Multi-Resolution for Disparity Estimation with Convolutional Neural Networks

Samer JAMMAL*, Tammam TILLO† and Jimin XIAO*
* Liverpool University, UK
E-mail: Samer.Jammal@liverpool.ac.uk
† Libera Universit di Bolzano-Bozen, Itally
E-mail: Tammam.Tillo@unibz.it
‡ X'ian Jiaotong Liverpool University, China
E-mail: Jimin.Xiao@xjtlu.edu.cn

Abstract—**Disparity estimation is the process of obtaining the depth information from the left and right views of a particular scene. A recent work based on convolutional neural network (CNN) has achieved state-of-the-art performance for the disparity estimation task. However, this network has some limitations for measuring small and large disparities, which compromises the accuracy of the obtained results. In this paper, a multi-resolution framework with a three-phase strategy to generate high quality disparity maps is proposed, which handles both small and large displacements and retains the details of the scene. The first phase up/down-samples the images to several different resolutions to improve the matching process between CNN feature maps where scaled information is obtained for objects with various sizes and distances. The second phase uses a deep CNN to estimate the disparity maps using the resampled versions, and each version is suitable for a specific range of disparities. Finally, the best fitting disparity map is adaptively selected. To the best of our knowledge, our framework is the first to exploit multiple resolutions of the stereo pair with convolutional neural network for disparity estimation. Significant performance gain is achieved with this proposed method, the mean absolute error is reduced to** 3.40 **from** 5.66**, the DispNetC performance for the Sintel dataset.**

Index Terms—**Disparity Estimation, Convolutional Neural Networks, Multi-Resolution, Up/down-sampling, Stereo Vision, Depth Estimation.**

I. Introduction

Disparity estimation and 3D reconstruction from stereo images is an important area of research in both multimedia and computer vision communities. Several approaches have been introduced to improve the accuracy and reduce the computational cost of disparity estimation [1]. A dense disparity map is particularly significant in the fields of 3D Television (3DTV) [2], Free Viewpoint Television (FTV) [3], autonomous driving [4], robotics [5], 3D modeling, object recognition [6]. A typical disparity estimation algorithm involves four stages: matching cost computation, cost aggregation, optimization, and disparity refinement [1]. The main challenge is known as the correspondence problem: given two images of the same scene, the pixels in the first image to the pixels of the second one must be matched.

Recently, Convolutional Neural Networks have been successfully applied to many tasks in the area of computer vision, including object classification [7] and detection [8]. New models have been designed to provide per-pixel estimations like semantic segmentation [9], depth estimation from a single image [10] and disparity estimation from a pair of stereo images [11]. In [11], a deep CNN that performs the matching process between the left and right CNN feature maps is employed. However, in this network pooling is applied to input images which reduce the resolution of the features before the matching process. In addition, the maximum displacement in the correlation layer is static. Therefore, this network is unable to estimate reliable and accurate disparity maps for close objects with large displacements as well as far objects with small disparities.

Moreover, the displacements between left and right views can range from small to large values, this makes most stereo methods suffer in estimating precise and accurate disparity map that cover large disparities as well as small disparities and keep the details for the whole scene. This makes each method suitable to estimate the disparities for some datasets and failed on others. To overcome the problem, a multi-resolution scheme is proposed as a possible solution for handling both large and small disparities for real scenarios.

This work aims to generate an accurate disparity map from stereo images by providing a new approach for handling small and large displacements. The input images are down/up-sampled with different factors to generate different resolutions: each pair is suitable for disparity estimation with a specific displacement range between the left and right views. High resolution pairs allow to improve the matching process and the proper estimation of disparities for far and small objects, whereas, low resolution pairs are used to estimate the disparities for close objects. After that process, several disparity maps with different resolutions are estimated using the state-of-the-art deep CNN provided in [11]. Finally, the best disparity map is adaptively selected. The experimental results demonstrate that there is significant performance gain over the state-of-the-art approach [11].

The remainder of this paper is organized as follows. Related work is reviewed in Section II while, the proposed framework is described in Section III. We present and discuss the role of different resampling and our results in Section IV. Finally, we present our conclusions in Section V.

II. Related work

Stereo vision has been an active research field of computer vision for decades and a lot of efforts have been made to improve the performance of disparity estimation. Typical stereo methods generally fall into two broad categories: local methods [12], [13], [14] tend to be applied in real-time applications where computational efficiency is valued over accuracy. On the other hand, global approaches [15], [16] provide an accurate results but with larger runtime. Between these two categories some methods provide a trade-off between accuracy and speed such as SGM algorithm [17].

Convolutional neural networks have been applied to the task of depth map prediction from a single image. Eigen et al. [18] use a multi-scale CNN to estimate the depth from a single image, where a coarse scale network is used to perform depth prediction with low resolution, and a fine scale network is trained to perform local depth refinements with higher resolution. Laina et al. [10] introduce a deep fully convolutional neural network with residual learning to model the mapping problem between monocular images and depth maps. However, estimating the depth from a single monocular image is still difficult due to the limited geometry information included in a single image.

Convolutional neural networks have been also investigated as a possible solution to the stereo correspondence problem. Zagoruyko and mokadis [19] exploited CNN architectures to learn directly from image patches a general similarity function. In particular, a two channels siamese models were used, reporting results related to stereo matching as specific case of image matching. Zbontar et al. [20], [17] used a convolutional neural network to determine corresponding patches in left and right traditional stereo images. These reconstruction methods include semi-global block matching for the estimation of the depth of each pixel with a series of post-processing steps, which renders the model computationally intensive. To reduce the computation time [21] exploited a matching network with a product layer that computes the inner product between the two representations of a siamese architectures.

In [11], a CNN is used to predict the disparity map from stereo images. This CNN includes a contracting part that progressively decreases the spatial size of the convolutional features, providing large receptive fields for higher-level convolutional layers, which in turn, enables the network to capture more global information. However, the pooling in the contracting part reduces the resolution of the predicted map. In an attempt to help the network in the matching process, a correlation layer that performs the matching between the left and right features maps is employed. One limitation results from the contracting part, the network loses the ability to retain the details in the estimated map for the far objects with small disparities. Moreover, the maximum displacement in the correlation layer is 40 pixels, which corresponds to 160 pixels in the input images. This renders the network unsuitable to estimate large disparities of foreground objects as shown in Fig.1 where the network find difficulty to estimate large displacements in the first row and could not preserve the details for far objects with small disparities in the second row.

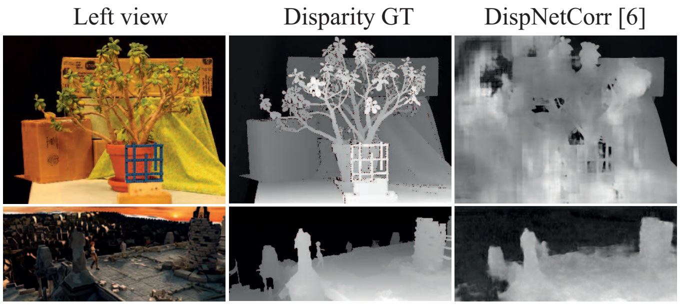

Fig. 1. Examples showing how the DispNetCorr [11] fails, in the first row, to estimate the disparity map for close objects with large disparities (over 480 pixels), and for far objects with small disparities (less than 13 pixels), in the second row.

Furthermore, the network in [11] was trained on FlyingThings3D dataset, and any modification on this network to handle larger disparities requires a new training set with larger disparities. However, the available datasets with large disparities such as Middlebury 2014 [22] are very small which might lead to heavy overfitting.

The proposed method solves this problem by using an adaptive spatial resolution method, where the input stereo images are resampled to several resolutions. Several disparity maps are generated from stereo images of different resolutions with the best is adaptivity selected as the final disparity map.

This approach allows to hand large range of disparities and integrates the convolutional neural network with handcrafted processing. Nevertheless, it is worth mentioning that the proposed approach could also be used with other stereo-based approaches for disparity estimation. The main contributions of this paper compromises of the following items: First, an architecture of the multi-resolution approach is proposed, the network takes a pair of RGB stereo images as input, and outputs an accurate disparity map that fits the objects' distances in the scene. Moreover, an evaluation of the proposed model is provided.

III. Proposed Work

A. CNN Architecture

Given a pair of rectified left and right views, a typical stereo algorithm computes the disparity using the original stereo pair without any down/up-sampling. In our work, we introduce several modifications to the network architecture compared to [11]. Firstly, we add a resampling layer to the CNN architecture to resample the input images to different resolutions. Secondly, since the disparity map estimated using DispNetC has one-quarter the resolution of the input images (half horizontally and half vertically), another resampling layer is introduced to generate a disparity map with the same size as the original stereo images. The integrated framework is shown in Fig.2.

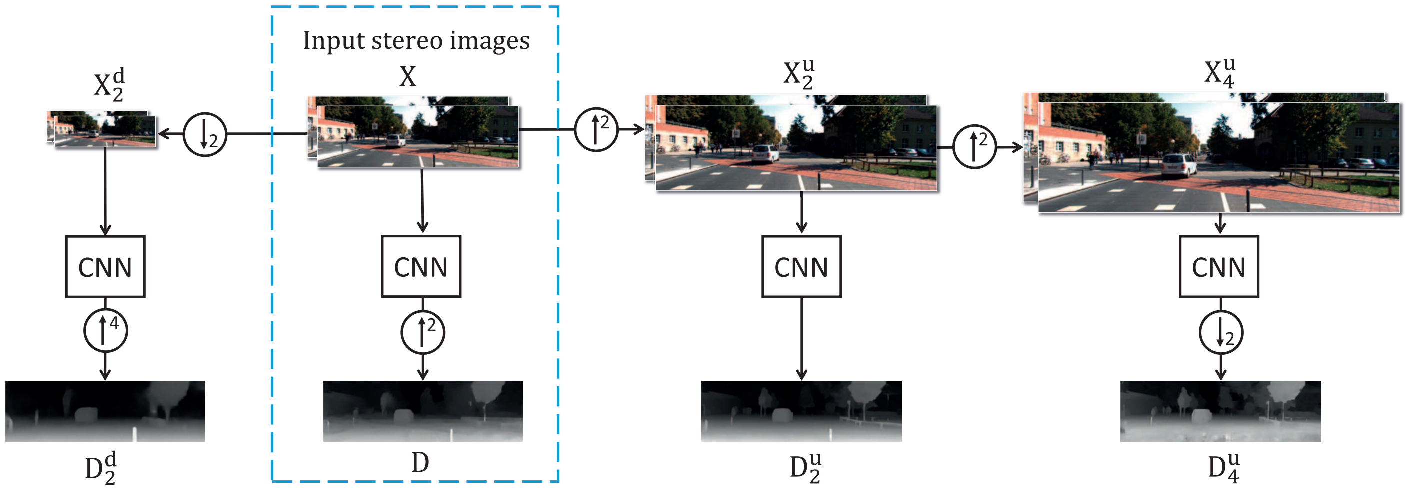

Fig. 2. The multi-resolution framework. The input stereo images are down/upsampled to several resolutions to improve the matching process between CNN feature maps. From each resmapled version a disparity map is generated, each map is suitable for a specific disparity range. Then according the displacements range in the input stereo images a the most fitting disparity map will be selected.

While disparity estimation needs precise per-pixel localization, it also requires finding correspondences between the stereo images. This involves learning to match them at different locations in the two images. Down/upsampling of the input images to several resolutions improves the matching process between CNN feature maps by using the scaled information for objects with various sizes and distances.

B. Proposed method

The proposed method is able to produce a precise disparity map for both close and far objects. We tackle the problem of estimating small disparities by up-sampling the input images with different factors as follows: Let X be the input stereo images of our method with size of $H \times W$. We upsample the inputs using several factors s_1, s_2, ... s_n. Let u_k be the up-sampling method with factor s_k, $X^u_{s_k}$ denote the up-sampled versions of X with size of $H^u_{s_k} \times W^u_{s_k}$, where $H^u_{s_k} = s_k \times H$ and $W^u_k = s_k \times W$. Let *DispNetC* be the CNN that generates the disparity map from $u_k(X)$. The estimated disparity is given by the following equation:

$$D^u_{s_k} = \mathit{DispNetC}(u_k(X)) \quad (1)$$

In DispNetC, the feature maps used in the matching process are obtained after two convolutional layers with stride 2, in other words the features are obtained after downsampling with a factor 4 which removes many details especially for the far objects. Using the upsampled images $X^u_{s_k}$ aims to use the upscaled information for the generation of features with larger resolutions keeping the fine details for the far objects before the matching process. Consequentially, a more accurate disparity map will be estimated. In a similar way, we solve the problem of estimating large disparities by downsampling the input images using several factors y_1, y_2, ... y_n. Let u_k be the upsampling method with factor y_k, $X^d_{y_k}$ denote the up-sampled versions of X with size of $H^d_{y_k} \times W^d_{y_k}$, where $H^d_{y_k} = H/y_k$ and $W^d_{y_k} = W/y_k$. The estimated disparity for the downsampled version is given by the following equation:

$$D^d_{y_k} = \mathit{DispNetC}(d_k(X)) \quad (2)$$

The multi-resolution versions require an interpolation technique that allows the generation of different different disparity maps of the same scene. Some interpolation techniques have been defined: linear interpolation, Bessel interpolation, Hermite interpolation, and so on. We have used bicubic interpolation method.

Therefore, the network makes a series of disparity predictions, starting from the lowest resolution, as an initial stage for the proposed approach. Because the DispNetC estimates a disparity map with one-quarter the resolution of the input images, we resampled the predicted disparity maps to have the same resolution of the original stereo pair $H \times W$.

Consequentially, a three-phase strategy for generating a high quality disparity map based on previous steps is used as follows:

- Firstly, the input stereo images are down-sampled by factor 2 with an initial disparity map estimated using the CNN.
- Secondly, the maximum disparity in the estimated disparity map is measured.
- Then according to the maximum disparity value of the predicted map a suitable resampled version of the input images $H_k \times W_k$ that fits the disparity range will be selected as input of the CNN. This is because each of the resampled version is suitable for a specific range of disparities. In other words, a down/upsampled version will be adaptively chosen to match the disparity range.

IV. EXPERIMENTS

In this section, we compare the performance of the proposed approach with other state-of-the-art works.

	Sintel			FlyingThings3D	KITTI2015	Middlebury2014
	disparities≤40	disparities≤80	disparities>160	disparities≤80	disparities≤80	disparities>160
Original stereo images	1.26	1.99	14.29	0.88	1.59	18.82
Down-sampled by 2	2.70	4.35	**8.04**	2.48	2.97	**9.65**
Up-sampled by 2	0.63	**1.26**	30.65	**0.49**	**1.30**	30.36
Up-sampled by 4	**0.44**	1.49	69.82	1.25	4.45	85.12

TABLE I
DISPARITY ERRORS. ALL MEASURES ARE MEAN ABSOLUTE ERRORS (IN PIXELS) OF THE SEVERAL RESAMPLED VERSIONS COMPARED TO DISPNETC [6]. THE RESULTS INCLUDE THE ERRORS FOR THE IMAGES WITH SMALL AND LARGE DISPARITIES AVAILABLE IN EACH DATASET. FOR EXAMPLE IN THE MIDDLEBURY 2014 DATASET ALL THE IMAGES CONTAIN CLOSE OBJECTS WITH DISPARITIES LARGER THAN 160 PIXELS.

A. Real and Synthetic Datasets

The evaluation was performed on real stereo benchmark datasets including the Middlebury dataset [22] which contains 81 image pairs with a disparity that reaches 600 pixels and the KITTI2015 dataset [23] with 200 image pairs. The evaluation was also performed on synthetic datasets including FlyingThings3D which contains 4760 frames for testing as well as the Sintel dataset [24] which contains 1064 image pairs. We evaluated disparity estimation accuracy using the mean absolute error (MAE) which is one of the standard error measures used for Middlebury stereo evaluation:

$$MAE = \frac{1}{T} \sum_{i \in T} |d_i - d_i^{gt}| \quad (3)$$

where T is the total number of the pixel set, d_i and d_i^{gt} are estimated and ground-truth disparity values for pixel i.

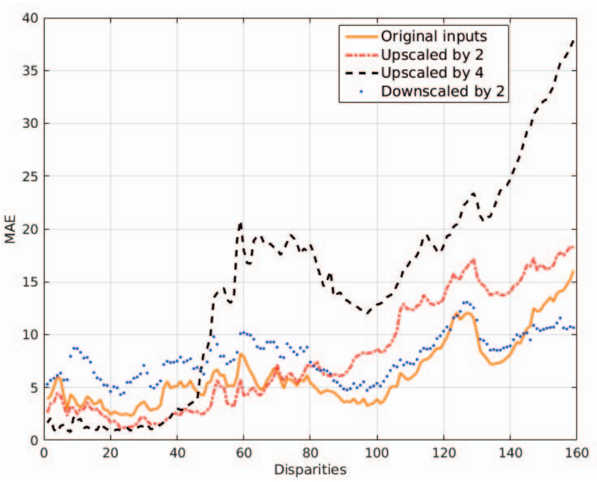

Fig. 3. MAE versus disparity analysis for the Sintel dataset.

B. Results

In the reported experiments, the input stereo images were horizontally and vertically up-sampled by factor 2 and 4 to improve the disparity for far and small objects, and the input images were down-sampled with a factor 2 for large disparities. We found that up-sampling the input images by 2 in general leads to good performance. Meanwhile, with upsampling by factor 2 the obtained disparity has the same size of the input stereo images without down/up-sampling after the disparity estimation. For down/up-sampling, the *bicubic* method is used. We ran several experiments to analyze the MAE versus disparities for the different resampling factors.

We found that each resampled version is suitable for a specific range of disparities, as shown in Fig. 3 for the Sintel. The MAE results are obtained using the whole Sintel dataset. We observe that:

- Upsampling by factor 4 is suitable if the disparities range in the input stereo images is between [0, 40].
- Upsampling by factor 2 is suitable if the disparities range is between [40, 80].
- the original input stereo images are suitable if the disparities range is between [80, 160].
- Downsampling by factor 2 is suitable for [0, *max_disp*] where *max_disp* > 160.

Table I reports the mean absolute error for different resampled versions compared with the original stereo images without resampling. The results show that whereas the upsampled versions improve the accuracy for small disparities. The downsampled versions helps to enhance the estimated disparity for foreground objects with large disparities.

We evaluated our approach with several existing disparity methods on KITTI 2012, KITTI 2015, Driving, FlyingThings3D and Sintel datasets. We evaluate the performance of two networks DispNets and DispNetsCorr1D in [11], Zbontar and LeCun [25] and the popular Semi-Global Matching [17] approach with a block matching implementation. Results are shown in Table II. The results show that the proposed approach has superior performance compared with the other methods on both synthetic and real stereo images for close and far objects. The DispNetC was trained using the FlyingThings3D dataset with original size of the input setereo images which leads to slightly better performance than our method.

For visual results, Fig. 4 depicts some disparity estimation examples using different resampling factors in comparison with DispNetC. In the first row, the stereo images pair comes from the Middlebury dataset where the maximum disparity

TABLE II
DISPARITY ERRORS. ALL MEASURES ARE MEAN ABSOLUTE ERRORS (IN PIXELS) OF THE SEVERAL RESAMPLED VERSIONS COMPARED TO SEVERAL OTHER METHODS. THE RESULTS INCLUDE THE ERRORS FOR THE IMAGES WITH SMALL AND LARGE DISPARITIES AVAILABLE IN EACH DATASET.

Method	KITTI 2012 train	KITTI 2015 train	Driving	FlyingThings3D test	Sintel clean train	Time
Proposed method	**1.50**	**1.57**	**12.3**	1.79	**3.40**	0.62s
DispNetC[11]	1.75	1.59	16.12	**1.68**	5.66	0.06s
DispNet [11]	2.38	2.19	15.62	2.02	5.38	0.06s
SGM [25]	10.06	7.21	40.19	8.70	19.62	1.1s
MC-CNN-fst [17]	-	-	19.58	4.09	11.94	0.8s

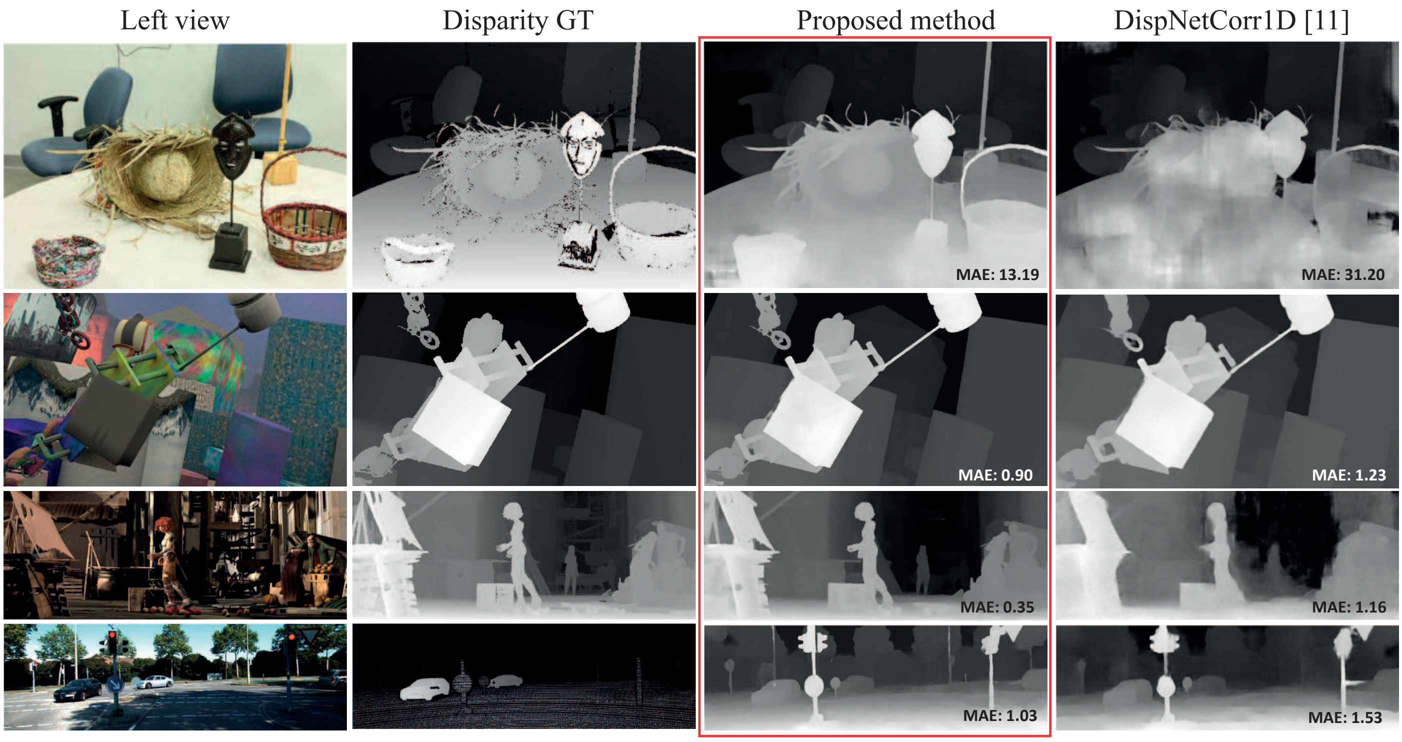

Fig. 4. Examples of disparity estimation using our method on both real and synthetic stereo datasets. Each column from left to right: left view, ground truth disparity, disparity estimated using our method and the disparity estimated using DispNetC. Rows from top to down: Middlebury 2014, FlyingThings3D (clean), Sintel(clean), KITTI 2015.

is 430 pixels. Since the maximum disparity is larger than 160, we used the down-sampled version by 2 to estimate the disparity. Our results are shown in the third column, where the network was able to measure the disparity accurately. In the second and fourth rows where the disparities are around 80 pixels, we found that using the up-sampled version by a factor 2 provides the most accurate results. In the third row, the disparities are less than 40; therefor up-sampling with a factor 4 is adopted, allowing the recognition of the far objects in detail. DispNetC fails to estimate reliable disparities for large and small disparities.

C. Various Resampling Methods

Resampling is used for several different purposes in computer vision. When resampling an image into a different size of the orginal size, an interpolating function is used to preserve the image quality. Since in our approach we down/upsample the image to different resolutions in this section, we investigate and compare several well known interpolation methods such as bicubic, nearest neighbor, bilinear, No-alteration(keeping the original pixels with no modification and an average value is used for the down/upsampling) interpolation method.

We evaluated the performance of these resampling methods on two datasets. Table II reports the mean absolute error, We can notice that: bicubic is the best for downsampling whereas the nearest method is the best for upsampling.

TABLE III
DISPARITY ERRORS. COMPARISION BETWEEN DIFFERENT RESAMPLING METHODS.

Resampling method	KITTI 2015			KITTI 2015		
	DownSampling-2	UpSampling-2	UpSampling-3	DownSampling-2	UpSampling-2	UpSampling-3
bicubic	3.00	1.82	3.78	3.02	1.54	2.33
bilinear	3.01	1.87	3.77	3.02	1.54	2.23
nearest neighbor	3.03	1.82	3.63	3.03	1.54	2.25
No-alteration	3.03	1.85	3.95	3.06	1.56	2.44
Average	3.02	1.84	3.78	3.03	1.55	2.31

V. CONCLUSION

In this paper, a three-phase method for estimating accurate disparity map for close/far objects with large/small displacements has been proposed. Firstly, the input stereo images are resampled to different resolutions, then a deep CNN is used to predict several disparity maps from each pair of the resampled versions. High resolution pairs allow the proper estimation of the disparities of far objects, whereas, low resolution pairs are employed to estimate the disparities for close objects. Experiments performed on the Middlebury2014 and KITTI2015 benchmarks demonstrate the accuracy of the proposed methodology. Future work will be devoted to generate a CNN-based merging approach.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (NO. 61501379 and NO. 61210006) and the Jiangsu Science and Technology Programme (BK20150375).

REFERENCES

[1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vision*, vol. 47, no. 1-3, pp. 7–42, Apr. 2002. [Online]. Available: http://dx.doi.org/10.1023/A:1014573219977

[2] L. Zhang and W. J. Tam, "Stereoscopic image generation based on depth images for 3d tv," *IEEE Transactions on Broadcasting*, vol. 51, no. 2, pp. 191–199, June 2005.

[3] Y. Mori, N. Fukushima, T. Fujii, and M. Tanimoto, "View generation with 3d warping using depth information for ftv," in *2008 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, May 2008, pp. 229–232.

[4] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, June 2012, pp. 3354–3361.

[5] D. Murray and J. J. Little, "Using real-time stereo vision for mobile robot navigation," *Auton. Robots*, vol. 8, no. 2, pp. 161–171, Apr. 2000. [Online]. Available: http://dx.doi.org/10.1023/A:1008987612352

[6] S. Helmer and D. Lowe, "Using stereo for object recognition," in *2010 IEEE International Conference on Robotics and Automation*, May 2010, pp. 3121–3127.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, p. 2012.

[8] C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection," in *Advances in Neural Information Processing Systems 26*, C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2013, pp. 2553–2561. [Online]. Available: http://papers.nips.cc/paper/5207-deep-neural-networks-for-object-detection.pdf

[9] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *CVPR (to appear)*, Nov. 2015.

[10] I. Laina, C. Rupprecht, V. Belagiannis, F. Tombari, and N. Navab, "Deeper depth prediction with fully convolutional residual networks," *CoRR*, vol. abs/1606.00373, 2016. [Online]. Available: http://arxiv.org/abs/1606.00373

[11] N.Mayer, E.Ilg, P.Häusser, P.Fischer, D.Cremers, A.Dosovitskiy, and T.Brox, "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation," in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, arXiv:1512.02134. [Online]. Available: http://lmb.informatik.uni-freiburg.de//Publications/2016/MIFDB16

[12] L. De-Maeztu, S. Mattoccia, A. Villanueva, and R. Cabeza, "Linear stereo matching," in *A13th International Conference on Computer Vision (ICCV2011)*, November 6-13 2011.

[13] J. Jiao, R. Wang, W. Wang, S. Dong, Z. Wang, and W. Gao, "Local stereo matching with improved matching cost and disparity refinement," *IEEE MultiMedia*, vol. 21, no. 4, pp. 16–27, Oct 2014.

[14] Y. Geng, "Local stereo matching based on information entropy of image," *3D Research*, vol. 7, no. 3, p. 27, 2016. [Online]. Available: http://dx.doi.org/10.1007/s13319-016-0103-x

[15] J. Sun, N.-N. Zheng, and H.-Y. Shum, "Stereo matching using belief propagation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 7, pp. 787–800, Jul. 2003. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2003.1206509

[16] L. Wang and R. Yang, "Global stereo matching leveraged by sparse ground control points," in *CVPR 2011*, June 2011, pp. 3033–3040.

[17] J. bontar and Y. Lecun, "Stereo matching by training a convolutional neural network to compare image patches," *Journal of Machine Learning Research*, vol. 17, 4 2016.

[18] D. Eigen, C. Puhrsch, and R. Fergus, "Depth map prediction from a single image using a multi-scale deep network," *CoRR*, vol. abs/1406.2283, 2014. [Online]. Available: http://arxiv.org/abs/1406.2283

[19] S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," *CoRR*, vol. abs/1504.03641, 2015. [Online]. Available: http://arxiv.org/abs/1504.03641

[20] J. Zbontar and Y. LeCun, "Computing the stereo matching cost with a convolutional neural network," *CoRR*, vol. abs/1409.4326, 2014. [Online]. Available: http://arxiv.org/abs/1409.4326

[21] W. Luo, A. Schwing, and R. Urtasun, "Efficient deep learning for stereo matching," in *CVPR*, June 2016.

[22] D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nešić, X. Wang, and P. Westling, *High-Resolution Stereo Datasets with Subpixel-Accurate Ground Truth*. Cham: Springer International Publishing, 2014, pp. 31–42.

[23] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *International Journal of Robotics Research (IJRR)*, 2013.

[24] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A naturalistic open source movie for optical flow evaluation," in *European Conf. on Computer Vision (ECCV)*, ser. Part IV, LNCS 7577, A. Fitzgibbon et al. (Eds.), Ed. Springer-Verlag, Oct. 2012, pp. 611–625.

[25] H. Hirschm, "Stereo processing by semiglobal matching and mutual information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 328–341, 2007.