The University Of Sheffield.

RGB-D DATA FUSION IN COMPLEX SPACE

Ziyun Cai and Ling Shao

University of Sheffield

IEEE International Conference on Image Processing

September 2017

The University Of Sheffield.

Background of RGB-D data:

The acquisition of RGB-D data:

Past:

Early range sensors (such as Konica Minolta Vivid 910, Faro Lidar scanner, Leica C10 and Optech ILRIS-LR)

Disadvantages:

They are expensive and difficult to use for researchers in a human environment. No much follow-up research at that time.

Recently:

With the release of the low-cost 3D Microsoft Kinect sensor1 on 4th November 2010, acquisition of RGB-D data becomes cheaper and easier.

Then:

The investigation of computer vision algorithms based on RGB-D data has attracted a lot of attention in the last few years.

The University Of Sheffield.

Examples of RGB-D data:

RGB data and depth data:

Pixel value: the distance from the camera to the object

The University Of Sheffield.

RGB-D DATA FUSION:

Many RGB-D fusion methods have been proposed to extract RGB-D features.

K. Lai, L. Bo, X. Ren, and D. Fox. A large-scale hierarchical multi-view rgb-d object dataset. In IEEE International Conference on Robotics and Automation (ICRA), pages 1817–1824, 2011.

R. Socher, B. Huval, B. Bath, C. D. Manning, and A. Y. Ng. Convolutional-recursive deep learning for 3d object classification. In Neural Information Processing Systems, pages 665–673, 2012.

S. Song and J. Xiao. Deep sliding shapes for amodal 3d object detection in rgb-d images. 2016.

…

The University Of Sheffield.

Limitations:

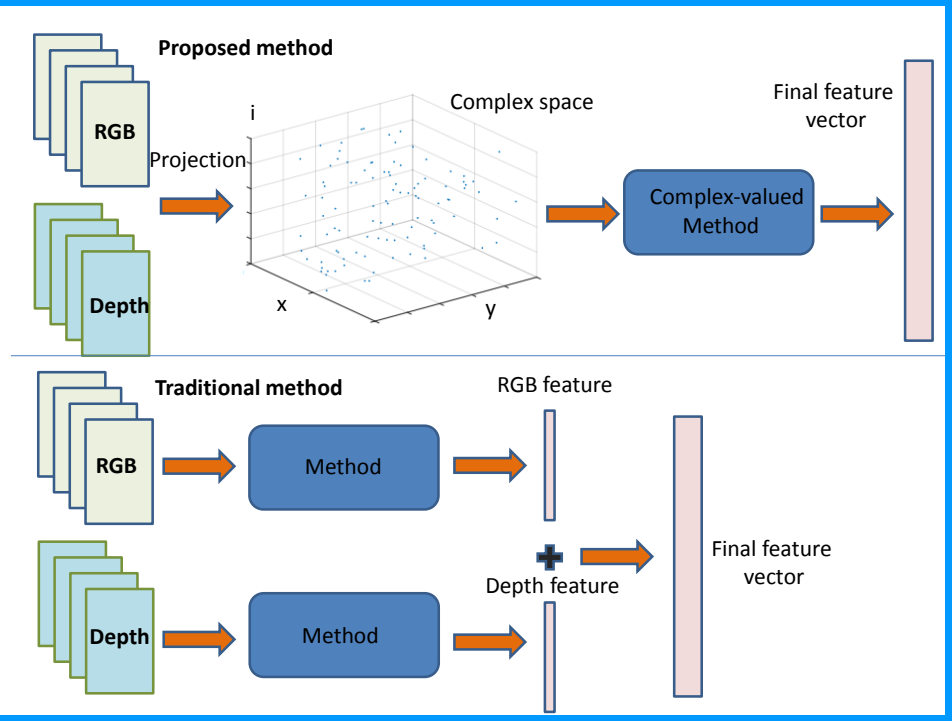
Some methods just learn features from RGB and depth separately and then simply concatenate them together as RGB-D features or encode these two kinds of features, which cannot explore the correlation between the RGB pixels and their corresponding depth pixels.

The correlation and complementary property between RGB and depth are ignored.

The University Of Sheffield.

Our method:

To better explore the correlation between the RGB pixel and the corresponding depth pixel, and take advantage of the complementary property, we first project raw RGB-D data into a complex space and then jointly learn features from the fused RGB-D images. The correlated and individual parts of the RGBD information in the new feature space are well combined.



The flow chart shows the difference between our fusion method and some traditional fusion methods.

Motivation:

Our fusion method can also be considered as representing the data closer to the nature of the data.

1) In physics, the range data correspond to the phase change and color information corresponds to the intensity.

2) From computer vision view, the feature representations are expected to satisfy low mutual information and also show a lot of variations. The fused RGB-D data should be treated holistically.

The University Of Sheffield.

C-SIFT：

Beyond this, we also modify the classical SIFT into complex valued SIFT (**C**-SIFT) to evaluate our fusion method. It is worthy to note that **C**-SIFT is just an example to show the advantages of the fusion method. CNNs, DBNs or other methods can be introduced into complex space as well.

The University Of Sheffield.

Fusion Methodology：

RGB image can be considered as amplitude measurements, which depends on the nature illumination, e.g. , sun light. According to depth images, the pixel values of the depth images always mean the distance from the camera to the observed objects. The depth image is often considered as the phase change measurement, which depends on the measured scattering received from the active illumination with the sensor, e.g. , laser. The phase can be regarded as actual distance.

We define **IR(x,y)** as the RGB image,
ID(x, y, d) as the depth image,

where d = d(x, y) , x and y are the image coordinate points, d is the depth value on the coordinate (x, y) .

The University Of Sheffield.

The fused complex-valued image function is expressed as:

$$f(x,y,d) = I_R(x,y) + I_D(x,y,d)e^{i\phi(x,y,d)},$$

where phase $\phi = \phi(x,y,d) \in [0, 2\pi)$

Moreover, with the representation of complex number, the fused complex valued image can be represented as Polar representation and Cartesian representation:

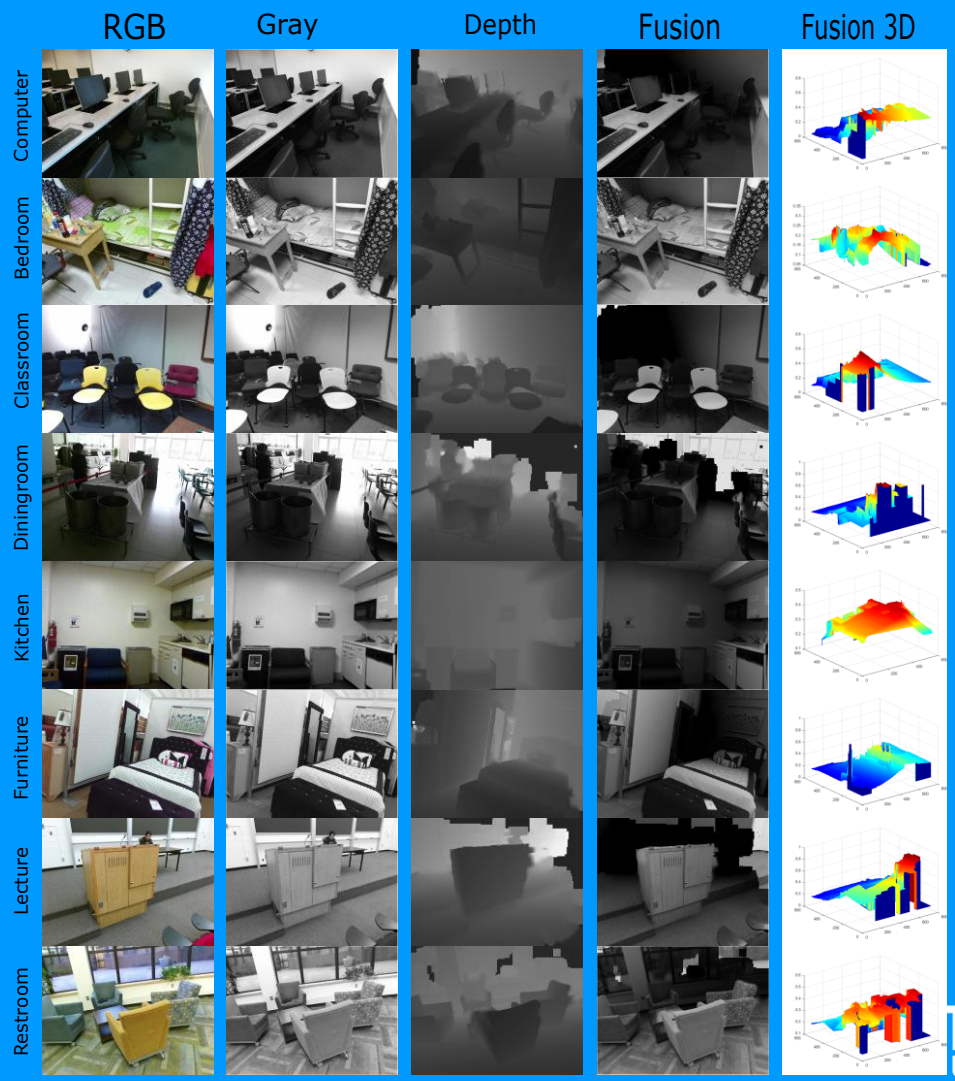
$$I_f^P = |f|e^{iarg(f)}, \quad I_f^C = Re(f) + iIm(f),$$

The University Of Sheffield.

Since we normalize all complex-valued images, we can obtain max $|f| = 1$.

Some random example images from 8 different scenes which include computer room, bedroom, classroom, dining room, kitchen, furniture room, lecture room and restroom from top to bottom.

Note that all the RGB images mentioned in our methodology are first converted into gray images.



The University Of Sheffield.

Comparison：

The comparison among the representations (RGB images, depth Images, fused RGB-D Polar representations and fused RGB-D Cartesian representations) from three aspects:

1）Mutual Information and Independence
2）Feature Distribution
3）Euclidean KS-distance to Uniformity

The fused RGB-D Cartesian representation is the optimal image representation among the four image representations.

The University Of Sheffield.

Complex-valued SIFT：

In this section, we modify classical SIFT into complex-valued SIFT (**C**–SIFT).

In the local extrema detection step, different from SIFT which detects the local extrema through comparing its **26** real-valued neighbors, Complex-valued SIFT chooses to compare the module **m** among these neighbors.

The module can be calculated as：

$$m^2 = Re(f)^2 + Im(f)^2$$

It can make sure that the color information and the depth information are all considered when choosing the keypoints.

The University Of Sheffield.

Experimental Setup

Datasets：
NYU Depth V1 dataset and SUN RGB-D dataset

We hope to show the advantages of our fused method under the same conditions without the influence from the methodological difference.

The University Of Sheffield.

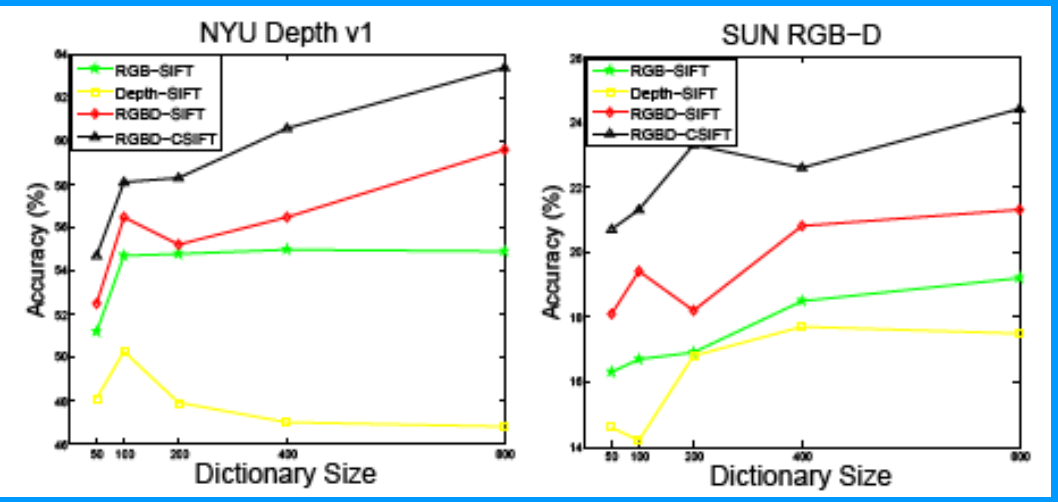
Experimental results:

Accuracies (%) for scene classification on NYU Depth V1 and SUN RGB-D datasets.

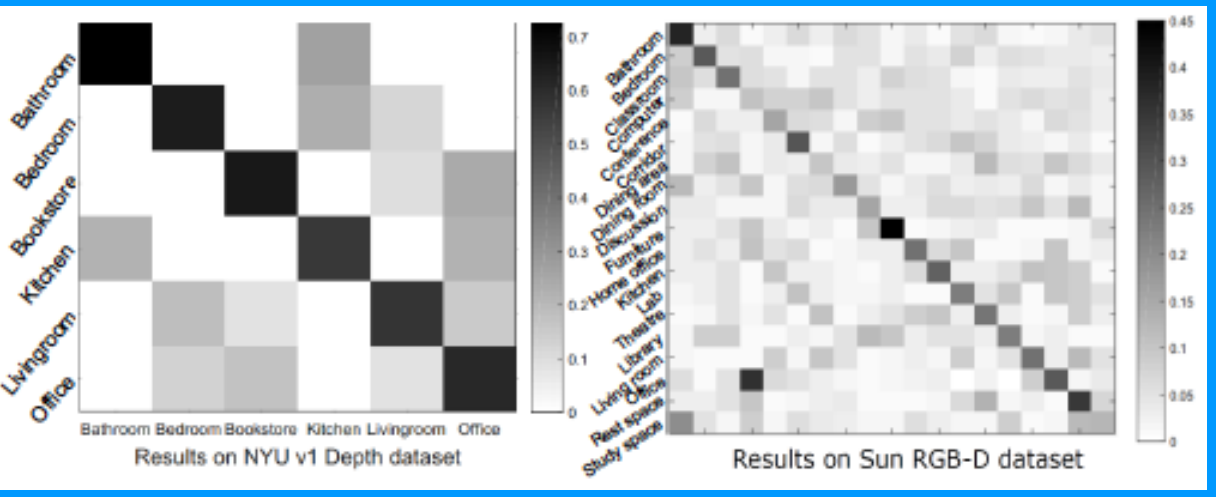
Methods	NYY Depth V1		SUN RGB-D	
	SIFT	AlexNet	SIFT	AlexNet
RGB	55.0	60.1	19.2	22.3
Depth	50.3	54.2	17.7	20.6
RGB-D	59.6	65.7	21.3	28.7
Our method	**63.4**	**70.1**	**24.4**	**31.2**

For the deep features, we use the NYU depth V2 RGB-D dataset with more than 200 K frames from the 249 training video scenes for learning the fused images initial AlexNet. Then we fine-tune the fused images from NYU Depth V1 and SUN RGB-D on this initial model, to extract the features of the fc-7 layer.

The University Of Sheffield.



Scene classification performance on NYU Depth v1and SUN RGB-D datasets with different dictionary sizes.



Confusion matrices about our fusion method results on NYU Depth V1 and SUN RGB-D datasets.

The University Of Sheffield.

Conclusion

A new RGB-D fusion method for fusing RGB-D images is proposed, which can better reveal the correlation between the RGB pixels and the depth pixels, taking advantage of the complementary property. The experimental results show that our method achieves competing performance against the classical fusion methods.

The University Of Sheffield.

Future Work：

We hope CNNs, DBNs or other methods can be introduced into complex space as well.

One recent work：

Trabelsi, C., Bilaniuk, O., Serdyuk, D., Subramanian, S., Santos, J. F., Mehri, S., Rostamzadeh, N., Bengio, Y., and Pal, C. J. (2017). Deep complex networks.

The University Of Sheffield.

Take CNNs for example：

A complex valued CNN model can be built with complex input and weights.

1. The optimization method for this network should be handled.

2. Back propagation algorithm can also be modified.

3. The loss function is real-valued with complex weights in the complex case, which cannot be differentiable everywhere. (e.g. Wirtinger derivatives)

4. The labels are real-valued. (e.g. a projection layer can be added as a special case of an activation function layer)

The University Of Sheffield.

Thank you !

