DEEP HIGH DYNAMIC RANGE IMAGING USING DIFFERENTLY EXPOSED STEREO IMAGES

Shashaank Aswatha Mattur
shashaank.aswatha.mattur@univ-poitiers.fr

Mohamed-Chaker Larabi
chaker.larabi@univ-poitiers.fr

CNRS, University of Poitiers, XLIM, UMR 7252, France

Abstract

- High dynamic range (HDR) image formation from low dynamic range (LDR) images of different exposures is a well researched topic in the past two decades
  - Most of the developed techniques consider differently exposed LDR images that are acquired from the same camera view point
  - Assumes the scene to be static long enough to capture multiple images
- Proposed technique: Addresses the problem of HDR imaging from differently exposed LDR stereo images
  - An encoder-decoder based convolutional neural network (CNN)
  - Does not require the LDR stereo images to be explicitly rectified and disparity corrected before merging to HDR image
    - Unlike conventional stereo matching methods
- Training and evaluation
  - An existing benchmark dataset of HDR stereo images, DML-HDR
- The experiments have shown some interesting results in comparison to the state-of-the-art approaches
  - End-to-end network is found to perform equally well on LDR images that are obtained from both stereo framework and single viewpoint

Objectives and Motivation

- High dynamic range (HDR) images capture more information of a scene than their corresponding low dynamic range (LDR) images with a better representation of real world illumination
- Most of the existing commercial HDR image acquisition systems capture HDR images by merging multiple LDR images that are captured from the same camera center sequentially
  - For dynamic scenes, such sequential capturing method introduces non-linear deformations between LDR images due to motion of objects in the scene
- The conventional approach for HDR imaging from stereo LDR images is a pipe-lining of two stages: rectification and/or disparity correction, and HDR merging
  - These geometric corrections make the HDR merging process computationally expensive and time consuming
  - Heavy networks

Literature

- Assume the scene to be static, and capture multiple LDR images of the scene at different shutter exposures that are combined to a HDR image by radiance mapping[1,2,3]
  - Since the LDR images are acquired from the same camera center sequentially, they lose structural consistency between them in a dynamic scene
- Align the sequential LDR images of a dynamic scene at a pivot point and locally register them for non-linear distortions, and then converted them to their linear exposure values and combine to obtain the HDR image using a deep convolutional neural network (CNN)[4,5]
  - Not plausible for abruptly changing dynamic scenes due to loss of coherency between the LDR images
    - The non-linear warping introduces some artifacts for abrupt scenes
- Use differently exposed LDR stereo images for forming a HDR image[6,7,8,9]
  - Require explicit geometric processes of rectification and disparity correction of stereo image pairs to account for perspective transformation between them
  - The HDR merging of the stereo images usually uses piecewise radiance mapping

1. P. E. Debevec and J. Malik, "*Recovering High Dynamic Range Radiance Maps from Photographs*", Int. Conf. Comp. Graph. and Interactive Techniques, 1997
2. P. Sen et. al., "*Robust Patch-based HDR Reconstruction of Dynamic Scenes*", Trans. Graph., vol. 31, no. 6, 2012
3. S. W. Hasinoff et. al., "*Burst Photography for High Dynamic Range and Low-light Imaging on Mobile Cameras*", Trans. Graph., vol. 35, no. 6, 2016
4. N. K. Kalantari and R. Ramamoorthi, "*Deep High Dynamic Range Imaging of Dynamic Scenes*", Trans. Graph., vol. 36, no. 4, 2017
5. Q. Yan et. al., "*Multi-Scale Dense Networks for Deep High Dynamic Range Imaging*", Winter Conf. on Applications of Comp. Vis., 2019
6. M. Batz et. al., "*High Dynamic Range Video Reconstruction from a Stereo Camera Setup*", Sig. Proc.: Image Comm., vol. 29, no. 2, 2014
7. S. Ning, et. al., "*HDR Image Construction from Multi-exposed Stereo LDR Images*", Int. Conf. on Image Proc., 2010
8. W. J. Park et. al., "*Stereo Vision-Based High Dynamic Range Imaging Using Differently-Exposed Image Pair*", Sensors, vol. 17, no. 7, 2018
9. J. Bonnard, et. al., "*Disparity-based HDR Imaging*", arXiv CoRR, vol. abs/1905.07918, 2019

Methodology

- An encoder-decoder based deep-CNN architecture for merging two differently exposed LDR stereo images into a HDR image
- Does not explicitly require the tasks of rectification and disparity correction
- Resolves rectification and disparity correction, within a tailored range, inherently while forming the HDR image using two stages of network
  - Feature extraction stage
  - Feature fusion stage

Methodology

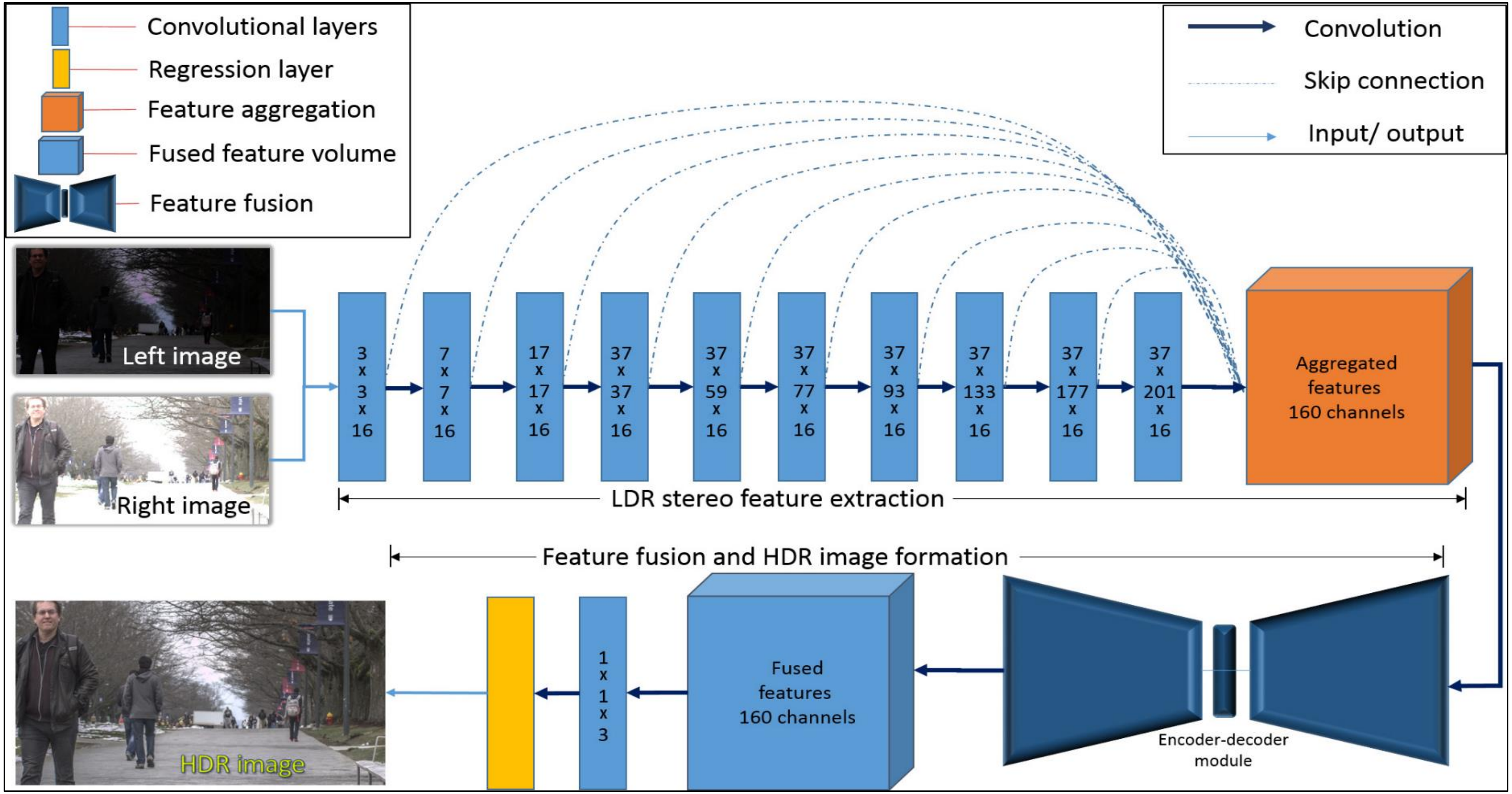

Fig.1: Proposed network for HDR image formation by corresponding LDR stereo image pairs with different exposures

Methodology

- Image references
  - Left camera view: Low exposure ($I_{l^{low}}$)
  - Right camera view: High exposure ($I_{r^{high}}$)
  - HDR image view: Left camera view (I_{hdr})
- The network learns the function $f(\bullet)$ by regressing over the HDR image of left camera view ($I_{l^{hdr}}$)
  - $I_{hdr} = f(I_{l^{low}}, I_{r^{high}})$
    - Subject to preset γ-correction
  - Loss function
    - $\Lambda_1 = \left\| I_{hdr} - I_{l^{hdr}}{}^{(\gamma)} \right\|_1$, the L_1 distance between I_{hdr} and γ-corrected $I_{l^{hdr}}$ and $\gamma = \frac{1}{2.2}$
- The fused HDR image is obtained by regressing directly on the reference HDR image of left camera view
  - No explicit consideration of disparity map or transformation function between the LDR images
  - Analogously seen to address three issues together: rectification, disparity correction, and HDR fusion
    - Separately dealt by individual steps in conventional approaches
- Results are evaluated with an inverse γ-mapping to match with the original scale of high dynamic range

Methodology

- Feature extraction
  - Dense features computed from the concatenated LDR stereo images, $H \times W \times 6$
  - Employ 10 consecutive 2-D convolutional layers
  - At each layer, 16 filters of stride 1 are used (refer Fig.1 for filter size, $h \times w$, in each layer)
    - Aggregation by residual connections after ten layers
  - Feature volume assembly of $H \times W \times 160$ for each image pair
  - Spatial support of feature volume
    - Spans a spatial range of 0 - 100 pixels in horizontal direction for disparity correction
    - Spans a spatial range of 0 - 30 pixels in vertical direction for rectification
    - Realized by varying the filter sizes that gradually increases with the depth of the network
      - Provides different extents of spatial coverage at different layers that are aggregated by depth concatenation
- Feature fusion
  - Computed by simple encoder-decoder module
  - The fused feature volume assembly of $H \times W \times 160$ is reduced to a dimension of $H \times W \times 3$ using 1×1 filters, which is regressed over the reference HDR image

Methodology

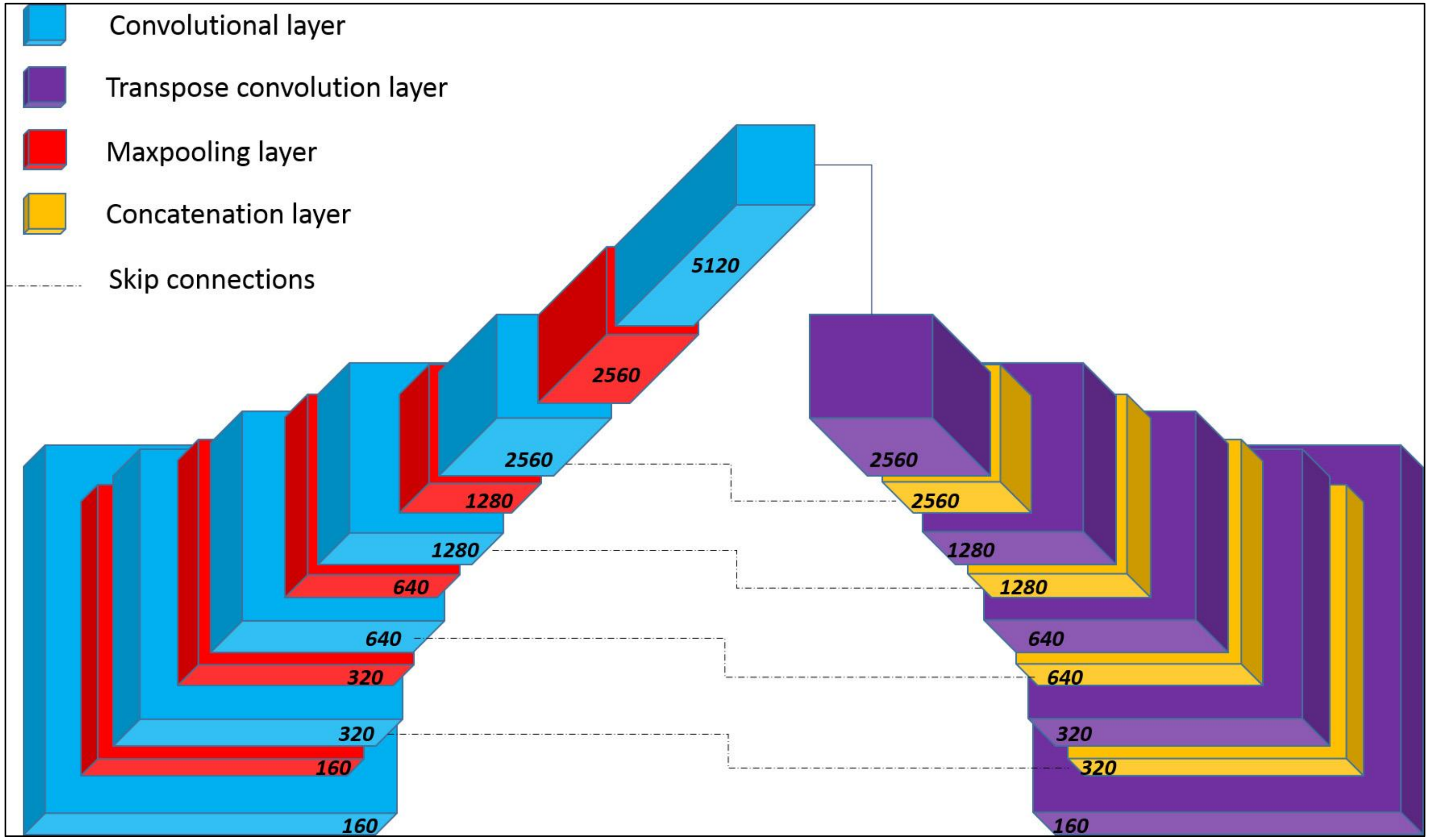

Fig.2: Layer representation of encoder-decoder module used in our network.

Dataset

- The network is trained using a stereo HDR video dataset, DML-HDR[1,2]
- The DML-HDR dataset consists of several stereo HDR videos that are acquired by RED SCARLET-X HDR motion footage capable camera
- Each video consists of hundreds of frames at Full HD resolution with a frame rate of 30 fps
- The image frames in the dataset have varied characteristics comprising of different composition, illumination, ambience, objects, etc., that are captured in both indoor and outdoor environments
- Data augmentation
  - Random rotations between -2° to 2°
  - Trained with image patches of size 256×512 that are randomly extracted from the training images with repetition
- HDR image visualization: Tone mapped to fit the dynamic range of normal displays[3]

1. A. B. Dehkordi, et. al., "*Compression of high dynamic range video using the HEVC and H.264/AVC standards*", Int. Conf. Heterogeneous Networking for Quality, Reliability, Security and Robustness, 2014
2. M. Azimi et. al., "*Evaluating the performance of existing full-reference quality metrics on high dynamic range (HDR) video content*", 2018
3. Z. Farbman et. al., "*Edge-Preserving Decompositions for Multi-Scale Tone and Detail Manipulation*", Trans. Graph., vol. 27, no. 3, 2008

Results

Fig.3: Example of ‘Walking on Snow’

Results

Fig.4: Example of 'Strangers'

Results

Fig.5: Example of ‘UBC’

Results

- Qualitative assessment
  - Manual observation in comparison against the ground truth data
  - No visibly implausible artefacts are observed in the HDR rendered images obtained from the proposed technique
- Quantitative assessment
  - Three quality metrics of HDR visual data evaluation: PSNR, SSIM, and HDR-VDP-2

Methods	PSNR	SSIM	HDR-VDP-2
Ning et. al.[2]	39.87	0.83	72.23
Bonnard et. al.[4]	43.28	0.93	89.37
Park et. al.[3]	40.93	0.86	78.58
Batz et. al.[1]	44.18	0.91	88.29
Proposed approach	45.96	0.95	93.89

1. M. Batz et. al., "*High Dynamic Range Video Reconstruction from a Stereo Camera Setup*", Sig. Proc.: Image Comm., vol. 29, no. 2, 2014
2. S. Ning, et. al., "*HDR Image Construction from Multi-exposed Stereo LDR Images*", Int. Conf. on Image Proc., 2010
3. W. J. Park et. al., "*Stereo Vision-Based High Dynamic Range Imaging Using Differently-Exposed Image Pair*", Sensors, vol. 17, no. 7, 2018
4. J. Bonnard, et. al., "*Disparity-based HDR Imaging*", arXiv CoRR, vol. abs/1905.07918, 2019

Results

Fig.6: Example of image acquisition by a stereo setup comprising of two mobile phones

Conclusion

- A two staged deep CNN based technique for forming HDR images using differently exposed LDR stereo image pairs is presented
  - Feature extraction phase and feature fusion phase
  - The network learns a function for combining the stereo LDR image pairs to a higher dynamic range
- Designed to handle geometric corrections like rectification and disparity correction between the LDR image pairs implicitly within a specific range
- The results are qualitatively and quantitatively shown to be at par with state-of-the-art techniques
- Future work
  - Future directions are laid towards forming HDR videos using LDR stereo framework
    - Applicability to render high quality and flicker-free HDR videos is yet to be explored

Acknowledgment: This research is funded by CPER NUMERIC - e-Creation project - Cinema