University of Michigan
EECS 442: Foundations of Computer Vision
Fall 2020. Instructor: Andrew Owens.

Problem Set 2: Signal Processing

Posted: Wednesday, Sept. 16, 2020 **Due:** Wednesday, Sept. 23, 2020

For Problem 2.1(b), 2.2(a), 2.3(c), please submit your written solution to Gradescope as a .pdf file. For Problem 2.1(a-b), 2.2(b), 2.3(a), please submit your solution to Canvas as a notebook file (.ipynb), containing the visualizations that we requested. Before submitting, please make sure to rename the file to `uniqname_umid.ipynb`.

The starter code can be found at:
https://colab.research.google.com/drive/1zhN_A84o5Hj2524SdeE7_6FD2OTfynxN?usp=sharing

We recommend editing and running your code in Google Colab, although you are welcome to use your local machine instead.

Problem 2.1 *Image blending*

Image pyramids are image representations useful for many downstream applications. This problem uses pyramid image processing to blend two images. [1]

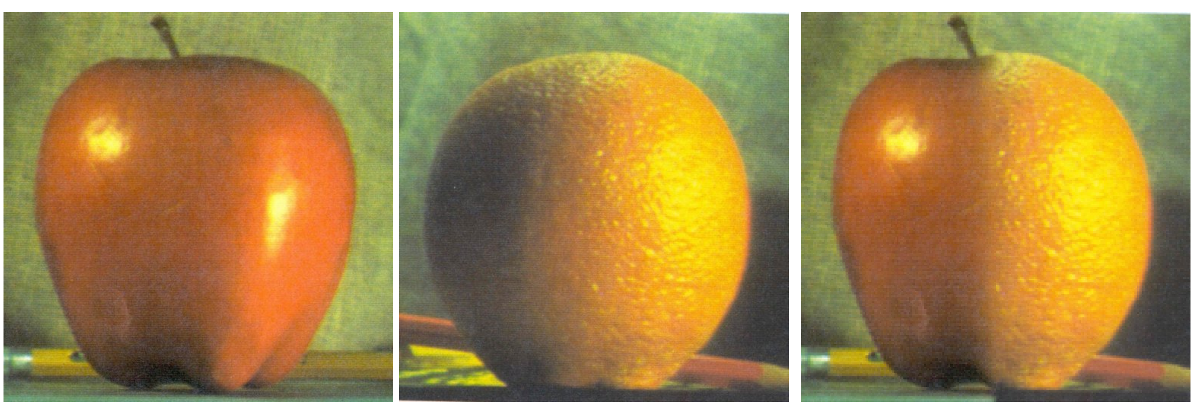

Figure 1: Blending with a Laplacian pyramid of 6 levels. Note that your result may look different than ours.

(a) As a first step, implement functions `pyr_up`, `pyr_down`, `gaussian_pyramid`, `laplacian_pyramid`, `reconstruct_img` to build a Laplacian pyramid from one image and show that you can reconstruct back the original image. Please use Gaussian kernels with the same size for `pyr_up` and

[1] This problem was originally based on a problem set by William Freeman and Antonio Torralba.

`pyr_down`. The only difference is that the kernel for `pyr_up` will be the one used for `pyr_down` multiplied by `4`. You can set the standard deviation of the Gaussian kernel as $\sigma = 1$. Please plot the original image, the Laplacian pyramid, and the reconstructed image. Please use a Laplacian pyramid with 4 levels. (Hint: `np.insert` may come in handy when implementing `pyr_up`.) **(4 points)**

(b) Implement functions `pyr_blend(im1, im2, mask, num_levels)` that takes as input two images and a binary mask (indicating which pixels to use from each image) and produces the Laplacian pyramids with `num_levels` levels for blending the two images. Use your function to blend the orange and apple image we provide in the Colab notebook. Plot the blended images with `num_levels` $\in \{1, 2, 3, 4, 5, 6\}$. Please describe the difference between the blended images with different levels of Laplacian pyramid: how does the result change as you use more pyramid levels?

To obtain color images, you can apply the blending to each color channel independently. In our implementation, this did not require any extra code (the same code worked on single-channel and multi-channel images due to `numpy` broadcasting), but your implementation may differ. **(2 points)**

(c) ***(Optional)*** Use your code to blend your own images. If you would *not* like your blending results shown in class, please let us know!

Problem 2.2 *Fourier Transform*

(a) Please match images on the left column of Figure 2 to its spectrum on the right. Provide your answers in a colon separated format. For example, 1:A, 2:E, ... **(1 point)**

(b) Convolve the provided image with a Gaussian filter: i) using direct convolution in the spatial domain, and ii) product in the frequency domain (via the convolution theorem). To perform DFT and inverse DFT, use `fft2` and `ifft2` from `scipy.fft`. For 2D convolution, we use `scipy.signal.convolve2d`. **(1 point)**

Problem 2.3 *JPEG image compression*

The dean has escalated his mysterious feud with the JPEG Standards Committee: beginning next week, UMich will, by his order, be entirely JPEG-free. Fearing the huge increase in storage costs from dealing with uncompressed images, U-M ITS has created a simplified version of JPEG that they plan to deploy to all of their data centers over the weekend[2]. They have turned to you for help to complete one tiny piece of the system: the frequency decomposition.

JPEG image compression takes advantage of the fact that human vision is less sensitive to high frequency components than to low frequency components. It is therefore possible to discard some high frequency components without significantly reducing the visual image quality. In this problem, you will implement a variation of the Fourier transform called the *discrete cosine transform* (DCT), and explore it via simple visualizations. After implementing

[2] Currently they have named it MPEG, but—fearing a two-front confrontation with the MPEG Video Committee—they have recently decided to rename it MichPEG

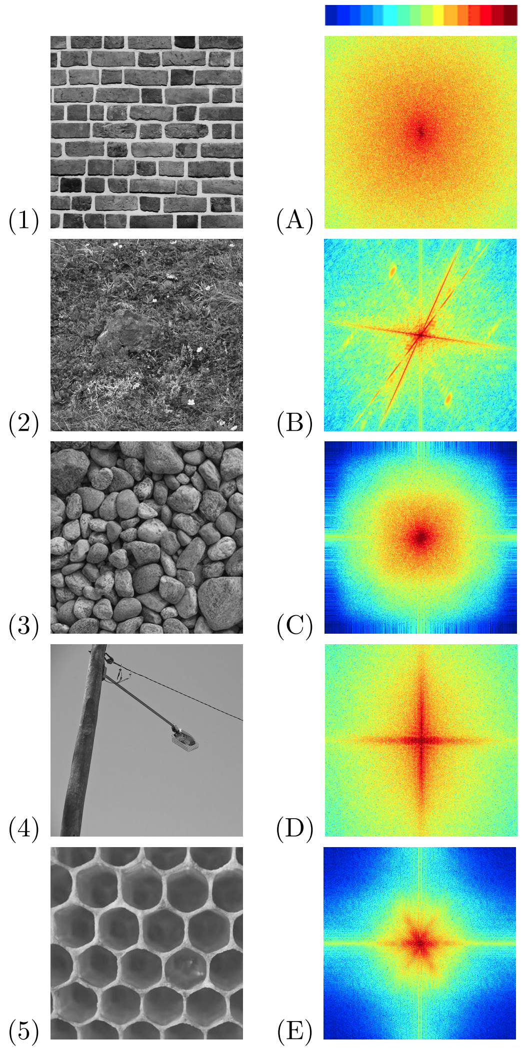

Figure 2: Images and DFT magnitude.

the DCT, you can plug it into U-M ITS's new compression system to confirm that the model leads to significantly smaller images.
Note: Despite the large amount of code we have given you for the JPEG compression system, this problem should only require a few lines of code!.

JPEG image compression can be divided into several steps:

1. Break the image into tiny 8×8 patches.
2. For each patch, apply the Discrete cosine transform (DCT) to obtain a frequency decomposition.
3. Quantization: high frequency DCT coefficients with small magnitude are set to zero.
4. Finally, we losslessly compress the quantized DCT coefficients. Since these contain a large number of zeros, they will compress much more easily than before quantization. This step uses Huffman coding (see `here` for more information, if you are interested)

Recovering the image from the encoded file follows the reverse of the above steps:

1. Decoding the Huffman-coded file to get the quantized DCT coefficients
2. De-quantization, where quantized DCT coeffecients are scaled back to the actual scale
3. Inverse DCT to reconstruct the image.

By following the above steps, we will compress a grayscale image into a compact encoded file, examine its new file size, and then decode that compact file and reconstruct the image. You will implement the DCT transform part (step 1). We have implemented the remaining code for you. Please note that the amount of code that you'll need to write is quite minimal!

Discrete Cosine Transform The DCT is very closely related to 2D Discrete Fourier Transform (DFT), which also describes how much of each frequency component an image contains. It produces results that are entirely real-valued (rather than complex), and it is often better-suited to processing images than the DFT (see Szeliski 3.4.2 for more information) . The 2D DCT of an $M \times N$ matrix A is given by:

$$B_{pq} = \alpha_p \alpha_q \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A_{mn} \cos \frac{\pi(2m+1)p}{2M} \cos \frac{\pi(2n+1)q}{2N}, \quad \begin{matrix} 0 \le p \le M-1 \\ 0 \le q \le N-1 \end{matrix} \tag{1}$$

The values B_{pq} are called the *DCT coefficients* of A. The inverse 2D DCT is given by:

$$A_{mn} = \sum_{p=0}^{M-1} \sum_{q=0}^{N-1} \alpha_p \alpha_q B_{pq} \cos \frac{\pi(2m+1)p}{2M} \cos \frac{\pi(2n+1)q}{2N}, \quad \begin{matrix} 0 \le m \le M-1 \\ 0 \le n \le N-1 \end{matrix} \tag{2}$$

where in both (2) and (3), α_p and α_q are given by:

$$\alpha_p = \begin{cases} 1/\sqrt{M}, & p = 0 \\ \sqrt{2/M}, & 1 \le p \le M-1 \end{cases} \qquad \alpha_q = \begin{cases} 1/\sqrt{N}, & q = 0 \\ \sqrt{2/N}, & 1 \le q \le N-1 \end{cases} \tag{3}$$

(a) A core component of 2D DCT is the 2D DCT basis, which is a set of 2D sinosiodal images with different spatial frequencies of the form $\cos \frac{\pi(2m+1)p}{2M} \cos \frac{\pi(2n+1)q}{2N}$. Implement function `build_2D_DCT_basis` that will compute this basis image set. Visualize the basis image set using the given code.[3] **(1 point)**

(b) ***(Optional)*** `DCT_2D` and `IDCT_2D` for performing 2D DCT and inverse 2D DCT are implemented using the built 2D DCT basis, read through the code of these two functions and make sure you understand them. **(0 point)**

(c) Now, the full JPEG compression system should be ready to run! In the encoding part of the code, try image compression using different *quantization tables*, including `dc_only`, `first_3`, `first_6` and `first_10`. Each of these tables discards a different set of frequency components; tables that discard more components produce a more compressed, but lossier, image. For each quantization table, report the compress ratio (size of the compressed file divided by size of uncompressed file). Also, please briefly comment on their reconstructed image quality and file size, and explain why it is the case. Hint: Take a look at the function `load_quantization_table` and function `quantize` to understand what they are doing. **(1 point)**

[3]Remarkably, if you were to use principal component analysis (PCA) to find a low-rank approximation to a collection of natural images, the results would look very similar to the DCT basis!

(d) ***(Optional)*** Read the Quantization and Huffman coding and decoding code if you want to understand JPEG more thoroughly. Try out the direct Huffman coding the image pixel values without DCT code block in the end. This will show that Huffman coding along is not able to compress the image effectively. **(0 point)**