ISSN (Online) : 2319 – 8753
ISSN (Print)   : 2347 - 6710

IJIRSET

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization,   Volume 3, Special Issue 1, February 2014

International Conference on Engineering Technology and Science-(ICETS'14)

On 10th & 11th February Organized by

Department of CIVIL, CSE, ECE, EEE, MECHNICAL Engg. and S&H of Muthayammal College of Engineering, Rasipuram, Tamilnadu, India

Virtual Mirror Rendering and Recognition of 3D-Objects using an RGB-D Camera

Gowthami.D[1], Dr.S.M.Ramesh[2]

PG Scholar, Department of ECE, BIT-Sathyamangalam, Tamilnadu, India [1]

Associate Professor, Department of ECE, BIT-Sathyamangalam, Tamilnadu, India [2]

ABSTRACT: Mirror is probably the most shared optical device in our daily life. The virtual mirror rendering by means of a combined camera-display system improved with virtual scenes and objects, has a wide range of applications from cosmetics to medical interventions.Challenges like wide angle viewing of the environment, accurate viewpoint tracking and rendering, as well as real time demonstration to deliver immediate visual feedback are involved in the realistic simulation of a mirror. In this paper, we suggest virtual image rendering system using a commodity structured light RGB-D cameras.These cameras offer the depth information for viewpoint tracking and rendering of scene from altered prospective. A novel depth denoising and completion algorithm is proposed to precise the problem of missing and erroneous depth measurements by structured light cameras. A scalable client server architecture is hosted to render the dynamic scene captured by the static camera network in online with the 3-D background prototype generated by offline.In addition to this labeling of objects are done for the images which must be in a recognized manner. Experimental results demonstrate effective performance of our method in which it provides reliable depth estimation, image rendering and statistics identification.

Index Terms—Virtual Mirrors, depth image denoising, image reconstruction, client-server systems, Markov random field, RGB-D system,3-D scene scanning, Dataset, RANSAC plane fitting, Label me tool, SHIFT descriptors, EMK,PCA.

I.INTRODUCTION

Real-time 3D reconstruction and rendering of humans captured by a set of cameras is important in a number of applications, such as virtual try-on and interactive mixed reality systems. These applicationscreate a virtual mirror by displaying a live representation of the user embedded in an artificial environment. Image-based methods are frequently used to perform reconstruction and rendering but suffer from poor performance at higher resolutions. Previous implementations computed every output image from scratch and therefore did not reach their highest potential efficiency.In this paper; we propose a camera-display system having a network of RGB-depth (RGB-D) sensors that senses the 3-D world. The depth information given by the camera is used to track the viewpoint, and then renders the dynamic scene by tracing the light ray from each scene point to the point of reflection on the virtual mirror, and finally determines the proper position and color values on the display surface at which the virtual reflected ray passes through before reaching the viewpoint.

Compared with other RGB-D rendering systems, our proposed system has the following technical contributions:1) A novel graphical model based on depth denoising and completion algorithm to modify the missing and erroneous depth measurements caused by the low-cost Microsoft Kinect sensors. 2) A realistic mirror visual effect by incorporating three key features: viewpoint dependent content, wide field of view, and 3-D based rendering.3)To provide real-time performance, we avoid the computational intensive surface meshing process and base our design on a 3-D point cloud which is faster in terms of both view acquisition and rendering.

The proposed algorithm performs reconstruction and rendering simultaneously, which is more efficient than executing both steps separately. When users move in a virtual mirror scenario, their motions are typically smooth and slow compared to the camera frame rate..With the suggested efficient image-based rendering algorithms, various new applications become feasible, such as high fidelity tele presence systems or virtual mirrors. For advanced applications, like virtual try-on systems and mixed reality augmentations, the rendered image of a user must be augmented with virtual garments and accessories.

In this paper we present techniques for RGB-D based object recognition and detection and demonstrate that combining color and depth information can substantially improve the results achieved on our dataset. The techniques are evaluated at two levels. Category level recognition and detection involves classifying previously unseen objects as belonging in the same category as objects that have previously been seen (e.g., coffee mug). Instance level recognition and detection is identifying whether an object is physically the same object that has previously been seen. The ability to recognize and detect objects at both levels is important if we want to use such recognition systems in the context of tasks such as service robotics. For example, identifying an object as a generic coffee mug or as Amelias coffee mug can have different implications depending on the context of the task.

II.RELATED WORK

2.1 VIRTUAL MIRROR RENDERING SYSTEM

Virtual mirror systems are not new concept in computer vision. Several research groups have developed virtual mirror prototypes. Though they differ in some aspects, most of them implement simple appearance modification with a limited viewpoint. Darrell et al. describe a virtual mirror interface that applies different graphical effects on the face of the viewer in front [2]. In [3], Kitanovski and Izquierdo propose a virtual facial modification program using a user defined 3-D face model with Kalman-filter based tracking. However, their works do not consider the viewpoint's influence on rendering a large virtual mirror. Francois and Kang propose a hand-held mirror simulation device in [4]. Although they consider the impact of viewpoint change, their system uses a simplified model for rendering by assuming that the 3-D world is parallel to the mirror. Such assumption significantly limits the range of possible perspective change and is not suitable for rendering large mirror. In [5], Hilsmannet al. proposes a dress-fitting system where users can virtually themselves in different clothes and see how they look from the virtual mirror. Since the target objects (clothing) in these systems are already known, they usually have a precomputed model from either an existing 3-D model or a collection of large training data.. A recent paper by Strakaet al. describes a system that allows the user to view him- or herself anywhere within a 360◦ field of view. Multiple RGB cameras are mounted around the user to capture the data for 3-D reconstruction. However, only the person is rendered in their proposed system without any background, thereby neglecting an essential characteristic of a mirror . In [6], Garroet al. presents an interpolation scheme for depth super-resolution. A high resolution RGB camera is used to guide the up sampling process on the depth image. To interpolate the missing depth pixel, the scheme uses neighboring depth pixels mapped into the same color segment as the target pixel. This method relies strongly on the extrinsic alignment between the color and depth image. In [7], Wang et al. propose a stereoscopic in-painting algorithm to jointly complete missing texture and depth by using two pairs of RGB and depth cameras. Regions occluded by foreground are completed by minimizing an energy function.

2.2 OBJECT RECOGNITION SYSTEM

The availability of public image repositories on the Web, such as Google Images and Flickr, as well as visual recognition benchmarks like Caltech 101, Label Me [12] and Image Net has enabled rapid progress in visual object category and instance detection in the past decade. Unlike many existing recognition benchmarks that are constructed using Internet photos, where it is impossible to keep track of whether objects in different images are physically the same object, our dataset consists of multiple views of a set of objects. This is similar to the 3D Object Category Dataset presented by Savarese et al. [8], which contains 8 object categories, 10 objects in each category, and 24 distinct views of each object. The RGB-D Object Dataset presented here is at a much larger scale, with RGB and depth video sequences of 300 common everyday objects from multiple view angles totaling 250,000 RGB-D images. The objects are organized into a hierarchical category structure using Word Net hyponym/ hypernym relations. We performed extensive comparative study between our scheme and theirs.The results are given in later sections.

III. VIRTUAL MIRROR MODELING AND SIMULATION

3.1 SYSTEM DESIGN

Methodology: Read RGB Image or Read converted images from any videos. Extract Foreground image from Back ground using contour and edge detection methods. On the extracted Foreground image apply depth analysis and depth correction. After that Enhance foreground image and fuse with appropriate background image of same identity.

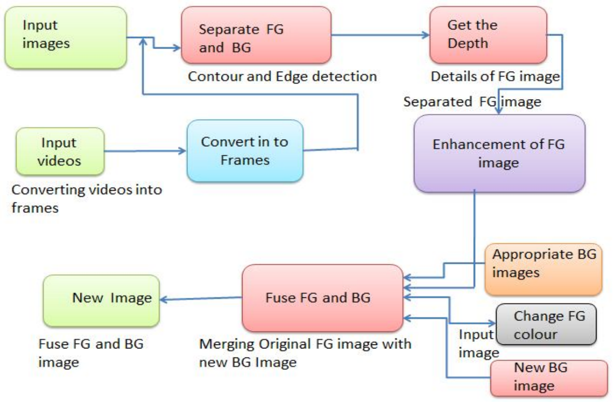


Fig 1: Rendering of virtual mirror image.

3.2 RGB IDENTIFICATION

Basically the project will identify the RGB values of the image for feature extraction purpose. There are some properties available for RGB Images. HSL , HSV and HSI are the most common cylindrical coordinate representations of points in an RGB colour model. These representations rearrange the geometry of RGB in an attempt to be more intuitive and perceptually relevant than the Cartesian (cube) representation.

3.3 CONTOUR AND EDGE DETECTION

The contour and edge detection will handshake with the segmentation method for foreground and background separation. An adaptive and compact background model that can capture structural background motion over a long period of time under limited memory. This allows us to encode moving backgrounds or multiple changing backgrounds. The capability of coping with local and global illumination changes. unconstrained training that allows moving foreground objects in the scene during the initial training period; layered modelling and detection allowing us to have multiple layers of background representing different background layers.

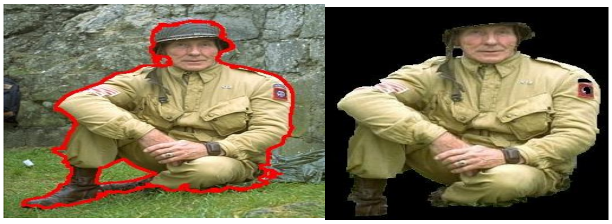


Fig 2: Separation of FG and BG in an image

3.4 DEPTH ANALYSIS

Our proposed algorithm for denoising and completing depth pixels is based on foreground/background separation. By separating pixels into foreground and background, the generated image can be completed in a guided manner. Missing depth pixels are interpolated or in-painted by neighboring depth pixels from the same group to prevent smearing of object boundaries. In addition, the foreground-background separation admits a low-complexity and better-quality rendering of the mirror image – foreground pixels can be first extracted and projected onto the display image while the remaining region are filled by pre-captured backgrounds, thereby avoiding the rendering of the entire frame and filling background holes caused by the change of viewpoint.

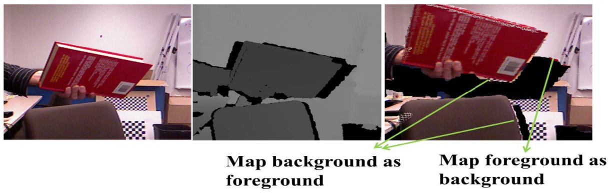


Fig.3: Error in RGB-D systems: (a) RGB Image (b)Depth Image (c) Virtual View.

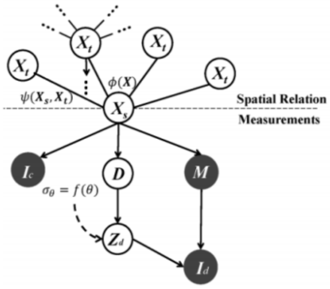


Fig 4: Foreground-Background Depth Measurement Model.

3.5 FUSE IMAGES

Now we having a new foreground image and background image, while fusing we will get a new fused image with new background. The fusion methods such as averaging, Brovey method, principal component analysis (PCA) and IHS based methods fall under spatial domain approaches. The disadvantage of spatial domain approaches is that they produce spatial distortion in the fused image. Spectral distortion becomes a negative factor while we go for further processing, such as classification problem. Spatial distortion can be very well handled by frequency domain approaches on image fusion.

Fig 5: Different virtual mirror views by using different 3-D background and inserting novel 3-D objects.

IV. ROBOTIC RECOGNITION OF 3D OBJECTS

Over the last decade, the availability of public image repositories and recognition benchmarks has enabled rapid progress in visual object category and instance detection. Today we are witnessing the birth of a new generation of sensing technologies capable of providing high quality synchronized videos of both color and depth,by the RGB-D (Kinectstyle) camera. With its advanced sensing capabilities and the potential for mass adoption, this technology represents an opportunity to dramatically increase robotic object recognition, manipulation, navigation, and interaction capabilities. In this paper, we introduce a large-scale, hierarchical multi-view object dataset collected using an RGB-D camera. The dataset contains 300 objects organized into 51 categories and even we describe the dataset collection procedure and introduced techniques for RGB-D based object recognition and detection, also demonstrating that combining color and depth information substantially improves quality of results.

4.1 RGB OBJECT DATASET COLLECTION

The RGB-D Object Dataset contains visual and depth images of 300 physically distinct objects taken from multiple views. Each of the 300 objects in the dataset belongs to one of the 51 leaf nodes in this hierarchy, with between three to fourteen instances in each category. The dataset is collected using a sensing apparatus consisting of a prototype RGB-D camera manufactured by Prime-Sense[11] and a firewire camera from Point Grey Research. Using this camera setup, we record video sequences of each object as it is spun around on a turntable at constant speed. Data was recorded with the cameras mounted at three different heights relative to the turntable, at approximately 30_, 45_ and 60_ above the horizon. Each video sequence is recorded at 20 Hz and contains around 250 frames, giving a total of 250,000 RGB + Depth frames in the RGB-D Object Dataset. The video sequences are all annotated with ground truth object pose angles between [0; 2π] by tracking the red markers on the turntable.

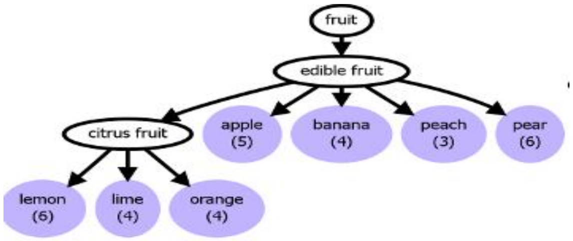


Fig 6: RGB-D Object Dataset object hierarchy

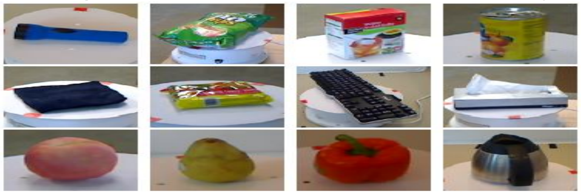


Fig 7: Objects from the RGB-D Object Dataset.

4.2 SEGMENTATION

Without any post-processing, a substantial portion of the RGB-D video frames is occupied by the background.

The first step in segmentation is to remove most of the background by taking only the points within a 3D bounding box.we perform RANSAC plane fitting [9] and vision-based background subtraction given by Kaew-TraKulPong et al. [10] to produce segmentation. These methods are very good at segmenting out the edges of objects. Since depth-based and vision-based segmentation each excel at segmenting objects under different conditions, we combine the two to generate our final object segmentation. Finally a filter is run on this segmentation mask to remove isolated pixels.

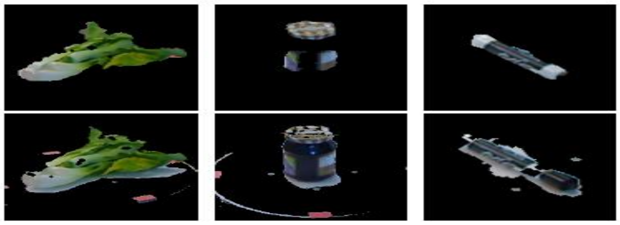


Fig 8: Combined depth and visual segmentation.

4.3 VIDEO SCENE ANNOTATION

Dataset also includes 8 video sequences of natural scenes. The scenes cover common indoor environments, including office workspaces, meeting rooms, and kitchen areas. The video sequences were recorded by holding the RGB-D camera at approximately human eye-level while walking around in each scene. Each video sequence contains several objects from the RGB-D Object Dataset. We demonstrate that the RGB-D Object Dataset can be used as training data for performing object detection in these natural scenes. Traditionally, the computer vision community has annotated video sequences one frame at a time. A human must tediously segment out objects in each image using annotation software like the LabelMe annotation tool [12]. We propose an alternative approach. Instead of labeling each video frame, we first stitch together the video sequence to create a 3D reconstruction of the entire scene, while keeping track of the camera pose of each video frame. We label the objects in this 3D reconstruction by hand. Finally, the labeled 3D points are projected back into the known camera poses in each video frame and this segmentation can be used to compute an object bounding box.

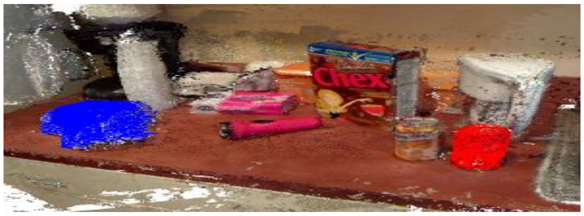


Fig. 9 :3D reconstruction of a kitchen scene.

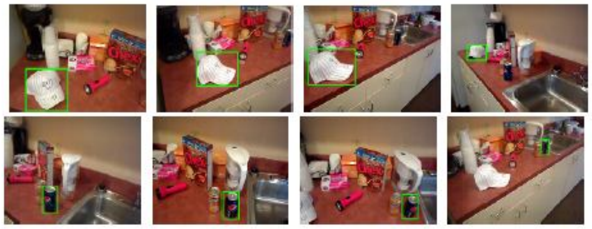


Fig.10: Ground truth bounding boxes of cap and soda can got by labeling the scene reconstruction

4.4 OBJECT RECOGNITION USING THE RGB-D OBJECT DATASET

In object recognition the task is to assign a label to each query image. A set of images are annotated with their ground truth labels and given to a classifier, which learns a model for distinguishing between the different classes. We evaluate object recognition performance at two levels: category recognition and instance recognition.1)In category level recognition, the system is trained on a set of objects. At test time, the system is presented with an RGB and depth image pair containing an object that was not present in training and the task is to assign a category label to the image (e.g. coffee mug or soda can).2) In instance level recognition, the system is trained on a subset of views of each object. The task here is to distinguish between object instances (e.g. Pepsi can, Mountain Dew can, or Aquafina water bottle). For category recognition, we randomly leave one object out from each category for testing and train the classifiers on all views of the remaining objects. For instance recognition, we consider two scenarios:

- Alternating contiguous frames: Divide each video into 3 contiguous sequences of equal length. There are 3 heights (videos) for each object, so this gives 9 video sequences for each instance. We randomly select 7 of these for training and test on the remaining 2.

ISSN (Online) : 2319 – 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization, Volume 3, Special Issue 1, February 2014

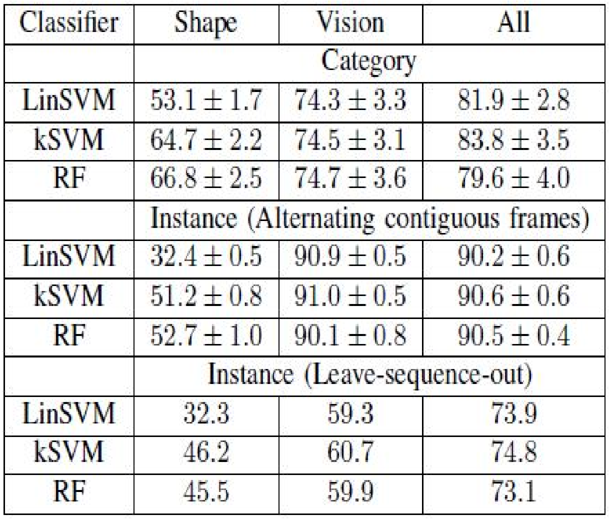
International Conference on Engineering Technology and Science-(ICETS'14)

On 10th & 11th February Organized by

Department of CIVIL, CSE, ECE, EEE, MECHNICAL Engg. and S&H of Muthayammal College of Engineering, Rasipuram, Tamilnadu, India

- Leave-sequence-out: Train on the video sequences of each object where the camera is mounted 30º and 60º above the horizon and evaluates on the 45º video sequence.

We present object recognition results on the RGB-D Object dataset using several different classifiers with only shape features, only visual features, and with both shape and visual features.



Classifier	Shape	Vision	All
	Category		
LinSVM	53.1 ± 1.7	74.3 ± 3.3	81.9 ± 2.8
kSVM	64.7 ± 2.2	74.5 ± 3.1	83.8 ± 3.5
RF	66.8 ± 2.5	74.7 ± 3.6	79.6 ± 4.0
	Instance (Alternating contiguous frames)		
LinSVM	32.4 ± 0.5	90.9 ± 0.5	90.2 ± 0.6
kSVM	51.2 ± 0.8	91.0 ± 0.5	90.6 ± 0.6
RF	52.7 ± 1.0	90.1 ± 0.8	90.5 ± 0.4
	Instance (Leave-sequence-out)		
LinSVM	32.3	59.3	73.9
kSVM	46.2	60.7	74.8
RF	45.5	59.9	73.1

Fig .11: Category and instance recognition performance of various classifiers

For object detection, there are many potential negative examples. The trained classifier is used to search images and select the false positives with the highest scores (hard negatives). These hard negatives are then added to the negative set and the classifier is retrained. we show precision-recall curves comparing detection performance with a classifier trained using image features only (red), depth features only (green), and both(blue).Our current single-threaded implementation takes approximately 10 seconds to run the four object detectors to label each scene. Both feature extraction over a regular grid and evaluating a sliding window detector are easily parallelizable. We are confident that a GPU-based implementation of the the described approach can perform multi-object detection in real-time.

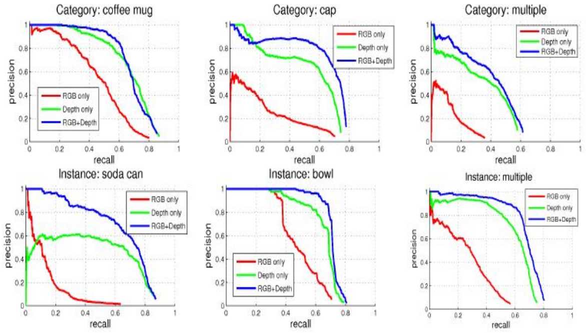


Fig.12: Precision-recall curves comparing performance with image features only (red), depth features only (green), and both (blue).

V. EXPERIMENTAL RESULTS

5.1 PROCESS TO RENDER VIRTUAL IMAGES

A number of experiments are conducted to demonstrate the effectiveness and accuracy of our proposed system. First step is to get the original input frames from three Kinects such as left view, right view, and bottom view. Then extract the synthesized foreground view from the subject's viewpoint and demonstrate different virtual mirror views by using different 3-D background. In addition to that novel 3-D objects are inserted to support virtual try on applications..

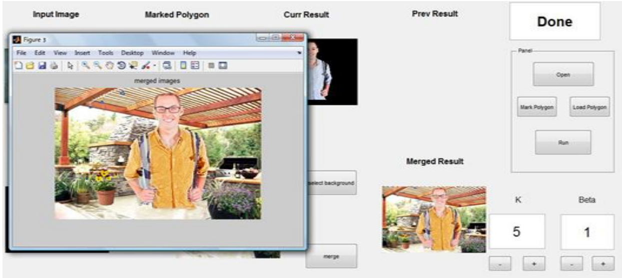


Fig 13: Rendered virtual mirror image view

5.2 PROCESS TO RECOGNIZE 3D OBJECTS

A large-scale, hierarchical multiview object data set collected using an RGB-D camera. The first step in segmentation is to remove most of the background by taking only the points within a 3D bounding box. In addition to the views of objects recorded using the turntable, the RGB-D Object Dataset also includes the video sequences of natural scenes. we annotated these natural scenes with the ground truth bounding boxes of objects in the RGB-D Object Dataset. We evaluate object recognition performance at two levels termed as category recognition and instance recognition using multi category and instance detectors.

5.2.1 MULTI CATEGORY LEVEL DETECTION

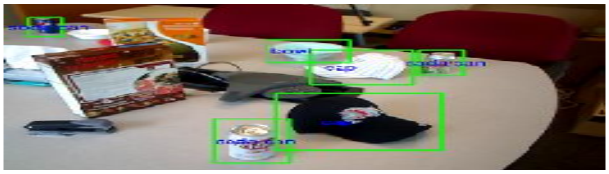


Fig 14:Category level object Recognization

5.2.2 MULTI INSTANCE LEVEL DETECTION

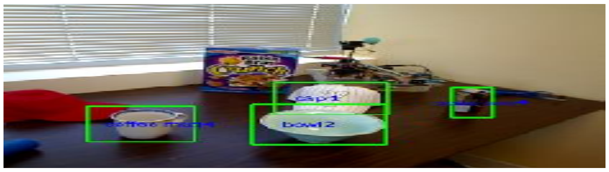


Fig 15:Instance level object Recognization

VI. CONCLUSION

In this paper, we have presented a framework for rendering virtual mirror and recognizing 3D objects by using modern PCs and commodity RGB-D cameras such as the Microsoft Kinect. In virtual mirror rendering, The initial depth data has been enhanced using a depth denoising and completion algorithm which takes advantage of creating a complete novel probabilistic background/foreground separation to eliminate outliers and complete missing values. To support a large mirror surface and wide viewing angle, an off-line background scanning is used to capture the background environment.Dynamic RGB-D data captured by each client is used to estimate the viewpoint and create a 3-D point cloud to render a viewer-dependent mirror image. The server aggregates all the partially rendered mirror images to compute the final result. We are currently exploring GPU implementation and algorithmic speedup on the multiple-round belief propagation in depth denoising. For our future work, the visual perception could be enhanced by rendering stereo views. In recognizing of 3D objects, we have presented a large-scale, hierarchical multi-view object dataset collected using an RGB-D camera. We have shown that depth information is very helpful for background subtraction, video ground truth annotation via 3D reconstruction, object recognition and object detection. This can be implemented in future days by integrating RGB-D Object Dataset and a set of tools in to the Robot Operating System for accessing and processing the dataset in real environment.

REFERENCES

[1] J. Shen, S-C S. Cheung, and J. Zhao, “Virtual mirror by fusing multiple RGB-D cameras,” in Proc. APSIPA Ann. Summit Conf., Dec. 2012,pp. 1–9.

[2] M. Biehl, A. Ghosh, and B. Hammer, “Learning vector quantization: The dynamics of winner-takes-all algorithms,” Neurocomputing, vol. 69, nos. 7–9, pp. 660–670, Mar. 2006.

[3] V. Kitanovski and E. Izquierdo, “3-D tracking of facial features for augmented reality applications,” in Proc. Int. Workshop Image Anal.Multimedia Interact. Services, vol. 2. Apr. 2011.

[4] A. R. J. Francois and E.-Y. E. Kang, “A handheld mirror simulation,” in Proc. Int. Conf. Multimedia Expo, vol. 2. Jul. 2003, pp. 745–748.

[5] A. Hilsmann and P. Eisert, “Realistic cloth augmentation in single view video,” in Proc. Vis., Modell., Visualizat. Workshop, 2009, pp. 55–62.

[6] V. Garro, C. dal Mutto, P. Zanuttigh, and G. M. Cortelazzo, “A novel interpolation scheme for range data with side information,” in Proc.IEEE Vis. Media Prod. Conf., Nov. 2009, pp. 52–60.

[7] L. Wang, H. Jin, R. Yang, and M. Gong, “Stereoscopic inpainting: Joint color and depth completion from stereo images,” in Proc. IEEE CVPR, Jun. 2008, pp. 1–8.

[8] P. Felzenszwalb, D. McAllester, and D. Ramanan.A discriminatively trained, multiscale, deformable part model.In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008.

[9] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM, 24(6):381–395, 1981.

[10] P. Kaewtrakulpong and R. Bowden.An improved adaptive background mixture model for realtime tracking with shadow detection. In European Workshop on Advanced Video Based Surveillance Systems,2001.

[11] PrimeSense. http://www.primesense.com/.

[12] B. Russell, K. Torralba, A. Murphy, and W. Freeman. Labelme: a database and web-based tool for image annotation. International Journal of Computer Vision, 77(1-3), 2008.