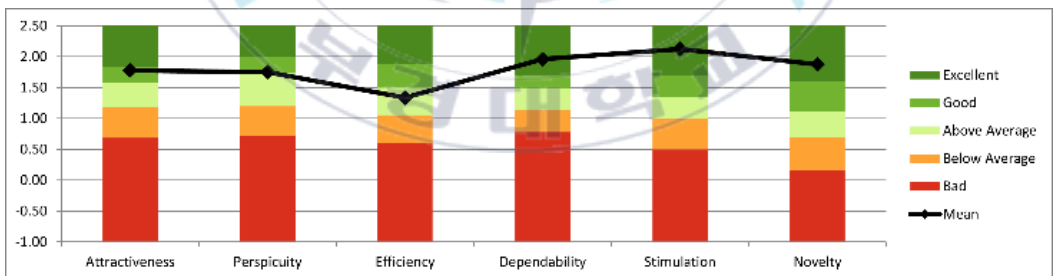
The results of the t-test between "Image only (I)" and "Image with contour overlay (I-O)" showed a statistically significant difference with a Bonferroni-corrected p-value less than 0.05, indicating that the usability of "Image with contour overlay (I-O)" is superior to that of "Image only (I)". The results of the t-test between "Image only (I)" and "Image with contour overlay and voice instructions (I-O-V)" also showed a statistically significant difference with a Bonferroni-corrected p-value less than 0.05, indicating that the usability of "Image with contour overlay and voice instructions (I-O-V)" outperforms that of "Image only (I)". The t-test results between "Image with contour overlay (I-O)" and "Image with contour overlay and voice instructions (I-O-V)" showed no statistically significant difference with a Bonferroni-corrected p-value greater than 0.05. Based on these results, we can conclude that "Image with contour overlay (I-O)" and "Image with contour overlay and voice instructions (I-O-V)" have better usability compared to "Image only (I)". However, no significant difference was observed between I-O and I-O-V.



I-O



I-O-V



**Figure 11 UEQ results and benchmarks for each interface.**

### 3. Evaluation using User Experience Questionnaire (UEQ)

The User Experience Questionnaire (UEQ) was used to evaluate the system's usability. The average UEQ scores for each interface are shown in Fig 11. The benchmark evaluation of the

user experience for the proposed method, "Image with contour overlay and voice instructions (I-O-V)", showed Attractiveness (Good), Perspicuity (Good), Efficiency (Above Average), Dependability (Excellent), Stimulation (Excellent), and Novelty (Excellent), indicating above-average user experience in all categories except for efficiency, which was rated below Good. A paired t-test was conducted to verify whether there were significant differences in user experience between interfaces, applying Bonferroni correction to address multiple comparisons. The comparison results of UEQ scores between interfaces are shown in Table V.

**Table V COMPARISON OF UEQ SCORES FOR EACH INTERFACE.**

	I vs I-O	I vs I-O-V	I-O vs I-O-V
Attractiveness p-Value	0.186	0.1137	0.6917
Perspicuity p-Value	0.2454	0.0915	0.9345
Efficiency p-Value	0.0954	0.0243 *	0.8265
Dependability p-Value	0.4362	0.0987	0.4534
Stimulation p-Value	0.1434	0.1434	0.3995
Novelty p-Value	0.1374	0.1374	1.069

From the results of each evaluation item, it was shown that "Image with contour overlay and voice instructions (I-O-V)" is more effective in terms of efficiency compared to "Image only (I)". No significant differences were observed in other items.

#### 4. User Feedback

##### 1. Image Only(I)

The interface "Image Only (I)" received many comments stating that it was unclear which berry should be thinned and that it is difficult to be used in real situation. Furthermore, it was pointed out that the visibility of the displayed image was poor unless the position of the smart glasses was completely adjusted for user. Additionally, the instructions were unclear, making it difficult to perform tasks while looking at the image, and this mode was considered the most difficult to use.

## 2. Image with Contour Overlay(I-O)

The interface "Image with Contour Overlay (I-O)" received positive evaluations for making it easier to identify which berry to thin thanks to the superimposing of the contour lines. However, it was noted that distinguishing between nearby berries was difficult, and the contour lines often became misaligned when the grape clusters were moved. Some users felt that while the overlaying was not 100% successful, it was still useful in practical application.

Those with experience using smart glasses found it is easy to use, but there were also some comments indicating that it was sometimes unclear what actions to take.

## 3. Image with Contour Overlay and Voice Instructions(I-O-V)

The interface "Image with Contour Overlay and Voice Instructions (I-O-V)" was evaluated positively for making it easy to understand the timing of taking photos, thanks to the voice instructions, allowing quick operations. Many users found it easier to grasp the timing of recognition compared to modes without voice instructions and found it easy to use. Although some struggled with target berry identification, once identified, operations could be performed quickly. The function that played a sound when the correct berry was cut was well-received, although some beginners found the voice instructions slightly irritating. While the contour lines made it clear which berry to thin, some users found it difficult to identify the berries that were on the back side of the cluster.

## 5. DISCUSSION

### **Average Time to Thin One Grape**

The evaluation experiment showed no significant difference in the average time to thin one berry between every two interfaces. This is likely due to the small number of participants, resulting in no significant differences between the interfaces. However, the trend in values indicated that the proposed method, "Image with Contour Overlay and Voice Instructions (I-O-V)", had the shortest average time to thin one grape. The average number of times the target berries were not thinned was 5.3 for "Image Only (I)", 2.5 for "Image with Contour Overlay (I-O)", and 3.2 for "Image with Contour Overlay and Voice Instructions (I-O-V)". These results suggest that users had difficulty determining which berry to thin with the "Image Only (I)" interface, leading to an increased number of instances where thinning was not performed, resulting in longer operation times.

Overall, the "Image with Contour Overlay and Voice Instructions (I-O-V)" interface may contribute to improving efficiency, and further experiments are needed to verify its effectiveness.

### **Evaluation of Usability and User Experience**

The usability evaluation was conducted using the System Usability Scale (SUS), and the user experience evaluation was conducted using the User Experience Questionnaire (UEQ). The results showed that the interfaces "Image with Contour Overlay (I-O)" and "Image with Contour Overlay and Voice Instructions (I-O-V)" had higher usability compared to "Image Only (I)". However, no significant difference was observed between I-O and I-O-V. In particular, the "Image with Contour Overlay and Voice Instructions (I-O-V)" interface was rated higher in terms of efficiency compared to "Image Only (I)". This indicates that overlaying contours on grapes improves usability, and voice instructions enhance efficiency and user experience.

## **User Feedback**

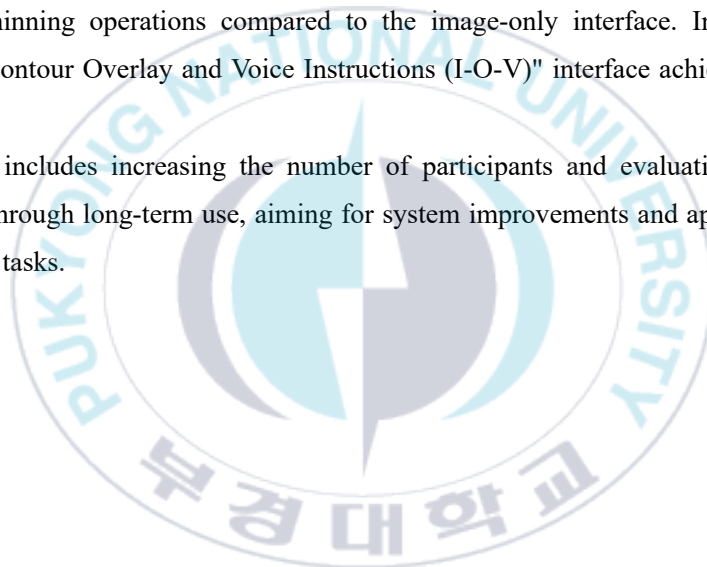
From the user feedback, the "Image Only (I)" interface was rated low in terms of visibility and unclear instructions, resulting in poor usability. On the other hand, the "Image with Contour Overlay (I-O)" interface improved visibility, but it was challenging to distinguish between nearby grapes, and the contour lines and grapes were often misaligned. The "Image with Contour Overlay and Voice Instructions (I-O-V)" interface was rated highly for ease of use and additional features, making it the most practical approach. However, some participants found the voice instructions slightly annoying.



## 6. CONCLUSION

This study proposed and evaluated the grape thinning support system that combines deep neural networks (DNN) and augmented reality (AR). The system is used to predict berries to be thinned and superimpose a virtual contour over the target berry, making it easy for users to identify the grapes. Additionally, the hand-tracking function was used to monitor the thinning operation in real-time and provide voice instructions to improve work efficiency and user experience. The evaluation experiment showed that the proposed system improved the usability of thinning operations compared to the image-only interface. In particular, the "Image with Contour Overlay and Voice Instructions (I-O-V)" interface achieved the highest usability.

Future work includes increasing the number of participants and evaluating the system's effectiveness through long-term use, aiming for system improvements and application to real grape thinning tasks.



## References

- [1] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
- [3] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016. Accessed: Jul. 10, 2024. [Online]. Available: [https://books.google.co.jp/books?hl=ja&lr=lang\\_ja&lang\\_en&id=omivDQAAQBAJ&oi=fnd&pg=PR5&dq=I.+Goodfellow,+Y.+Bengio,+and+A.+Courville,+Deep+Learning,+Cambridge,+MA:+MIT+Press,+2016.&ots=MOO\\_fsszWW&sig=fDQNJ3Y86kzfA9xZYMoGMsD85SY](https://books.google.co.jp/books?hl=ja&lr=lang_ja&lang_en&id=omivDQAAQBAJ&oi=fnd&pg=PR5&dq=I.+Goodfellow,+Y.+Bengio,+and+A.+Courville,+Deep+Learning,+Cambridge,+MA:+MIT+Press,+2016.&ots=MOO_fsszWW&sig=fDQNJ3Y86kzfA9xZYMoGMsD85SY)
- [4] P. Milgram and F. Kishino, “A taxonomy of mixed reality visual displays,” *IEICE Trans. Inf. Syst.*, vol. 77, no. 12, pp. 1321–1329, 1994.
- [5] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino, “Augmented reality: A class of displays on the reality-virtuality continuum,” in *Telem manipulator and telepresence technologies*, Spie, 1995, pp. 282–292. Accessed: Jul. 10, 2024. [Online]. Available: <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/2351/1/Augmented-reality-a-class-of-displays-on-the-reality/10.1117/12.197321.short>
- [6] P. Buayai, K. R. Saikaew, and X. Mao, “End-to-end automatic berry counting for table grape thinning,” *IEEE Access*, vol. 9, pp. 4829–4842, 2020.
- [7] P. Buayai, K. Yok-In, D. Inoue, H. Nishizaki, K. Makino, and X. Mao, “Supporting table grape berry thinning with deep neural network and augmented reality technologies,”

*Comput. Electron. Agric.*, vol. 213, p. 108194, 2023.

[8] D. Mourtzis, V. Zogopoulos, and F. Xanthi, “Augmented reality application to support the assembly of highly customized products and to adapt to production re-scheduling,” *Int. J. Adv. Manuf. Technol.*, vol. 105, no. 9, pp. 3899–3910, Dec. 2019, doi: 10.1007/s00170-019-03941-6.

[9] J. Chalhoub and S. K. Ayer, “Exploring the performance of an augmented reality application for construction layout tasks,” *Multimed. Tools Appl.*, vol. 78, no. 24, pp. 35075–35098, Dec. 2019, doi: 10.1007/s11042-019-08063-5.

[10] F. Lamberti, F. Manuri, A. Sanna, G. Paravati, P. Pezzolla, and P. Montuschi, “Challenges, opportunities, and future trends of emerging techniques for augmented reality-based maintenance,” *IEEE Trans. Emerg. Top. Comput.*, vol. 2, no. 4, pp. 411–421, 2014.

[11] P. Buayai *et al.*, “End-to-end inflorescence measurement for supporting table grape trimming with augmented reality,” in *2021 International Conference on Cyberworlds (CW)*, IEEE, 2021, pp. 101–108. Accessed: Jul. 10, 2024. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9599366/>

[12] O. Faugeras, *Three-dimensional computer vision: a geometric viewpoint*. MIT press, 1993. Accessed: Jul. 21, 2024. [Online]. Available: [https://books.google.co.jp/books?hl=ja&lr=lang\\_ja|lang\\_en&id=Aa6TTW9dWy0C&oi=fnd&pg=PR11&dq=Three-Dimensional+Computer+Vision:+A+Geometric+Viewpoint&ots=eYyOGt5bXd&sig=kjxDNVHHOW3Pc-vfdbHC3yAUBp0](https://books.google.co.jp/books?hl=ja&lr=lang_ja|lang_en&id=Aa6TTW9dWy0C&oi=fnd&pg=PR11&dq=Three-Dimensional+Computer+Vision:+A+Geometric+Viewpoint&ots=eYyOGt5bXd&sig=kjxDNVHHOW3Pc-vfdbHC3yAUBp0)

[13] J. Brooke, “SUS: a retrospective.,” *J. Usability Stud.*, vol. 8, no. 2, 2013, Accessed: Jul. 10, 2024. [Online]. Available: [http://uxpajournal.org/wp-content/uploads/sites/7/pdf/JUS\\_Brooke\\_February\\_2013.pdf](http://uxpajournal.org/wp-content/uploads/sites/7/pdf/JUS_Brooke_February_2013.pdf)

[14] M. Schrepp, A. Hinderks, and J. Thomaschewski, “Applying the User Experience

Questionnaire (UEQ) in Different Evaluation Scenarios,” in *Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience*, vol. 8517, A. Marcus, Ed., in *Lecture Notes in Computer Science*, vol. 8517. , Cham: Springer International Publishing, 2014, pp. 383–392. doi: 10.1007/978-3-319-07668-3\_37.

