

# Initial Investigation of Visual Marker Design for Activity Recognition using Near-infrared Images

JOSEPH KORPELA<sup>1,a)</sup> TAKUYA MAEKAWA<sup>1,b)</sup>

**Abstract:** This study investigates the design of reflective markers for use in activity recognition using low-exposure near-infrared (NIR) images. These reflective markers can be placed on everyday objects, e.g., kitchen knives, to recognize object-based activities, e.g., cutting vegetables. These visual markers must balance a need to keep the marker sizes small, to allow their placement on the everyday objects, while still being large enough for detection in the NIR images. The markers must also be identifiable when placed on curved surfaces, such as when wrapped around handles. Our investigation indicates that markers created using flexible retroreflective material with sizes as small as 7x7 mm can be attached to many surfaces and provide enough reflectivity to be detected in the NIR images used in this study.

## 1. Introduction

In an aging population, such as that of Japan, solutions are needed to reduce the burden on families and doctors caring for the elderly. One such solution comes from the field of human activity recognition. Human activity recognition allows caregivers to track the daily routines of the elderly using systems that collect data, e.g., video, and process it automatically using machine learning algorithms. With this data in hand, caregivers can easily monitor their patients with little burden on either party.

However, when using such systems, care must be taken to protect the privacy of those being monitored. In the case of video-based activity recognition, this can be quite difficult. Even if measures are taken to protect the video recorded, the patient may still feel uncomfortable being recorded, and may either refuse to adopt the technology or may avoid its use. Therefore, existing approaches to computer-vision-based activity recognition may not always be a suitable choice in caregiving situations.

In this paper, we examine a method for computer-vision-based activity recognition that uses a near-infrared (NIR) camera to capture images outside the visible spectrum that show only images of highly reflective surfaces. We propose using retroreflective material to create markers that can be attached to everyday objects and can be used to track the use of those objects as a means to performing activity recognition.

## 2. Related Work

Several previous studies has proposed methods for using computer vision to conduct human activity recognition. In [3], the authors examined human activity recognition using stereo cameras, with a goal of investigating a method for allowing robots

to recognize human activities conducted in a kitchen. In [4], the authors conducted human activity recognition using a combination of accelerometer data and RGB video data. In [1] and [2], the authors used video data that included depth information to conduct human activity recognition for kitchen related activities. Each of these studies relied on the collection of video in the visual spectrum and as such are not privacy preserving. In contrast, our method allows for activity recognition using a modified webcam that will preserve the user's privacy.

## 3. Detecting NIR Markers

Our method uses a 1080p webcam modified to remove its IR-blocking visible-light-passing filter, replacing it with a visible-light-blocking IR-passing filter. The modified camera is then used to capture video at 30 fps at an exposure of 3.9 ms, with the resulting images showing only the reflections from highly reflective surfaces. We then attach highly reflective markers to everyday objects that are used in our target activities. The markers are made of 3M Scotchlite 6160R White High Gloss Trim retroreflective material, a durable and highly reflective material designed for heavy-duty use in workplaces such as construction sites. Figure 1 shows example images captured using our setup. Object (b) is a 7 mm wide strip of retroreflective material wrapped around the handle of a small paring knife.

Using a marker as simple as object (b), it is possible to track an object over time to gather movement information for activity recognition. However, such a simple design does make it difficult to distinguish between the marker and background noise. For example, in Figure 1, both object (a) and object (b) have a similar shape in the NIR images. One way to better distinguish the marker from background noise is to use more complex marker designs, as shown in Figure 2. Using such designs makes it easier to distinguish between markers and background noise and also makes it possible to track multiple objects simultaneously, using

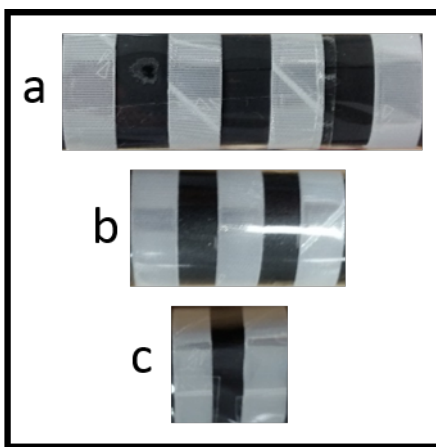
<sup>1</sup> Osaka University

<sup>a)</sup> joseph.korpela@ist.osaka-u.ac.jp

<sup>b)</sup> maekawa@ist.osaka-u.ac.jp



**Fig. 1** Three NIR images captured by our modified webcam. Object (a) is background noise and object (b) is a retroreflective marker placed on the handle of a paring knife. The images show the movement of the paring knife from right to left, with the middle frame showing object (a) being occluded by the paring knife.



**Fig. 2** Three example marker patterns that make it easier to distinguish markers from background noise and make it possible to identify multiple objects in a NIR image frame.

the markers to distinguish between the objects being tracked.

#### 4. Tracking NIR Markers Over Time

In order to perform activity recognition using our NIR data, we need to first be able to detect and track objects over time. To accomplish this, we first detect objects on a per frame basis by performing image preprocessing to remove noise and detect areas of high reflectivity in the images. We then use Kalman filters to track objects across frames in order to collect time series data for use in activity recognition.

We preprocess the raw NIR image in two main steps. The first step involves processing the image using a combination of thresholding, opening, and closing operations to remove background noise, remove small areas of reflectivity, and then build up regions of interest (ROIs) around the remaining areas of reflectivity. We then use the ROIs discovered in the first step to focus on areas of the raw image to preprocess using a combination of adaptive thresholding and histogram equalization to create a clearer image of each object. Finally, we compute the area and histogram for each potential marker, to use when comparing them to existing tracks.

We track markers across frames using Kalman filters, with the Kalman filters modeling movement based on the objects' veloci-

ties on the x and y axes. With each new frame, we start with a list of predicted new locations for existing tracks from our Kalman filters and a list of objects detected in the current frame. We compare the existing tracks to detected objects based on their histograms, areas, and distances from each other, and then assign detected objects to existing tracks using the closest matches. Existing tracks that are not matched to new objects for over 60 frames are deleted, while new objects that are not matched to existing tracks are assigned as new tracks. Figure 1 shows an example of object tracking, with objects (a) and (b) tracked across three frames. Object (a) is tracked across a short period of occlusion (frame 2) while object (b) is tracked as it moves across the work area, with the Kalman filter correctly tracking the two distinct objects even when they are in close proximity.

#### 5. Evaluation

We evaluated our method using 1080p NIR video recorded at 30 FPS in four different environments, with an average of 12.6 seconds recorded per environment. The camera was placed overhead each activity area at a height of 2.45 m. The marker used was a 7 mm wide strip of retroreflective material wrapped around the handle of a small paring knife. In each video, the marker was moved around the environment to test the ability of our Kalman filter to track the marker during use.

In all four environments, the Kalman filters were able to track the marker throughout its entire active period, despite short periods of occlusion. In addition, three of the four environments also included background noise, in the way of non-target objects that were reflecting NIR light from chrome surfaces. In each of these three environments, the Kalman filter was able to track the target marker without confusing it with background noise.

Table 1 shows three measures for the prediction errors of the Kalman filters in each of the four test environments. Comparing the average error in Table 1 to the *Avg Width* in Table 2, we can see that on average, the Kalman filters were able to predict new locations for the markers that fell within a few pixels of the markers edges. However, the *Max Error* was much higher, most likely due to the short periods during which the markers were occluded (Table 2 shows how markers were only visible for about 88% of the time on average). Even with the high *Max Error* for posi-

**Table 1** Prediction errors for Kalman filters when tracking NIR markers in four different environments. Error is measured as the distance in pixels between the predicted center of the marker and the actual center of the marker.

	Average Error	Max Error	Error Std Dev
Environment 1	5.20	31.76	4.75
Environment 2	5.17	66.41	6.92
Environment 3	5.40	106.78	10.98
Environment 4	6.32	56.85	7.23
Overall	5.35	106.78	7.64

**Table 2** Measures of marker visibility in each of the four environments. The Visible Ratio is measured as number of frames visible divided by total number of frames active. The Avg Width is computed as the square root of the marker area (in pixels).

	Visible Ratio	Avg Width (pixels)
Environment 1	0.89	7.43
Environment 2	0.93	7.67
Environment 3	0.80	7.74
Environment 4	0.99	8.21
Overall	0.88	7.64

tion estimates, the Kalman filters were still able to maintain their tracking after these errors.

## 6. Conclusion

This paper investigates a design for retroreflective markers for use in activity recognition with NIR video. We believe that

this method is useful for activities that require the use of tools and utensils, allowing for multiple objects to be tracked in the workspace without collecting visible-light video of the users, thus preserving their privacy. Since the reflective markers are made of the same material that is commonly sewn into work uniforms in occupations such as construction work, we believe that these durable markers will allow for activity recognition with low maintenance costs. In our future work, we plan to expand our marker design to allow for multiple marker types so that the visible objects can be identified by marker type.

**Acknowledgments** This study is partially supported by JST CREST.

## References

- [1] Ji, Y., Ko, Y., Shimada, A., Nagahara, H. and Taniguchi, R.: Cooking gesture recognition using local feature and depth image, *Proceedings of the ACM Multimedia 2012 Workshop on Multimedia for Cooking and Eating Activities*, ACM, pp. 37–42 (2012).
- [2] Lei, J., Ren, X. and Fox, D.: Fine-grained kitchen activity recognition using RGB-D, *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, ACM, pp. 208–211 (2012).
- [3] Rybok, L., Friedberger, S., Hanebeck, U. D. and Stiefelhagen, R.: The KIT Robo-Kitchen data set for the evaluation of view-based activity recognition systems, *Proceedings of the 11th IEEE-RAS International Conference on Humanoid Robots*, IEEE, pp. 128–133 (2011).
- [4] Stein, S. and McKenna, S. J.: Combining embedded accelerometers with computer vision for recognizing food preparation activities, *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ACM, pp. 729–738 (2013).