

Sensitivity Evaluation of Embedded Code Detection in Imperceptible Structured Light Sensing

Jingwen Dai and Ronald Chung

Department of Mechanical and Automation Engineering
The Chinese University of Hong Kong, Shatin, NT, Hong Kong

{jwdai, rchung}@cuhk.edu.hk

Abstract

We address the use of pre-trained primitive-shape detectors for identifying embedded codes in imperceptible structured light (ISL) sensing. The accuracy of the whole sensing system is determined by the performance of such detectors. In training-based methods, generalization of the training results is often an issue, and it is especially so when the work scenario could have substantial variation between the training stage and the operation stage. This paper presents sensitivity evaluation results of embedded code detection in ISL sensing, together with the associated statistical analysis. They show that the scheme of embedding imperceptible codes into normal video projection can be maintained effective despite possible variations on sensing distance, projection-surface orientation, projection-surface shape, projection-surface texture and hardware configuration. The finding indicates the feasibility of integrating the ISL method into robotic systems for operation over a wide domain of circumstances.

1. Introduction

Vision sensing has been widely applied in the area of robotics, such as visual SLAM, measurement, human robot interaction and visual servo. The improving performance and declining price of micro / pico projectors make it possible to use them on the robots. Assisted with cameras, projector-camera systems (PROCAMS) have emerged in many robotic platforms [7, 6, 8] for augmented reality (AR), human-robot interaction (HRI) and some other applications.

The adoption of structured light illumination has been proven to be an effective and accurate visual means for 3D reconstruction. With the help of 3D information, it is more accurate and efficient for robot to understand surrounding environment and to interact with human users. If some structured patterns could be embedded into normal video projection nonintrusively and imperceptibly, the projector

could serve as both a display device to show vivid video content and a 3D sensor to derive the 3D information of the objects near it. This leads to the concept of Imperceptible Structured Light (ISL) [10].

In ISL, code embedded images and their complements were projected temporally in high frequency. Due to limitation of human visual perception, the embedded code patterns can be made undetectable to the user, but cameras synchronized to the modulation are able to reconstruct the embedded codes for structured light sensing. Many researcher focused on how to determine the embedded intensity properly to guarantee the code imperceptibility [2, 3, 1]. However, few works focus on the decoding method in imperceptible code embedding configuration, especially when huge external noises exist.

In our previous work [4, 5], we proposed a novel approach to resolve the conflict between imperceptibility of the embedded codes and the robustness of code retrieval. We introduced noise-tolerant schemes to both the coding and decoding stages. At the coding end, specifically designed primitive shapes (cross, rhombus and sandglass) and large Hamming distance are employed to enhance tolerance toward noise. At the decoding end, pre-trained primitive shape detectors are used to detect and identify the embedded codes – a task difficult to achieve by segmentation that is used in general structured light methods, for the weakly embedded information is generally interfered by substantial noise. Extensive experiments including evaluations of both code imperceptibility and decoding accuracy show the effectiveness of the proposed system. Some applications in robotics, including sensing surrounding environment and touch-based human robot interaction on projection plane surface, were also illustrated in [4].

It is obvious that the performance of this method depends on the accuracy of pre-trained primitive shape detectors, which is determined by the training process to a great extent. Generally, for the training based methods, generalization of the training results is an issue, especially, when the scenarios between training stage and operation stage are quite

different. For applications on robotics, due to the different sensor-object localization, different projection surfaces, different surrounding environment and different hardware platforms, the generalization of the pre-trained detector is of great importance, since it is impractical even impossible to re-train the detector for different scenarios.

This paper aims at certifying the validity of our previous work in different robotic applications by evaluating the sensitivity of primitive detectors under different circumstances, including variations on sensing distance, projection surface orientation, projection surface shape, projection surface texture and hardware configuration.

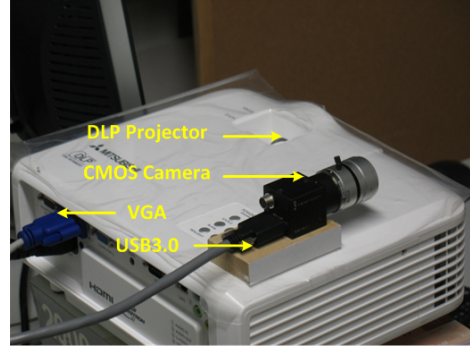
The remainder of this paper is structured as follows. In the next section, the hardware configuration of two projector-camera systems for evaluation is described. In Section 3, the benchmark of evaluation is depicted. The sensitivity analyses for primitive detectors under different circumstances are detailed in Section 4. Conclusion are offered in Section 5.

2. System Setup

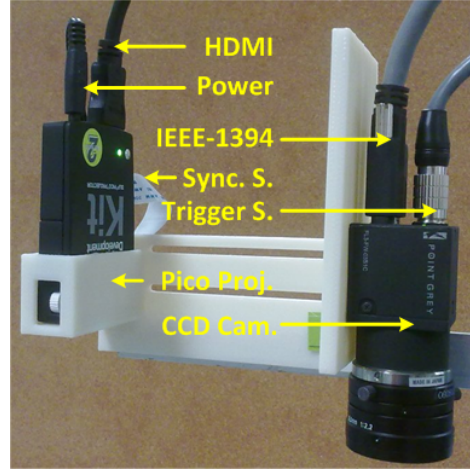
In order to evaluate the performance of primitive shape detector in different platforms, we set up two projector-camera systems using different equipments. The first one (*PROCAMS-A*) consisted of a consumer-level DLP projector (Mitsubishi EX240U projector) of 1024×768 resolution and $120Hz$ refresh rate, and a CMOS camera (Point Grey Flea 3 FL3-U3-13S2C with Myutron FV1520 f15mm lens) of 1328×1048 resolution and $120fps$. While the second one (*PROCAMS-B*) consisted of a Pico DLP projector with a native resolution of 640×480 and an interface for firmware configuration (TI DLP Pico Projector Development Kit 2), plus a CCD camera of 648×488 resolution at $120fps$ (Point Grey FL3-FW-03S1C camera with Myutron FV0622 f6mm lens).

For *PROCAMS-A*, we first fixed the camera and projector rigidly, and the projector and camera were connected to a desktop computer through VGA and USB3.0 interfaces respectively. Since there was no synchronization signal output in the consumer-level projector, the synchronization between projectors and cameras was implemented through software delay. The hardware configuration is shown in Fig. 1(a). For *PROCAMS-B*, the projector and camera were mounted on a special designed framework rigidly, and were connected to a laptop computer through HDMI and IEEE-1394 interfaces respectively, and the hardware trigger signal of the camera was connected to the sync. output of the projector for synchronization between them, which are illustrated in Fig. 1(b).

Moreover, the projector-camera systems were calibrated using an LCD monitor as the calibration object; the calibration method, detailed in [11], could derive the intrinsic and extrinsic parameters of the two instruments. Once the



(a) PROCAMS-A



(b) PROCAMS-B

Figure 1. Hardware configuration of two projector-camera systems.

experimental system was set up and calibrated, we could conduct further experiments.

3. Benchmark

As described in [4], in decoding stage, to resolve the low signal-to-noise ratio problem, we regard the primitive shapes as objects to "identify" and "detect" rather than "segment". The primitive shape detectors were trained through the approach proposed in [9] for its capability of processing images rapidly with high detection rates.

To collect the training samples, 500 color images with different contents were collected from Google Image, and 40 primitive shapes were embedded in those images at different positions to generate 500 pairs of projected images and complementary images. By projecting them to a white planar projection screen with small orientation variations, 500 subtraction images could be derived from image capture exercises. The sub-images containing primitive shapes were considered as positive training samples. The background were divided into small patches to generate negative

training samples. More detailed training procedures can be found in [4].

Another 500 color images were collected, through embedding-projection-capture cycle, 500 subtraction images for detector accuracy evaluation were generated. The ground-truth was obtained by manual labeling in the image data captured under binary pattern illumination. The scenario of testing sample generation was the same as training sample collection stage. Thus, this test results are considered as the benchmark for the sensitivity analysis in the next step. The performance of primitive detector are evaluated by hit rate (H), missing rate (M), false rate (F) and position error (E_d), which are formulated as

$$H = \frac{N_h}{N_t}, M = \frac{N_m}{N_t}, F = \frac{N_f}{N_t}, \quad (1)$$

$$E_d = \sqrt{\epsilon_X^2 + \epsilon_Y^2}, \quad (2)$$

$$\epsilon_X = \frac{1}{N_h} \sum_{i=1}^N |X_d - X_g|_i, \quad (3)$$

$$\epsilon_Y = \frac{1}{N_h} \sum_{i=1}^N |Y_d - Y_g|_i, \quad (4)$$

where N_t is the total embedded primitive shape number, N_h , N_m and N_f are the number of correct detections, missed detections and false detections respectively. ϵ_X and ϵ_Y are the average feature point detection errors along the x-axis and y-axis, (X_d, Y_d) and (X_g, Y_g) are the detected coordinate and ground-truth respectively. Qualitative results in some subtraction images¹ are presented in Fig. 2. The more detailed quantitative testing results are listed in Table 1. It is clear that the primitive shape detectors have excellent performance when the environment of operation stage is the same as that of training stage.

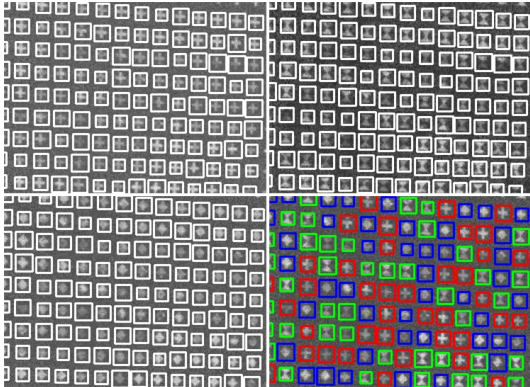


Figure 2. Some qualitative benchmark results.

¹All the subtraction images in this article are scaled to $[0, 255]$ for illustration purpose.

	H(%)	M(%)	F(%)	E_d (pixel)
C	94.53	3.95	1.52	1.632
R	95.21	3.59	1.20	1.833
S	95.50	3.63	0.87	1.542

Table 1. Benchmark for sensitivity evaluation.

4. Sensitivity Analysis

From the benchmark testing in Section 3, about 95% primitive shapes can be detected and identified correctly by pre-trained detectors, when the scenarios of training stage and operation stage are almost the same. For the applications in robotics, especially for mobile robot, it is impossible to have an operation circumstance that is the same as training stage. Therefore, the sensitivity and generalization of the pre-trained detector under different circumstances are of great importance for the performance of whole sensing system. In this section, a series sensitivity analysis will be conducted when some environmental factors are changed.

4.1. Working Distance

The working distance is the average distance from the projector-camera system to the object surface. When the intrinsic parameters of the projector and camera (focal length and resolution) are fixed, the size of the primitive shapes in subtraction image is determined by the working distance directly. In the configuration of training stage, the working distance is set as 800mm, the size of primitive shapes in image data is about 20 pixels. In the operation stage, the working distance is changed to 500mm, 1200mm and 1600mm, the focal length of procams is slightly adjusted to get sharp projection and clear capture. Some subtraction images with detection results are shown in Fig. 3, the size of the primitive shapes are around 15, 35 and 45 pixels respectively.

The detailed quantitative results are listed in Table 2. It is clear that when the working distance decreased to 500mm, the hit rates dropped, because it is difficult for primitive shape detectors to find small size shapes in image data. For the enlarged shapes in larger working distance, the performance of detectors are almost the same as the benchmark.

4.2. Projection Surface Orientation

Besides the size of the primitive shapes in image data, the distortions will also influence the performance of the pre-trained detectors. The distortions mainly come from the variations on the orientation of the projection surface w.r.t. the sensing system and the variations on the shape of the projection surface. First, the detector accuracy will be evaluated under different projection surface orientations.

In training data collection stage, the images are projected to a planar surface that is almost parallel to the image plane

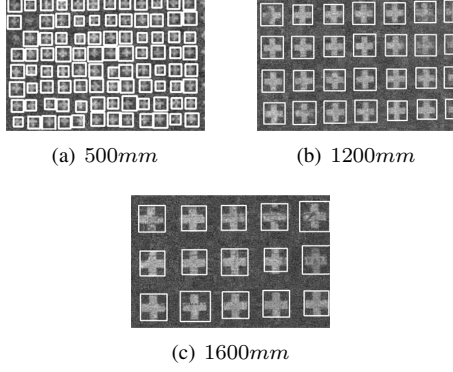


Figure 3. Cross shape detection in different working distances.

Distance	Pri.	H(%)	M(%)	F(%)	$E_d(\text{pixel})$
500mm	C	86.21	11.63	2.16	1.814
	R	85.83	12.57	1.60	1.836
	S	87.49	11.64	0.87	1.712
1200mm	C	94.44	4.32	1.24	1.728
	R	94.86	4.23	0.91	1.904
	S	94.49	4.62	0.89	1.572
1600mm	C	94.52	4.11	1.37	1.731
	R	95.06	3.92	1.02	1.910
	S	95.39	3.68	0.93	1.591

Table 2. Primitive shape detection accuracy in different working distance.

of the camera. Now in operation stage, the orientation of the surface is adjusted to 10° , 20° , 30° , 40° , 50° in yaw direction, as shown in Fig. 4. In each sub-image, the upper part is the captured image to show the extent of distortion, while the lower part is the magnified subtraction image of the subregion indicated by the rectangle in captured image. The detection results are also shown in the subtraction images. More detailed quantitative results are listed in Table 3.

In the testing results, when the rotation degree θ is small, i.e., $\theta = 10^\circ, 20^\circ$, the performance is almost the same as benchmark. With the increase of the rotation degree, the hit rates decrease slightly. When $\theta = 50^\circ$, more than 85% primitive shapes are still detected correctly, which satisfies the application requirements [4].

4.3. Projection Surface Shape

The alteration of projection surface shape will also result in the distortion of primitive shapes in image data. In training stage, the negative and positive sample were collected from the images projected to a planar surface. In this test, the projection surface are three different non-planar surfaces (convex paper, concave paper and plaster statue). Some test images and the statistical results are shown in Fig. 5 and Table 4 respectively. In all three surfaces, although the hit rates have small decrease, it is still sufficient

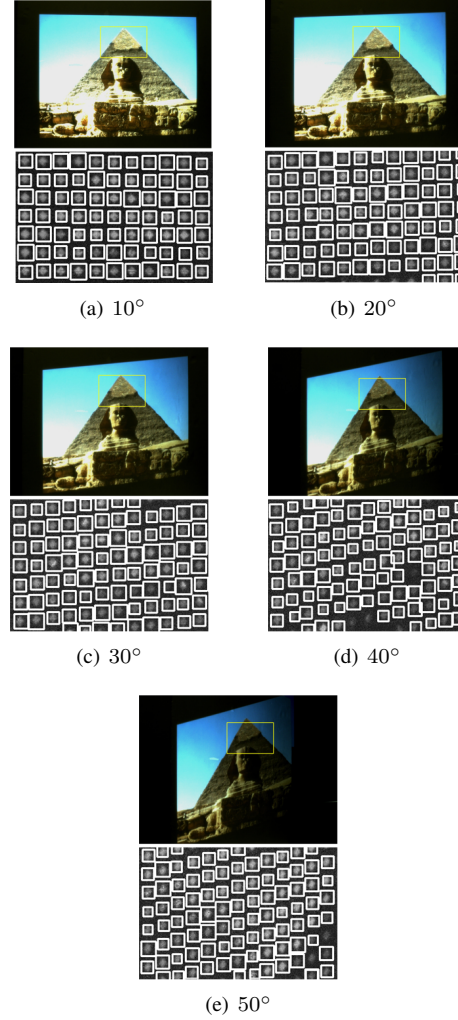


Figure 4. Rhombus shape detection in the projection surface with different orientations.

Orientation	Pri.	H(%)	M(%)	F(%)	$E_d(\text{pixel})$
10°	C	94.51	3.96	1.53	1.635
	R	95.08	3.60	1.22	1.845
	S	95.46	3.74	0.80	1.544
20°	C	94.50	3.96	1.54	1.634
	R	95.08	3.64	1.08	1.848
	S	95.43	3.77	0.80	1.564
30°	C	93.47	4.50	2.03	1.938
	R	92.15	6.37	1.48	2.141
	S	92.43	6.78	0.79	2.011
40°	C	90.19	7.70	2.11	2.414
	R	89.42	9.50	1.08	2.809
	S	91.23	7.87	0.90	2.374
50°	C	86.21	11.63	2.16	2.728
	R	85.83	12.57	1.60	2.904
	S	86.87	12.27	0.86	2.572

Table 3. Primitive shape detection accuracy in projection surface with different orientations.

Surface	Pri.	H(%)	M(%)	F(%)	$E_d(\text{pixel})$
Convex	C	93.53	4.86	1.61	1.756
	R	93.25	5.29	1.46	2.043
	S	94.14	4.85	1.01	2.122
Concave	C	93.64	4.84	1.52	1.762
	R	93.82	4.70	1.48	2.108
	S	93.76	5.41	0.83	2.135
Free-Form	C	84.81	13.33	1.86	2.028
	R	85.73	13.06	1.21	1.904
	S	86.09	13.03	0.88	2.075

Table 4. Primitive shape detection accuracy in projection surface with different shapes.

Texture	Pri.	H(%)	M(%)	F(%)	$E_d(\text{pixel})$
Green Paper	C	94.41	4.17	1.42	1.634
	R	95.19	3.66	1.15	1.836
	S	95.49	3.63	0.88	1.558
Cork Board	C	93.41	5.07	1.52	1.641
	R	94.25	4.43	1.32	1.850
	S	94.92	4.16	0.92	1.623
Poster	C	91.74	6.63	1.63	2.024
	R	90.28	8.25	1.47	1.996
	S	92.19	6.76	1.05	1.762

Table 5. Primitive shape detection accuracy in projection surface with different textures.

to derive correct correspondences for triangulation. In the plaster statue case, the missing detections are mainly found in the regions where the surface has sudden change.

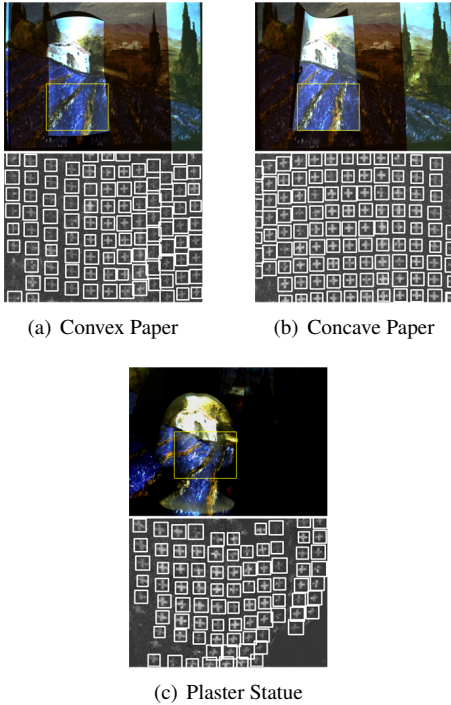


Figure 5. Cross shape detection in different projection surfaces.

4.4. Projection Surface Texture

The texture on the projection surface will affect the quality of captured images. In the benchmark training stage, the projection surface is textless and in white color. In the operation stage for test, the images are projected to a planar surface in green color, a cork board and a poster with text and images, as illustrated in Fig. 6. The quantitative results are listed in Table 5. The results indicate that the texture variation on the projection surface has little influence on the performance of primitive shape detectors, since in our

method [5] the decoding process was conducted in subtraction image, which would weaken the texture influence to a certain extent.

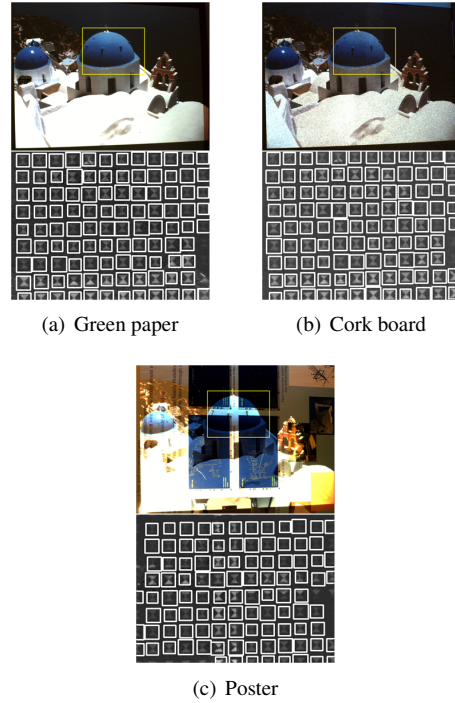


Figure 6. Sandglass shape detection in different projection surface textures.

4.5. Projector-Camera System

If the pre-trained detectors are used in another robot systems with different hardware configuration, the performance of the detectors would be affected, since the differences in the resolution of projector and camera (high vs. low), the camera sensor (CCD vs. CMOS) and the optical parameters (different lens) will change the appearance of the primitive shape in image data. In this test, the primitive detectors trained by the data collected from *PROCAMS-A*

are applied in *PROCAMS-B* during the operation stage.

Due to the low projector resolution in *PROCAMS-B*, the dimension of the original pattern image is too large for embedding, so we employ two method to solve this issue, the first one is to select a sub-region of the original pattern image as a new pattern image and the second one is to resize the original pattern image to coincide the projector resolution. Some detection results in the subtraction images derived from two different embedding methods are illustrated in Fig. 7(b) and 7(c). The quantitative results are also shown in Table 6.

Compared with the benchmark, it is obvious that the performance in *PROCAMS-B* degrades intensively, especially in the resized pattern case. By analyzing the missed and false detection cases, we find that the mistakes were mainly caused by large noise from the low luminance of the pico projector and the extremely small primitive shapes in image data.

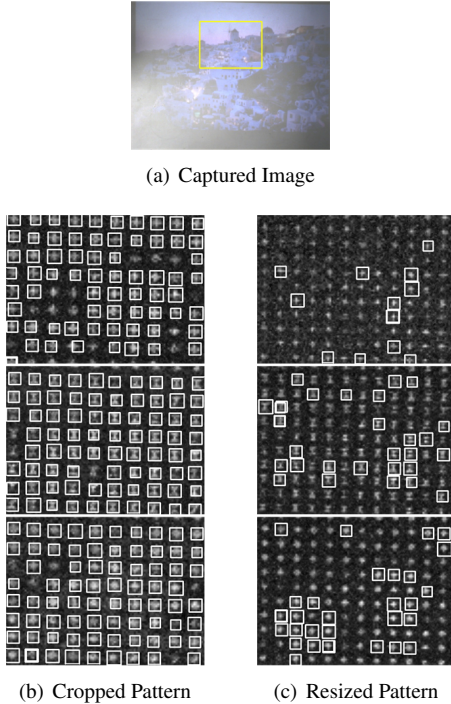


Figure 7. Primitive shape detection in *PROCAMS-B* with different embedding approaches.

5. Conclusion

We have presented sensitivity evaluation results of embedded code detection in imperceptible structured light sensing, together with the statistical analysis. They show that embedding imperceptible codes into normal video projection can be made effective even though there could be variations on sensing distance, projection-surface orientation, projection-surface shape, projection-surface texture

	Pri.	H(%)	M(%)	F(%)	$E_d(\text{pixel})$
Cropped Pat.	C	80.23	14.43	5.34	3.028
	R	79.93	14.17	5.92	2.981
	S	81.09	13.28	5.63	2.812
Resized Pat.	C	30.52	59.23	10.25	2.628
	R	30.63	58.03	11.34	2.913
	S	30.80	57.93	11.27	2.874

Table 6. Primitive shape detection accuracy in *PROCAMS-B* with different embedding approaches.

and hardware configuration. Our future work will be about extending imperceptible structured light sensing to a variety of robotic applications.

References

- [1] A. Grundhofer, M. Seeger, F. Hantsch and O. Bimber. Dynamic adaptation of projected imperceptible codes. In *Proc. of IEEE/ACM ISMAR*, pages 1–10, 2007.
- [2] Hanhoon Park, Byung-Kuk Seo and Jong-Il Park. Subjective evaluation on visual perceptibility of embedding complementary patterns for nonintrusive projection-based augmented reality. *IEEE Transactions on Circuits and Systems for Video Technology*, 20(5):687–696, 2010.
- [3] Hanhoon Park, Moon-Hyun Lee, Byung-Kuk Seo, Yoonjong Jin and Jong-Il Park. Content adaptive embedding of complementary patterns for nonintrusive direct-projected augmented reality. In *HCI international*, pages 132–141, 2007.
- [4] J. Dai and R. Chung. Embedding imperceptible codes into video projection and applications in robotics. In *Proc. of IEEE IROS*, 2012.
- [5] J. Dai and R. Chung. On making projector both a display device and a 3d sensor. In *Proc. of ISVC*, pages 654–664, 2012.
- [6] J. Park and J. Kim. Robots with projectors: an alternative to anthropomorphic hri. In *Proc. of ACM/IEEE HRI*, pages 221–222, 2009.
- [7] T. Matsumaru. Mobile robot with preliminary-announcement and display function of forthcoming motion using projection equipment. In *Proc. of IEEE ROMAN*, pages 443–450, 2006.
- [8] N. Linder and P. Maes. Luminar: portable robotic augmented reality interface design and prototype. In *Proc. of ACM UIST*, pages 395–396, 2010.
- [9] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. of IEEE CVPR*, pages 511–518, 2001.
- [10] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs. The office of the future: A unified approach to image-based modeling and spatially immersive displays. In *Proceedings of SIGGRAPH 98*, pages 179–188, 1998.
- [11] Z. Song and R. Chung. Use of LCD panel for calibrating structured-light-based range sensing system. *IEEE Transactions on Instrumentation and Measurement*, 57(11):2623–2630, 2008.