ROLL NO: EE19S042

ASSIGNMENT 1 :

TASK1:

SUMMERY OF HIERARCHICAL DEEP STEREO MATCHING ON HIGH-RESOLUTION IMAGES

- This Paper works on the problems like memory constraints ,speed limitations ,accuracy ,lack of bench mark data ,on demand processing ,which most of the State of the art (SOTA) methods struggle to process in High resolution images.
- Previous methods-
  - SGM and ELAS **efficient matching algorithm** lack in accuracy and speed requirement in real time case.
  - Zbontar et al. And Luo uses **Siamese networks** to extract path wise features.
  - **Disp-Net-based architecture** adopts encoder decoder scheme with skip-connections to compute multi scale cost volumes by correlation and apply 2D convolution /UP convolution to refine predictions from coarse to fine
  - **PWC-Net** uses coarse to fine architecture to wrap features and achieve SOTA result in optical flow estimation.
  - **GCNet** constructs hierarchical 4D feature volumes and processes them from coarse to fine using 3D convolution.
  - Data Augmentation is used for getting training data which can be help full for real world calibration and rectification. A hierarchical stereo matching architecture is being proposed which can process high resolution images. High resolution synthesized data for training and real data for testing is collected. Stereo augmentation is used to improve robustness of the model.
  - An end to end frame work is being proposed for searching correspondences which being incremented over course to fine hierarchy.
  - A high resolution encoder feature is processed with coarse to fine decoder feature volumes whose resolution gradually increases.
- **METHODS:**

1. **HIGH RESOLUTION STEREO MATCHING NETWORK**:
   - It first down sample high resolution images while extracting multi scale features and use potential correspondences to build up a pyramid of cost volumes that increases in resolution.
   - Coarse to fine design principle-

     1.Spatial Pyramid Pooling(SPP) increase receptive field and pooled features kept in native coarse resolution to reduce memory.

     2.3-D convolution used to process high resolution cost-volumes efficiently.

     3.Multi scale loss functions.

   - Feature pyramid encoder to extract descriptors for coarse-to- fine matching. Encoder-Decoder architecture with skip connections for efficient feature extraction.
   - 4D Feature Volume is constructed with differences between pair of potentially-matching descriptors along horizontal scanlines. Pyramid of 4 volumes each with increasing spatial resolution and increasing disparity resolution is constructed.

- Feature volume encoder,

  1.Filtered by 6conv3D blocks

  2.Volumetric Pyramid pooling is applied to generate features that captures global context for high resolution input.

  3.Output is trilinearly-upsampled so that it can be fused with next 4D feature volume in the pyramid.

  4.For on demand disparities output is processed with another conv3D block to generate 3D output cost volume.

- Losses are scaled to account for the increased disparity level at each pyramid level.

2. STEREO DATA AUGMENTATION: Three Asymmetric augmentation techniques are used as follows,
   a. Y-disparity augmentation
   b. Asymmetric chromatic augmentation
   c. Asymmetric masking used to force model to rely more on contextual cues.
3. HIGH RESOLUTION DATA SET:

   High Resolution Virtual stereo is collected using carla simulator under 4 weather condition

   High resolution real stereo bench mark is collected using high-resolution stereo cameras and LiDAR.

Bench mark protocol used for ideal sensing range under different driving speeds.

- **Metrics:**

  1) bad-4.0 (percentage of “bad” pixels whose error is great than 4.0), bad-2.0 and bad-1.0, which tolerates the small deviations such as quantization error.

  2) avgerr (average absolute error in pixels) and rms (root mean- square disparity error in pixels), which also takes into account the subpixel accuracy.

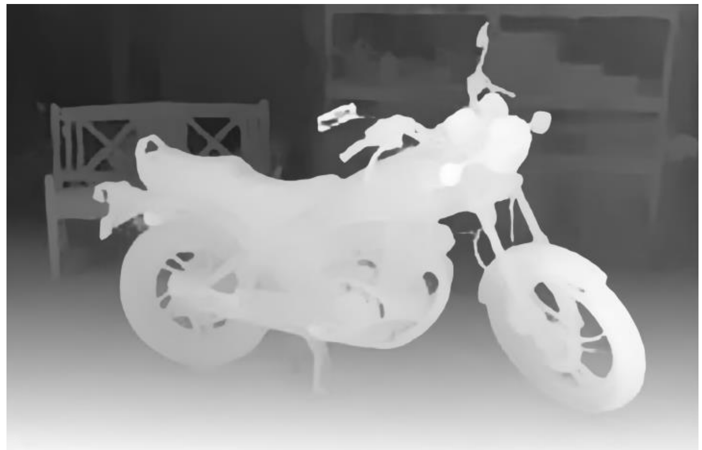
  3) A99 (99% error quantile in pixels), A95 and A90, which ignores large deviations while measuring the accuracy.

Referred paper link:
http://openaccess.thecvf.com/content_CVPR_2019/papers/Yang_Hierarchical_Deep_Stereo_Matching_on_High-Resolution_Images_CVPR_2019_paper.pdf

Link for Code : https://github.com/gengshan-y/high-res-stereo (Read me file to be referred from github folder)

Dataset Used: Middle_burry 2014 dataset

Disparity estimation using HSM algorithm

TASK 2 : The code for estimating disparity that are different from the ground truth and returns a percentage of the different pixels in the whole image using the choosen matching algorithm is being attached as Assignment1_EE19S042_Deep_learning-Task 2 folder.

TASK 3:

Matlab function that estimates disparity from the Middlebury stereo images with different block radii using stereo matching algorithm SAD is attached as .m file.

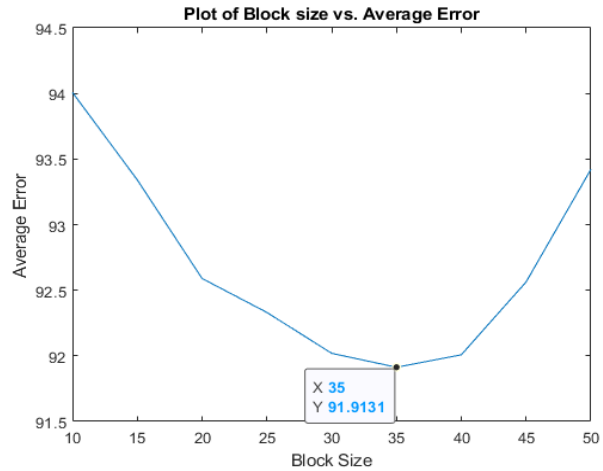
ART image from Middle Burry 2005.

LEFT

RIGHT

Disparity Image of ART

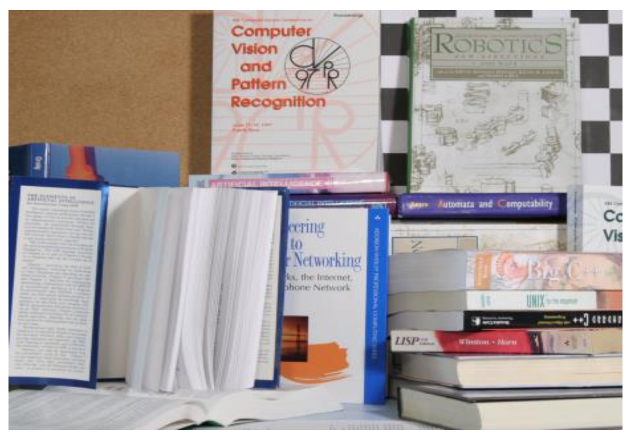
Plot of Block Size VS. Average Error of ART



Book image from Middle Burry 2005.

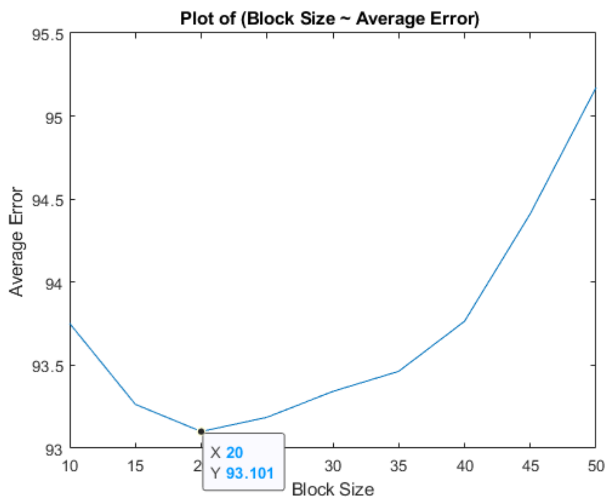
LEFT IMAGE OF BOOK

RIGHT IMAGE OF BOOK



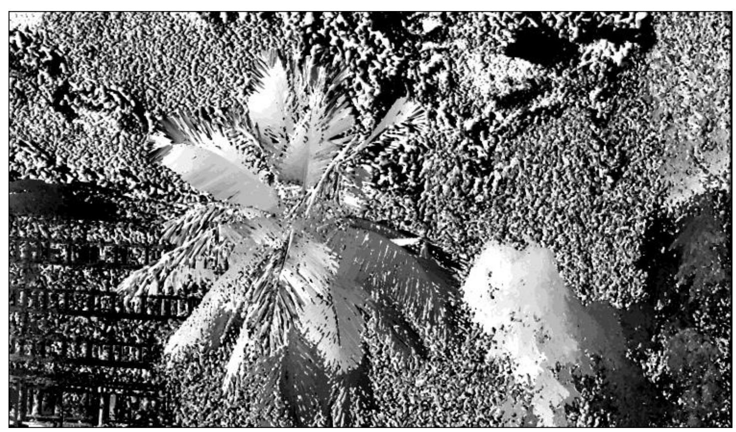
Disparity Image of ART

Plot of Block Size VS. Average Error of



The Minimum Average error came when block size is 20 .

DISPARITY IMAGE OF COCUNUT TREE:



TASK 4:

The code for rectification is attac

BEFORE RECTIFICATION :

ESB LEFT ESB RIGHT

AFTER APPLYING RECTIFICATION :

ESB ROAD LEFT ESB ROAD RIGHT

I. The Middle burry data depth map seems to be more continuous than the real image ones.
II. Yes ,there are some distinctive problems with some content type because of vertical disparity or luminance or camera orientation the disparity estimation being done was not coming properly. But after applying Rectification code , we could easily figure out the horizontal disparity.
III. For example as above the shown ESB road image has vertical disparity because of which the disparity-image initially coming