# Blind navigation with a wearable range camera and vibrotactile helmet

(author's name removed for double-blind review) X university 1@2.com

(author's name removed for double-blind review) X university 1@2.com

(author's name removed for double-blind review) X university 1@2.com

## ABSTRACT

We present a wayfinding system that uses a range camera and an array of vibrotactile elements we built into a helmet.

The range camera is a Kinect 3D sensor from Microsoft that is meant to be kept stationary, and used to watch the user (i.e., to detect the person's gestures). Rather than using the camera to look at the user, we reverse the situation, by putting the Kinect range camera on a helmet for being worn by the user. In our case, the Kinect is in motion rather than stationary.

Whereas stationary cameras have previously been used for gesture recognition, which the Kinect does very well, in our new modality, we take advantage of the Kinect's resilience against rapidly changing background scenery, where the background in our case is now in motion (i.e., a conventional wearable camera would be presented with a constantly changing background that is difficult to manage by mere background subtraction).

The goal of our project is collision avoidance for blind or visually impaired individuals, and for workers in harsh environments such as industrial environments with significant 3-dimensional obstacles, as well as use in low-light environments.

## Categories and Subject Descriptors

I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis—*Depth cues, Range data, Motion*; I.4.9 [**Image Processing and Computer Vision**]: Applications; H.1.2 [**Information Systems**]: Models and PrinciplesUser/Machine Systems[Human factors]

## Keywords

personal safety devices, blind navigation, Microsoft Kinect, depth sensor, human computer interface

*ACM MM 2011* Scottsdale, Arizona USA

## 1. INTRODUCTION

### 1.1 Conventional uses of Kinect

The Kinect, from Microsoft, was designed for use with Microsoft's XBOX360 gaming console. The Kinect allows the gamer to interact with games without the need for physical controls. It accomplishes this by tracking the gamer's movements and position in 3-Dimensional space, with respect to itself, in real-time. In normal use, the Kinect sits stationary and observes the gamer as he/she moves.

### 1.2 Reversing the role of user and camera

We propose the use of the Kinect in a different manner, where the Kinect moves with the user, so that it observs the world in a similar fashion as the user observes (or would have observed, in the case of a blind individual).

Rather than having the Kinect watch the user, the user uses it to watch their environments.

In our implementation, the Kinect is used to extract the 3-dimensional depth information of the environment being observed by the user. This depth information is passed to the user in the form of tactile feedback, using an array of vibrotactile actuators.

Microsoft's Kinect employs PrimeSense's 3-D sensing technology. PrimeSense's 3-D sensor uses light coding to code the scene volume, using active IR (infrared) illumination [**?**][**?**][**?**]. The sensor then uses a CMOS image sensor to read the coded light back from the scene. The coded light is processed by PrimeSense's SoC chip [**?**], contained in the 3-D sensor, to give the depth information.

### 1.3 Other head-mounted navigational aids

Most previous head-mounted navigational aids have used standard camera systems, to present tactile information to the user. One such example is called "seeing with the tongue" [**?**].

Standard camera systems work well for gesture recognition because the stationary background can be subtracted from the image, so that people can be clearly seen with simple computer image processing. However, when the camera is wearable, the background is constantly changing, making it difficult to separate distant background clutter from nearby objects.

Some specialized blind navigation aids such as the VibraVest[**?**]

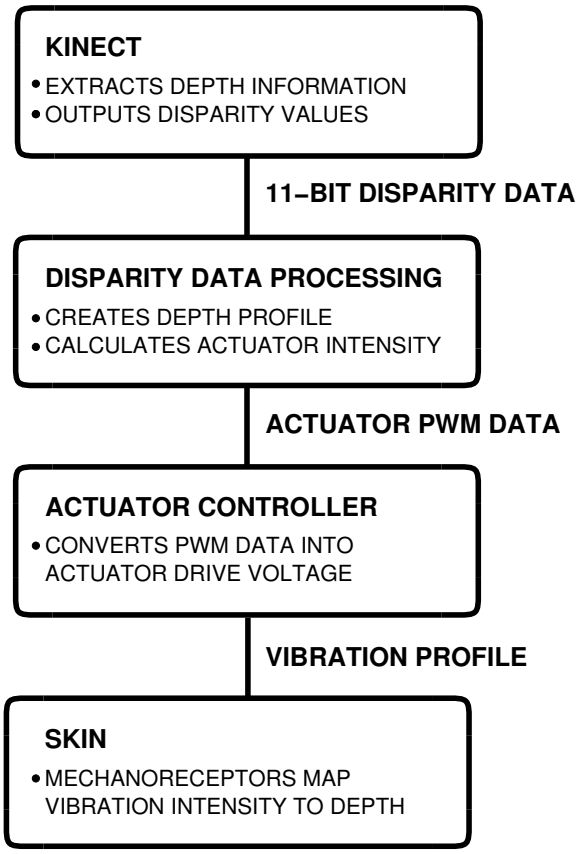


**Figure 1: System signal flow path overview**

provided 3D range information but required expensive special-purpose hardware such as a miniature radar system.

The Kinect is a new widely deployed commercial off-the-shelf technology that provides a low-cost solution to the problems associated with sensitivity to distant background clutter. Since background clutter is especially prevalent in a wearable camera situation, the technology used in the Kinect shows great promise in wearable vision systems.

# 2. PHYSICAL SETUP

Figure 1 shows the signal flow path in our system architecture.

Data is captured from the Kinect camera, processed, and supplied to an array of vibrotactice actuators.

Our goal is to convert depth information obtained using Kinect into haptic feedback so that users can perceive depth within a range that matters most for collision avoidance, while not being overwhelmed by distant background clutter.

The Kinect depth camera, coupled with a wearable computer running Openkinect drivers, was used to create a depth map of the image.

An array of six vibrating actuators mounted inside a helmet are controlled using the depth values using an algorithm that calculates the vibration intensity profile for each of these actuators. The intensity profile is transmitted to an Arduino microcontroller (also part of the wearable system), which drives each of the actuators using PWM (Pulse-Width Modulation). PWM allows voltage on the actuators to be regulated for varying degrees of vibration. Fig 1 shows how the varying degrees of vibrations are picked up by the mechanoreceptors present in the sensitive skin on the forehead of the usera. Using this system, the user has a sense of depth.

This sense of depth moves with the head in a natural manner. Thus, the user can scan the head back and forth to get a natural understanding of subject matter in their environment.

The general layout of our helmet is depicted in Fig 4.

We mounted the Kinect securely on top of a welding helmet. An array of 6 vibration actuators were positioned along the headstrap of the helmet. The helmet is placed on the head as shown in Fig 4.

For testing by sighted users, a dark welding shade was used, which could either entirely stop light from passing through, or, under computer program control, vary the amount of light passing through. In this way, the device could function as a switchable blindfold for testing purposes.

## 2.1 Vibrotactile actuators, and motor controllers

We used a set of vibrating actuator motors. The vision processing algorithm controls the motors through a serial connection to an Arduino microcontroller. These values correspond directly to PWM output from pins 2 to 7 on the Arduino. Each output pin is used as control signal in motor driver circuit which determines the actuator vibration response for AL3 to AR3 as shown in Fig 5.

For our setup, we used 10x3.4mm shaftless vibration motor for each of the actuators. The motor is rated to be driven at a maximum voltage of 3.6V. Therefore, we supplied the 3.6V power supply to the motor driver circuits. Depending on the PWM value from each of the Arduino pins, the corresponding actuator can be driven at voltages calculated as:

$$V_{\text{actuator}} = \text{PWM}/255 * 3.6, \quad \text{PWM} \in [0, 255] \qquad (1)$$

The actuator Voltage and Current response was tested to be linear. Based on this we determined that the vibrating actuator also had a linear reponse, when driven between voltages 0 to 3.6V.

# 3. REAL-TIME IMAGE PROCESSING TO CONTROL VIBRATION ACTUATORS

## 3.1 Distance map recovered from Kinect

We accessed the Kinect data in real-time with a Linux PC. The Kinect provides data in a proprietary data format, in what is called “disparity” values.

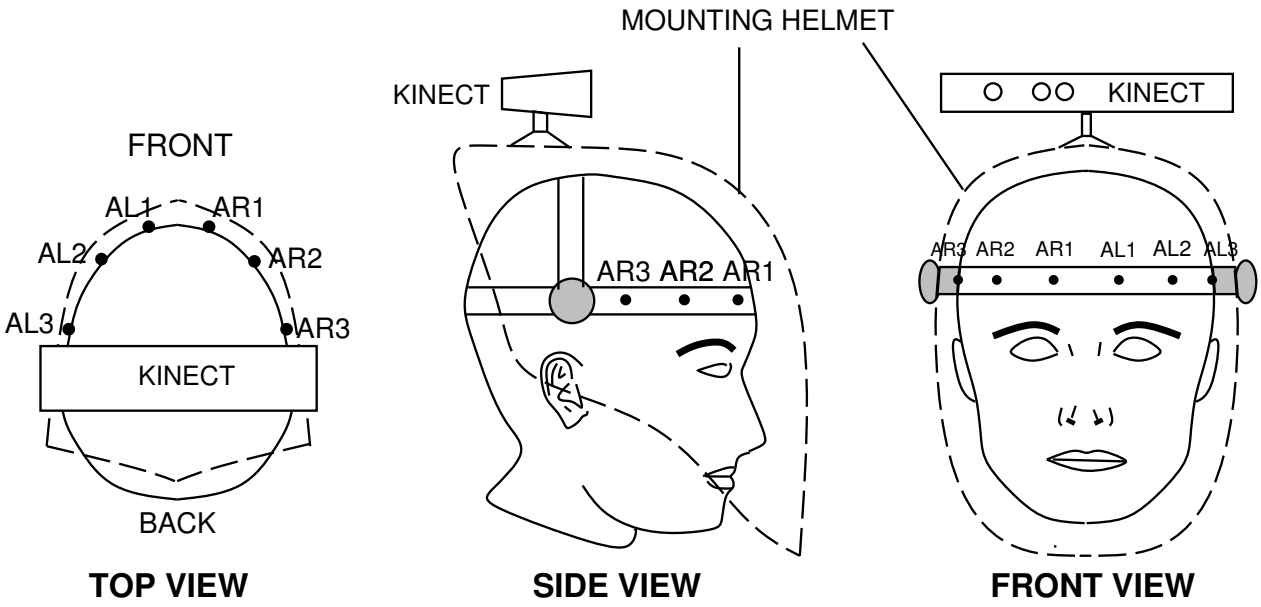


**Figure 4: Wearable sensor and actuator configuration on a helmet, showing placement of the motor actuators around the forehead.**

To recover the actual distance from the raw depth data in the proprietary format, we used the following conversion equation

$$\text{distance} = R = \frac{1}{\alpha \cdot \text{disparity} + \beta} \tag{2}$$

The parameters have been empirically found in [**?**] to be: $\alpha = -0.0030711016$ and $\beta = 3.3309495161$. As a result, the range extremities become:

| | disparity | distance |
|---|---|---|
| MIN distance detectable | 0 | 0.30 m |
| MAX distance detectable | 1030 | 5.96 m |

## 3.2 Partitioning the distance map

The Kinect operates with horizontal field of view of 57° horizontally and 43° vertically. It is able to measure disparity values beyond a critical distance of 0.3m.

At distances closer than 30cm, the Kinect is not able to measure disparity. The disparity values are calibrated such that the device is able to read values up to 6m without significant loss of resolution within acceptable error margin while operating indoors. We found this range of 0.3 to 6.0 metres to be useful for collision avoidance at typical walking speeds.

In our setup, the depth sensing region was divided into six smaller zones, three on the left (SL1, SL2, and SL3) and three on the right (SR1, SR2, SR3). Each of the zones corresponds to the vibration in one actuator.

## 3.3 Controlling a 1-dimensional array of actuators

Fig 6 shows the layout of sensing regions. This layout allows the user to scan their head back and forth and feel various objects as if they were pressing against their forehead. While not providing high resolution imaging, we found that it was possible to navigate a hallway and find various doors, doorways, etc., and also avoid collision with other people in a crowded hallway.

## 3.4 Transfer function with 1-to-1 mapping from each sensing region to each actuator

We desired objects in the visual field to cause the vibrations to become stronger as the objects get closer. In this way the system creates the sensation of objects pressing against the forehead at-a-distance, i.e. before collision occurs. The sensation increases in strength as collision is more eminent.

This can be accomplished by making the vibration (as sensed) be inversely proportional to distance, i.e. $V \propto 1/R$.

Alternately we can make the sensed vibration vary as $V \propto 1/R^2$, as with a force field such as a magnetic field. For example, when holding two magnets close to each other, the magnetic repulsion (or attraction) is inversely proportional to the separation distance squared. Thus we can mimic nature, in this regard, in order to make the effect comprehensible and intuitive.

We now partition the depth map for each of the different vibration actuators. For an inverse-square law, we make the

**Figure 2: Our wearable configuration with a Kinect and vibrotactile actuators mounted on a welding helmet. A welding helmet was chosen because of its comfort, its electronically controllable blindfold (which defaults to transparent upon power failure), and because, other work, we are using this system in welding applications.**

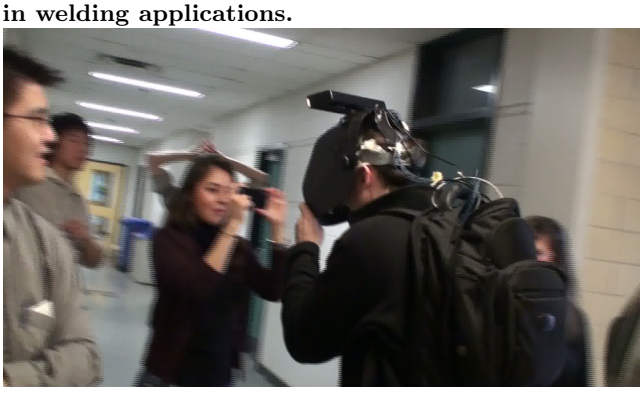

**Figure 3: The system is effective in helping with navigation through crowded corridors, for example. See http://wearcam.org/blindvision**

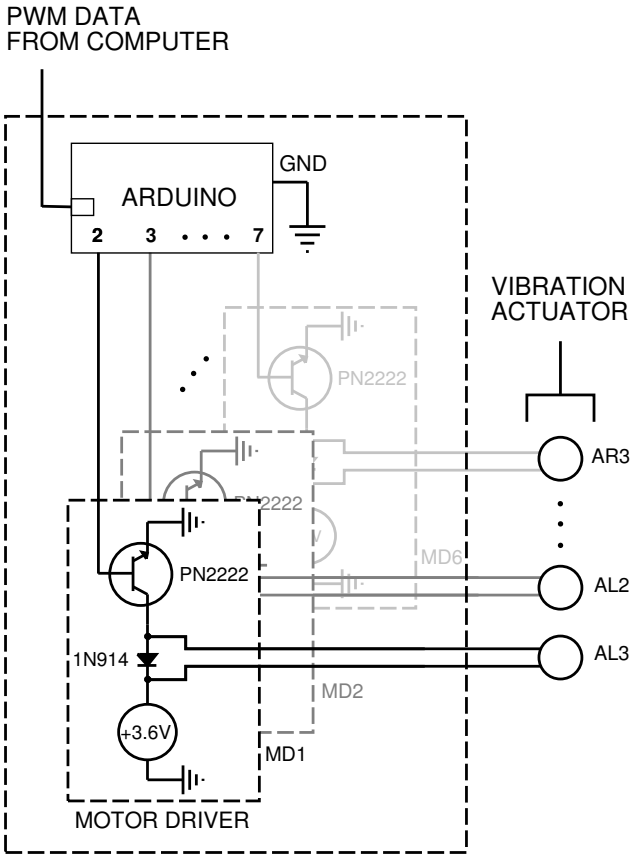


**Figure 5: Vibration actuator drive electronics**

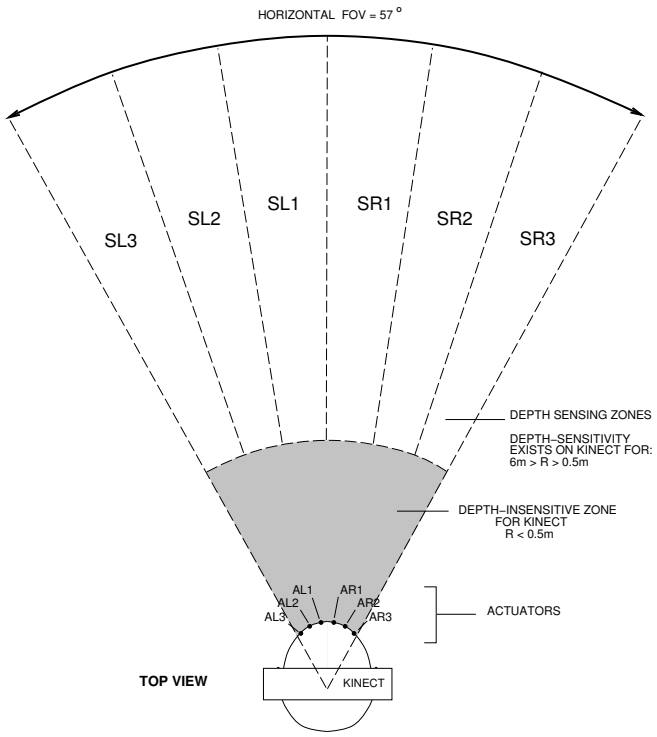


**Figure 6: Partitioning the depth sensing map into zones for the separate control of vibration actuators. The sensing zones have been numbered SL3...SR3, and the actuators numbered AL3...AR3.**

total vibration a weighted integral across the sensing region:

$$v_n = \frac{1}{S_{FOV}} \int_{S_{FOV}} \frac{1}{R^2(\theta,\phi)} a_{\theta,n}(\theta) a_{\phi,n}(\phi) dS \qquad (3)$$

for actuator $n$. $a_{\theta,n}$ and $a_{\phi,n}$ are aperture functions, weightings which vary depending on the horizontal and vertical locations, respectively. They are different for each sensing region $n$. $S$ is sensing surface in steradians.

We found empirically that the background noise from Eqn. 3, coming from objects in the background behind the subject matter, was distracting. Our second implementation simply used the closest object in zone, still weighted by the aperture functions:

$$v_n = \min_{S_{FOV}} \frac{1}{R^2(\theta,\phi)} a_{\theta,n}(\theta) a_{\phi,n}(\phi) \qquad n \in 1...N \qquad (4)$$

We also experimented with the following 1/R law, which gave an improved advance warning of faraway objects approaching ($>$ 3m).

$$v_n = \min_{S_{FOV}} \frac{1}{R(\theta,\phi)} a_{\theta,n}(\theta) a_{\phi,n}(\phi) \qquad n \in 1...N \qquad (5)$$

The result is a center-weighted mapping, as illustrated in Fig 7.

## 3.5 Fuzzy zone boundaries

It is also possible to go beyond a simple 1-to-1 mapping between each spatial sensing region and each actuator. For example, we experimented with making fuzzy boundaries on each sensing region, using a horizontal aperture function that extended beyond the boundaries of each sensing region (see horizontal aperture function in Fig 7). As a result, each actuator was slightly responsive to neighbouring sensing regions. The overlap and center-weighting, combined with natural exploratory motions of the user's head, gave some sub-pixel accuracy that allowed the user to sense some degree of fine detail.

## 3.6 Compensating for non-linear behaviour of motors and human perception

One challenge of our design is to convert the depth map values to another sensory mode such as tactile feedback using an actuator.

It is clear to see that using a linear model for mapping raw depth value to the actuator is inadequate for several reasons. First, the linear model does not handle the non-linearity in human perception. For many of the sensory modalities, our sensory perceptions are non-linear and have a highly compressive nature. For example, humans perceive loudness of sound in a logarithmic scale. This logarithmic scale recurs often in the human senses and comes from Weber's analysis of "just-noticeable differences" [**?**]. A perceived sensation $P$ results from fractional changes in a physical quantity $I$ as in:

$$\Delta P \propto \frac{\Delta I}{I} \qquad (6)$$

After setting $P = 0$ at the minimum perceptible physical quantity $I_0$, the solution becomes:

$$P = k \log\left(\frac{I}{I_0}\right) \qquad (7)$$

Weber found this logarithmic law to exist for the sense of touch [**?**]. Additionally, the raw depth data collected from the kinect are not a direct measurement of the actual distance in the real world, and a reconstruction of the actual depth value is required for fine calibration.

Since the non-linearities and the underlying interactions between the actuator and human perception are difficult to recover, we estimated these relationships experimentally by perform trials on different users. We have found that using an exponential decade function as follows provides adequate results, which also comforms with the non-linear relationship between human sensory and distance information we conjured previously.

$$w = (0.978)^{255d} \qquad (8)$$

where $d$ is the actual distance normalized to 1 with the maximum range, and $w$ the PWM (Pulse-width modulation) value which controls the electrical actuator.

Figure 8 shows the conversions and compensation we have introduced in the signal flow path. Notice that our system has aggregated the inverse of the non-linear responses of

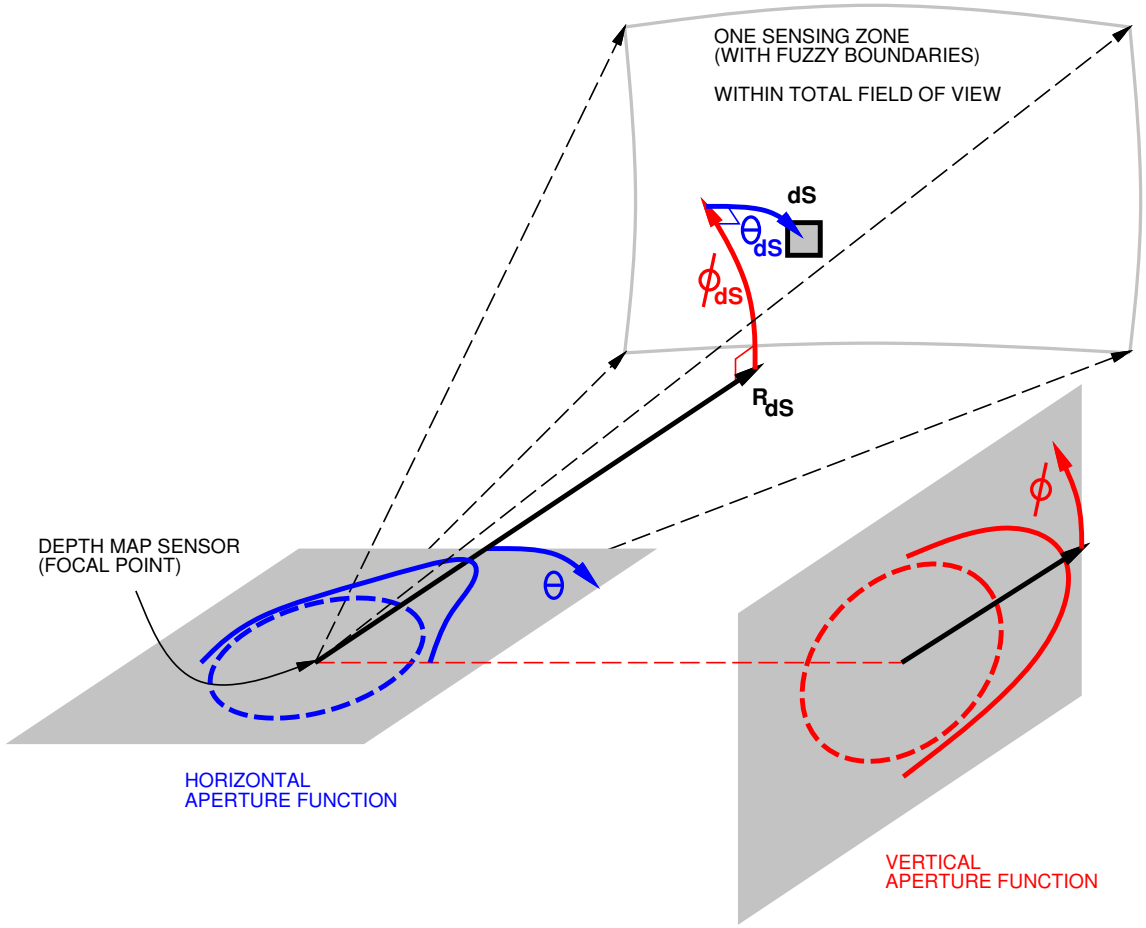


**Figure 7: The viewable area of the depth sensor was divided into zones, each of which corresponded to one of the actuator motors. The sensitivity profile was center-weighted as illustrated.**

the motor and electronics as well as human perception for simplicity.

With the proper calibration and compensation for non-linearity and sensory thresholds, users were able to learn the relationship between the distance and the vibration intensity after several minutes of training with the system.

### 3.7 2-dimensional mapping

In further variations of the system, we implemented various 2-dimensional arrays, such as a 3 by 4 array of 12 actuators, and a 2 by 6 array of 12 actuators (the Arduino has 12 PWM outputs). In further explorations, we also experimented with larger arrays using multiple microcontrollers. However, we found that a small number of actuators was often sufficient.

## 4. SUPPLEMENTARY VIDEO MATERIAL

Videos of the helmet in action can be viewed at:

`http://wearcam.org/blindvision/`

## 5. CONCLUSION

We have proposed a novel way of using the Microsoft Kinect 3-D camera, for navigation which we hope will someday assist the visually impaired.

Rather than having the Kinect observe a user, we put the Kinect on the user to observe the user's environment.

We found that the typical operating range of the Kinect (30cm to 6m) was well suited to indoor navigation in typical crowded corridors, and the like.

This preliminary work suggests that eyeglasses could be made using PrimeSense's 3-D sensing technology, for potential use by the visually impaired.

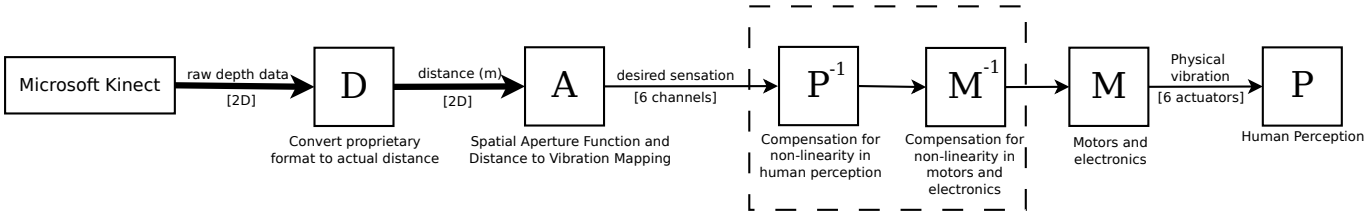


**Figure 8: This figure shows the conversion of the depth map data from the Microsoft Kinect to the actual physical vibration of the 6 actuators in the helmet. The underlying non-linear relationships in the raw depth sensor, motor and electronics, and human perceptions are estimated empirically. By aggregating the $L^{(-1)}$ and $P^{(-1)}$ functions, we can determine the mapping of the vibrating intensity to the optimal sensitivity range of human senses experimentally.**

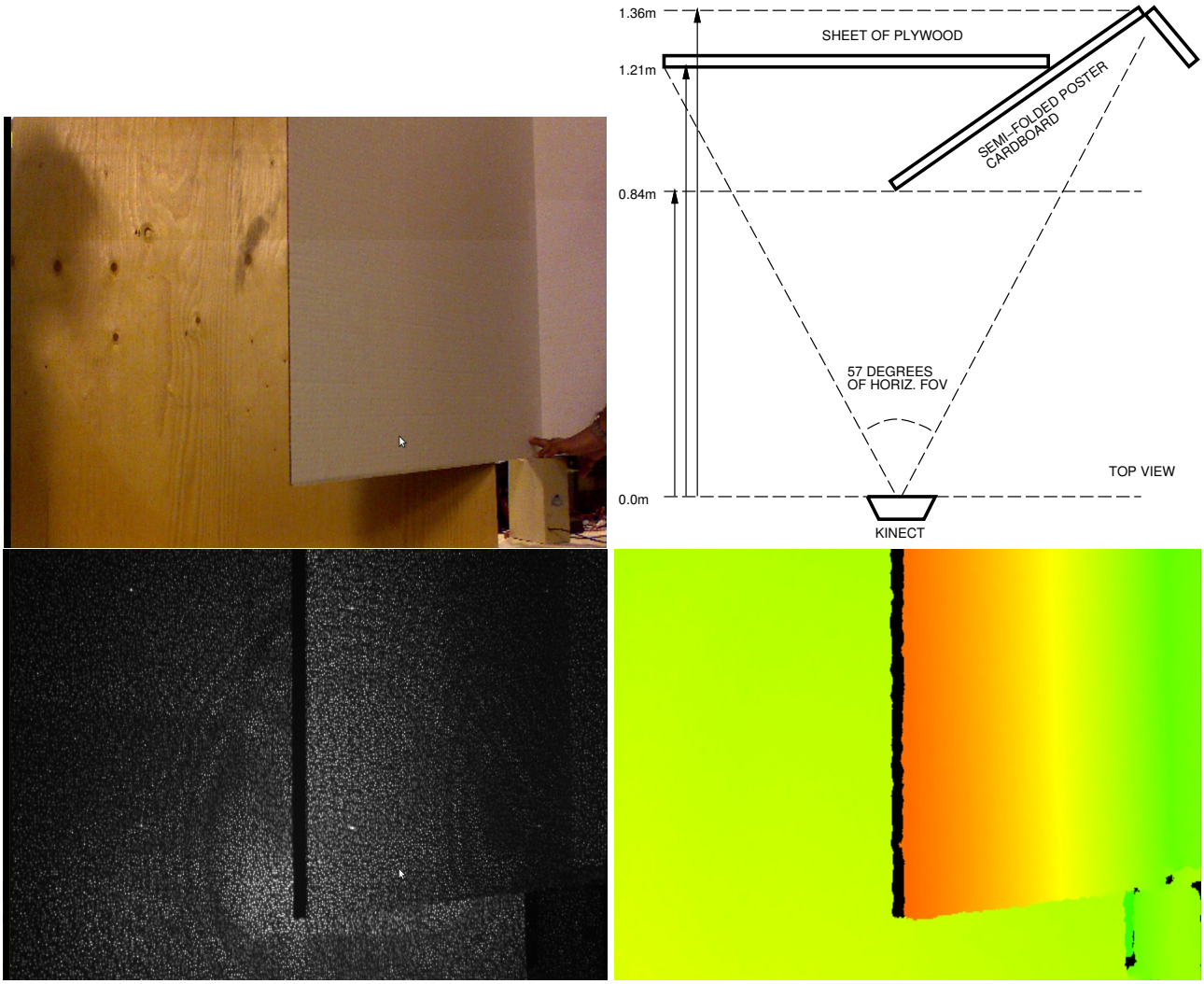


**Figure 9: Example with plywood and cardboard**