IEIE Transactions on Smart Processing and Computing, vol. 4, no. 6, December 2015
http://dx.doi.org/10.5573/IEIESPC.2015.4.6.379

Distance Measurement Using the Kinect Sensor with Neuro-image Processing

Kajal Sharma

Flat No 301, Building 206, Piorville Apartment, 24 Namyang-Dong, Seongsangu, Changwon, Korea kajal175@gmail.com

***** Corresponding Author:

Received October 14, 2015; Revised November 10, 2015; Accepted December 10, 2015; Published December 31, 2015

** Short Paper*

***Abstract*:** This paper presents an approach to detect object distance with the use of the recently developed low-cost Kinect sensor. The technique is based on Kinect color depth–image processing and can be used to design various computer-vision applications, such as object recognition, video surveillance, and autonomous path finding. The proposed technique uses keypoint feature detection in the Kinect depth image and advantages of depth pixels to directly obtain the feature distance in the depth images. This highly reduces the computational overhead and obtains the pixel distance in the Kinect captured images.

***Keywords*:** Kinect sensor, Neural network, Distance estimation

1. Introduction

Feature calculation and distance estimation is a key technique in many vision applications, such as video surveillance and autonomous path finding [1, 2]. In this paper, an efficient Kinect-based distance estimation technique is proposed, which uses color processing of depth image pixels and is efficient at estimating the distance of autonomous toy vehicles. In the literature, many algorithms determine object position in real time in image sequences [3, 4]. Other feature-detection approaches obtain the features of an object in tracking applications [5-8], but feature-based distance estimation with a color depth image remains an unsolved problem. Thus, an efficient and accurate distance estimation approach is described in this paper. A brief overview of the Kinect sensor and a flow diagram of the proposed approach are given in Section 2. Section 3 discusses the proposed method to detect object distance with color image processing. In sections 4 and 5, respectively, the results of the proposed method and the conclusion are given.

2. Kinect Sensor Overview

The Kinect sensor used for research purposes consists of an RGB camera, a depth sensor and a multi-array microphone running proprietary software, which provide full-body 3D motion capture, facial recognition and voice recognition capabilities. A flow diagram of the proposed approach is in Fig. 1.

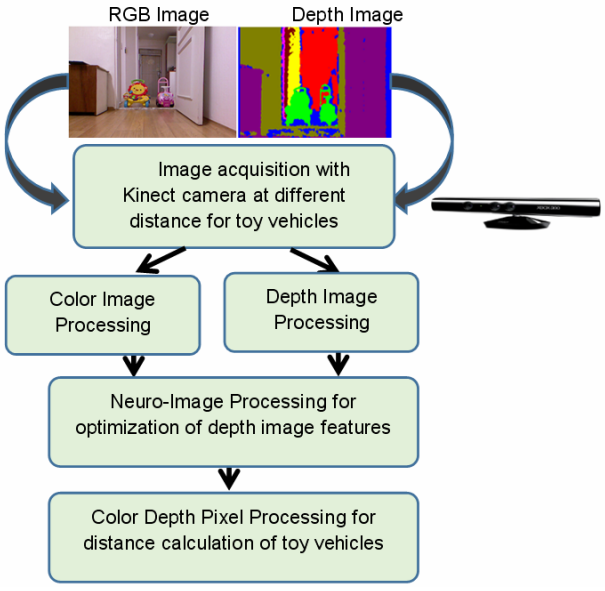


Fig. 1. Flow diagram of the proposed approach, and a Kinect overview.

The depth sensor uses infrared signals to create a digital 3-D model of an object, and the color camera captures rich visual information. However, it is extremely hard to extract dense depth from camera data alone, especially in indoor environments with very dark or sparsely textured areas. RGB-D cameras are sensing systems that capture RGB images along with per-pixel depth information. In the proposed method, depth estimation is done with the use of a Kinect sensor, and the distance of toy vehicles is determined by using the different colors.

3. Proposed Method to Detect toy Vehicle Distance with Kinect Depth-image Processing

The proposed approach presents a new and fast vehicle-distance calculation method in which the features are optimized with an unsupervised neural network. The proposed method presents unsupervised feature selection and category classification for application to a vision-based path-finding system. Fig. 2 shows different types of images that can be captured with a Kinect sensor: RGB images, depth images, infrared images, and color depth images. The network architecture of the proposed method extracts the feature points of the toy vehicles and calculates the descriptors using scale-invariant feature transform (SIFT). For any image, the SIFT algorithm is implemented by using a difference-of-Gaussian function to obtain the interest points that are invariant to scale and orientation. SIFT replaces the images with a set of scale- and orientation-invariant feature descriptors using gradient-orientation histograms.

For the selected keypoints, orientations are assigned to each keypoint location, based on local image gradient directions. SIFT divides the image region into 4 × 4 sub-regions, and sums the gradient strength in each sub-region. SIFT uses eight directions in each sub-region to generate an eight-dimensional vector. The local image gradients are transformed into 128 dimensional representations resulting in a keypoint descriptor. A SIFT descriptor is a 3-D spatial histogram of the image gradients, and each pixel gradient is formed by pixel location and gradient orientation. These descriptor vectors are passed to the feature-matching step with the proposed neuro-image processing to match the keypoints of different objects in different distance images.

The SIFT features are optimized using a self-organizing map (SOM), where all the SIFT descriptors and histograms of the selected SIFT descriptors are reduced and matched using SOMs. The optimization and reduction of the features of the toy vehicles are done with winning pixel estimation in the color-depth image. Features of toy vehicle images from mobile vehicles change with time; thus, the proposed method enables an unsupervised feature classification that requires no parameter setting for the number of category classifications. The output of the SOM network is represented as a topological category map on the Kohonen layer, and the bag of features (BOG) represents the selected optimized features in the image set.

Fig. 2. Different types of image that can be captured with a Kinect sensor: RGB images, depth images, infrared images, and color-depth images.

The features are clustered with the SOM, where the distance between the input feature vector and the weight vector is computed by

$$d_k(t) = \left\| x(t) - w_k(t) \right\| \text{ where } k = 1 \ldots n \tag{1}$$

where x is the input vector, w is the weight vector, and d denotes the distance vector (usually the Euclidean distance); n denotes the number of mapped features into a low-dimension gridmap. The best-matched neuron (bmn) is obtained by the estimation of minimum distance with the input feature vector and is calculated with (2):

$$d_{bmn}(t) = \min_k (d_k(t)) \tag{2}$$

Neuro-optimization is done in step 1 of Eq. (1) in terms of computational time; if the computation in Eq. (1) is done only on the nonzero values, the computation overhead is highly reduced. The 128-dimension descriptor vector is passed to Eq. (1) in order to pass for neuro-optimization, and only the winning pixels contribute in the matching process. The resulting descriptor set consists of a low-dimension winning feature vector in a dataset of different distance images. Eq. (1) can be recomputed with the following:

$$d_k(t) = \sum_{x_i \,!=0} x_i(t)(x_i(t) - 2w_{ki}(t)) + \sum_{i=1}^{n} w_{ki}(t)^2 \tag{3}$$

The computation overhead is reduced by taking into consideration only the nonzero feature values. The complexity is reduced from $O(N^2 * n)$ to $O(N^2 * f_{non\ zero} * n)$, and computation is done over the nonzero feature input vectors. The running time is reduced to nonzero values, and only $2 * xi(t) * w_{ki}(t)$ computations are required in each iteration; thus, the overall complexity is reduced to $O(N^2 * f_{non\ zero} * n)$.

The computational overhead can be further reduced in

Table 1. The Kinect camera parameters set for the experiments.

Parameters	Values
Image type	COLOR, DEPTH, INFRARED
Key components	Color camera, infrared (IR) emitter, IR depth sensor, tilt motor, microphone array, LED
Field of view	43 degrees vertical by 57 degrees
Color- and depth-image resolution	640 x 480 pixels
Frame speed	30 frames per second (FPS)
Tilt motor	Shifted upwards or downwards by 27 degrees
LED	Green LED indicates that the Kinect device drivers loaded properly

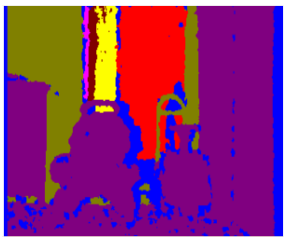
the neighborhood function calculation step by considering only the first three neurons having the shortest distance from the input feature vector. The contribution of the third neuron neighborhood feature element is less than 1/9, which reduces the overall computation, and thus, the update neighborhood phase is decreased to $O(N^2)$ with a constant of 6, since three weight vector columns and their squared components need to be updated.

The proposed method generates stable features, which contribute in path-finding for the autonomous toy vehicles. The distances of the toy vehicles are further estimated with the proposed Kinect color processing, and details with the results are given in the next section. The main contribution of this paper is the color processing of the depth pixels from the Kinect camera. The features obtained with the matching and neural-network reduction are stored in the database. The features obtained are invariant, and stable features are assigned distances based on the color processing of the depth pixels. With the neuro-image processing, only the invariant stable features are obtained, which are used to calculate the object distance from the Kinect camera. The proposed distance measurement from the camera for toy vehicles, and different colors assigned to the depth image, are detailed in Table 2. Thus, the neuro-optimized feature distance is obtained with the colors. Blue shows the near features, whereas yellow represents the far features. Thus, the vehicle position is determined with the colors, and the calculated distance is used to determine the vehicle location.

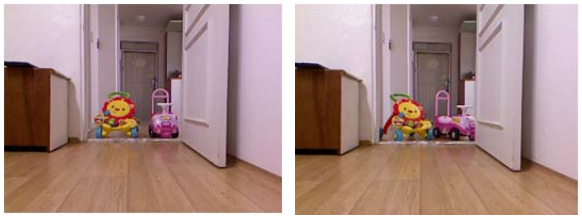
4. Experimental Results

The Kinect sensor device developed by Microsoft is used for toy-vehicle navigation in order to estimate the object depth with the Kinect methodology. Real-time images are captured with the Kinect sensor, and the corresponding color-depth image is captured. Stable features are extracted from the image pair of toy vehicles with a scale-invariant feature-matching technique. A color-depth image is generated that consists of color pixels having different distance values from the Kinect camera.

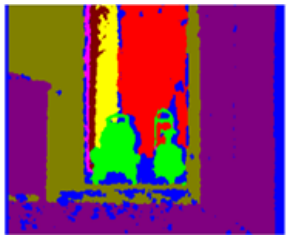
(a) Image pair captured with Kinect camera at different angles to obtain invariant features



(b) Depth image obtained with depth-pixel processing; purple indicates the distances of toy vehicle 1 and toy vehicle 2



(c) Image pair captured with Kinect camera, where toy vehicles are at a far distance



(d) Depth image obtained with depth-pixel processing; green indicates the distances of toy vehicle 1 and toy vehicle 2

Fig. 3. Results of the proposed algorithm: (a) and (b) show results of a captured image pair close to the camera, (c) and (d) show results of a captured image pair far from the camera; color in the depth image is used to estimate the distances of the toy vehicles.

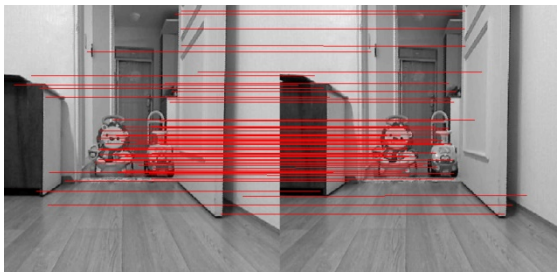
Table 1 shows the Kinect camera parameters that were set for the experiments. Fig. 3 shows the results of the proposed algorithm.

Redundant features and unstable features are removed with the neuro-optimization method given in Section 3. The output image consists of a depth image with object distance information, which is used for autonomous toy-vehicle navigation in terms of path finding. Table 2 shows the results of the proposed method, where the different colors are used to generate the color-depth image, and the distance is obtained with color-image processing.

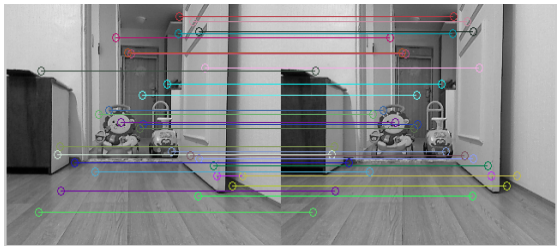
The proposed method was compared with recent matching techniques for features in images. For the experiments, a dataset of images at different distances was

Table 2. Distance measurements from the camera for toy vehicles; different colors in the depth image represent the objects' distances from the camera.

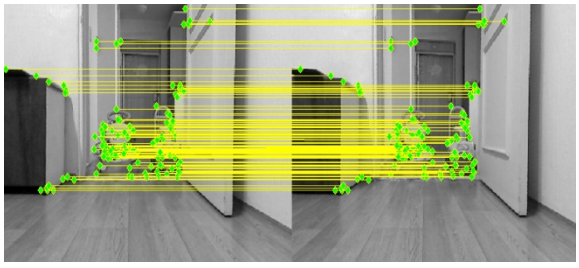
	Color and distance calculation with depth pixels	
	Color	Distance
Toy Vehicle Image 1	**Blue**	**Less than 1 meter**
Toy Vehicle Image 2	Purple	2 meters
Toy Vehicle Image 3	Green	3 meters
Toy Vehicle Image 4	Lime	4 meters
Toy Vehicle Image 5	Fuchsia	5 meters
Toy Vehicle Image 6	Yellow	6 meters



(a) The results of SIFT feature matching in image 5 and image 6



(b) The results of SURF feature matching in image 5 and image 6



(c) The results of the proposed neuro-image matching in image 5 and image 6

Fig. 4. The results of matching with SIFT, SURF, and the proposed method for toy vehicle image 5 and image 6, *dist = 2m*. The lines in the images show the matching features in the images.

captured with the Kinect sensor. The proposed method was compared with both SIFT and speeded up robust features (SURF) matching methods (Fig. 4). The results are shown in Table 3. The computation time for feature matching at different distances, ranging from 0.5 m to 6 m, are shown with different image pairs. The matched features are computed with SIFT, SURF, and the proposed method. The comparison results show the proposed method

Table 3. Computation time of feature matching with SIFT, SURF, and the proposed method

Dataset of image pair at different distances	Matching time of features in seconds		
	SIFT	SURF	Proposed method
Toy vehicle images 1 and 2; dist = 0.5 m	3.405934	2.575838	**0.033820**
Toy vehicle images 3 and 4; dist = 1 m	3.847004	2.757804	0.031442
Toy vehicle images 5 and 6; dist = 2 m	3.960434	2.825093	0.039153
Toy vehicle images 7 and 8; dist = 3 m	3.378901	2.512207	0.031566
Toy vehicle images 9 and 10; dist = 4 m	3.535563	2.718341	0.037940
Toy vehicle images 11 and 12; dist= 5 m	3.239341	2.806035	0.014134
Toy vehicle images 13 and 14; dist= 6 m	3.400691	2.118004	0.016603

dist = distance and m = meter

generates stable features at different distances with far less computation time. These features are marked as stable and are assigned distances based on the colors. Blue features show the object is near, and green shows the object is far away. See Figs. 3(b) and (d).

5. Conclusion

This work presents a novel method for distance estimation of toy vehicles with color processing of depth-image pixels. The proposed method generates different distance color images, with features optimized by neuro-image processing. This method is helpful for the development of various autonomous path-finding applications in robot vision fields.

References

[1] W. Hu et al., "A survey on visual surveillance of object motion and behaviors," IEEE Transactions on Systems, Man, and Cybernetics, vol. 34, no. 3, pp. 334-352 (2004). Article (CrossRef Link)

[2] T. L. Liu and H. T. Chen, "Real-time tracking using trust region methods," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 3, pp. 397-402 (2004). Article (CrossRef Link)

[3] J. Fan, X. Shen, and Y. Wu, "Scribble Tracker: A Matting-Based Approach for Robust Tracking," IEEE Transactions on Pattern Analysis and Machine

Intelligence, vol. 34, no. 8, pp. 1633-1644 (2012). Article (CrossRef Link)
[4] X. Li et al., "Graph mode-based contextual kernels for robust SVM tracking," IEEE International Conference on Computer Vision (ICCV), pp. 1156-1163 (2011). Article (CrossRef Link)
[5] S. Hare et al., "Efficient online structured output learning for keypoint-based object tracking," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1894-1901 (2012). Article (CrossRef Link)
[6] M. Grabner, H. Grabner, and H. Bischof, "Learning Features for Tracking," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8 (2007). Article (CrossRef Link)
[7] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110 (2004). Article (CrossRef Link)
[8] H. Bay et al., "SURF: Speeded Up Robust Features," Computer Vision and Image Understanding, vol. 110, no. 3, pp. 346-359 (2008). Article (CrossRef Link)

Kajal Sharma received a BE in computer engineering from the University of Rajasthan, India, in 2005, and MTech and Ph.D. degrees in computer science from Banasthali University, Rajasthan, India, in 2007 and 2010, respectively. From October 2010 to September 2011, she worked as a post-doctoral researcher at Kongju National University, Korea. From October 2011 to April 2013, she worked as a post-doctoral researcher at the School of Computer Engineering, Chosun University, Gwangju, Korea. Presently she is working as an independent researcher in Korea. Her research interests include image and video processing, neural networks, computer vision, and robotics. She has published many research papers for various national and international journals and conferences.