Name: Vaishnavi Ankaikar

Email: [ank.vaishnavi@gmail.com](mailto:ank.vaishnavi@gmail.com)

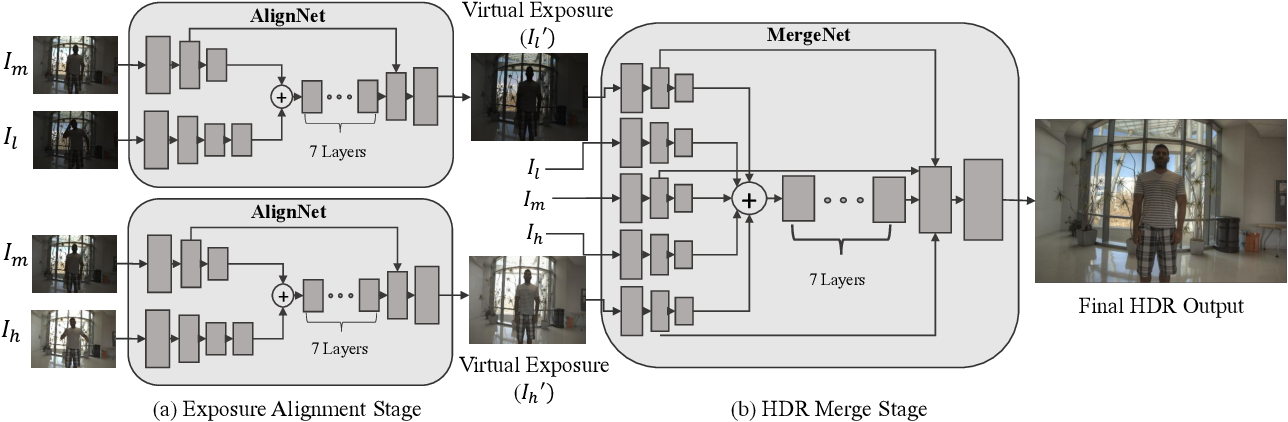
Ph no. 8380092207

**Report**

Problem 2 (Paper 1):

In [photography](https://en.wikipedia.org/wiki/Photography) and [videography](https://en.wikipedia.org/wiki/Videography), HDR or high-dynamic-range imaging is the set of techniques used to reproduce a greater range of [luminosity](https://en.wikipedia.org/wiki/Luminosity_function) than that which is possible with standard photographic techniques. Standard techniques allow differentiation only within a certain range of brightness. Outside this range, no features are visible because in the brighter areas everything appears pure white, and pure black in the darker areas. HDR is useful for recording many real-world scenes containing very bright, direct sunlight to extreme shade. [High-dynamic-range (HDR)](https://en.wikipedia.org/wiki/High_dynamic_range) images are often created by capturing and then combining several different, narrower range, [exposures](https://en.wikipedia.org/wiki/Exposure_(photography)) of the same subject matter. High dynamic range (HDR) imaging provides the capability to capture, manipulate and display real-world lighting, unlike traditional, low dynamic range (LDR) imaging. Several methods that can retarget LDR to HDR content are present. These methods make it possible to utilize and manipulate the vast amounts of LDR content within HDR pipelines and visualize them on HDR displays. However, such methods are primarily model-driven, use various parameters which make them difficult to use by non-experts, and are not suitable for all types of content. Recent machine learning advances for applications in image processing provide data-driven solutions for imaging problems, bypassing reliance on human expertise and heuristics. CNN’s are the current de-facto approach used for many imaging tasks, due to their high learning capacity as well as their architectural qualities which make them highly suitable for image processing. Deep Learning models, with their multi-level structures, are very helpful in extracting complicated information from input images. Convolutional neural networks are also able to drastically reduce computation time by taking advantage of GPU for computation, which many networks fail to utilize. A set of three input LDR images low exposure(Il), mid exposure(Im), and high exposure(Ih) is fed to the network, and the expected output is an HDR image free of ghosts and artifacts. Conventional multi-exposure fusion methods for HDR image generation often create unsightly ghosting artifacts in the presence of large scene motions. We propose to mitigate this problem by creating virtual exposures Ill and Ilh out of the image pairs (Im, Il) and (Im, Ih) respectively, and then blending (Ill, Im, Ilh ) to create ghost-free HDR images.

The final output generated is in two stages: a) The first stage (exposure alignment stage) involves the creation of virtual exposure images which match the structure of Im. Conventional methods for this task such as polynomial curve fitting fail to recover information in saturated regions. To overcome this limitation, we use a deep CNN (AlignNet), which can learn the required transformation and generate a better virtual exposure image. The virtual exposure images and the inputs are passed to the second stage,b) where another deep CNN(MergeNet) learns to generate an artifact-free HDR image.



Given an exposure guide image(Ie) and a structure guide image(Is), we aim to generate an output virtual exposure image(Io) using a CNN (AlignNet). The network has an encoder-decoder architecture with residual blocks, similar. The input images are encoded in separate branches for 2 layers using a stride of 2 to compress the information. An extra layer of stride 1 is added to the Ie branch to extract an abstract feature vector which is then appended to the feature vector produced by the Is branch and passed through 7 residual layers of stride 1. Finally, the output of the residual block is decoded back over 2 deconvolutional layers. We use batch normalization and leaky Relu activation function in the interior layers and tanh activation function in the output layer to scale the output to [−1, 1]. Skip connections [10] [11] are provided between the inner decoder layers and the encoder layers in the Im branch. Two instances of AlignNet were trained for (Im, Il) and (Il, Ih) pairs to generate Il and Ih respectively. The goal of the HDR merge process is to take the aligned LDR images, Il, lm, Ih, as input and produce a high-quality HDR image. Intuitively, this process requires estimating the quality of the input-aligned HDR images and combining them based on their quality. For example, an image should not contribute to the final HDR result in the regions with alignment artifacts, noise, or saturation. Generally, we need the aligned images in both the LDR and HDR domains to measure their quality. The images in the LDR domain are required to detect the noisy or saturated regions. Moreover, the images in the HDR domain could be helpful for detecting misalignments by measuring the amount of deviation from the reference image. Convolutional neural networks pre-trained for image classification have already learned to encode the perceptual and semantic information we would like to measure in our loss functions. So our loss functions depend on the network, the pre-trained network. We got two types of losses: a) Feature Loss b) Style Reconstruction Loss. Content loss (Lc) measures the fidelity of the generated output image (Io) with respect to the ground truth (Ig) target image. We use the mean of absolute differences for our content loss (L1 distance). Feature reconstruction loss(Lf ) provides a way to minimize the differences between high-level feature representations of the network-generated image and the ground truth. This loss helps in preserving the overall structure of the generated image.

**Conclusion:** We use a convolutional neural network to generate the HDR image from a set of images aligned with optical flow. To properly train the network, we proposed a strategy to produce a set of input LDR images.The proposed method learns HDR image generation over multiple stages as opposed to single stage end-to-end learning adopted by state-of-the-art deep HDR solutions. It handles scenes with large object motion much more effectively and is able to produce ghost-free images rich in dynamic range by hallucinating plausible details in regions with total occlusion. The proposed solution also scores better on several image quality metrics such as PSNR, SSIM, and HDR-VDP-2 as compared to the state of the art.

Yes, we can extend this approach to videos.

One of the major drawbacks of standard digital cameras is their inability to capture the full range of illumination in the scene. HDR imaging is now popular and available to the public through smartphone cameras and commercial software like Adobe Photoshop. On the other hand, HDR video remains out of reach for the public as the majority of approaches focus on specialized cameras. These cameras are often bulky and expensive since they need complex optical systems. To generate HDR videos using inexpensive off-the-shelf cameras, we can capture the input low dynamic range (LDR) sequences by alternating the exposure of each frame. The HDR video can then be reconstructed by recovering the missing HDR details at each frame, from the neighboring images with different exposure. We propose to address these problems by modeling both the alignment and HDR merge components using two sequential CNNs and train the two networks in an end-to-end fashion by minimizing the error between the estimated and ground-truth HDR frames. Alignment can be performed by estimating the flow between the neighboring frames and the current frame using a CNN which is specifically designed for this application and performs better than existing learning-based optical flow methods. These estimated flows are then used to warp the neighboring frames and produce a set of aligned images. As is common with deep learning systems, we need a large dataset of input LDR frames and their corresponding ground truth HDR frames to properly train our networks. We produce our training dataset by synthetically extracting the input LDR images from a set of HDR videos. To avoid overfitting to this synthetically generated dataset, we simulate the imperfections of standard digital cameras by adding noise to the input LDR frames and perturbing their tone. We propose the first deep learning approach to produce an HDR video from a sequence of alternating exposures. Then we present a flow network which is specifically designed for HDR video reconstruction application and performs better than the existing non-learning and learning-based optical flow approaches. •We apply necessary changes to the input and architecture of the merge network.