Development and Analysis of Hybrid Images

Project Report

Computer Vision Course (CS 419/619)

Submitted By-
Shashank Giri (160001054)
Ashish Kumar Gour (160002008)
Saurabh Sharma (160002053)

Report Summary -

1. Introduction
2. Motivation and Brief Background
3. Work Done-
    a. Image set used
    b. Producing Hybrid Images
4. Results and Analysis
5. Conclusion
6. References

1) Introduction-

In this project, we have exploited the multiscale perceptual mechanisms of human vision to create visual illusions (hybrid images) where two different interpretations of a picture can be perceived by changing the viewing distance or the presentation time.

Hybrid images are generated by superimposing two images at two different spatial scales: the low-spatial scale is obtained by filtering one image with a low-pass filter; the high spatial scale is obtained by filtering a second image with a high-pass filter. The final image is composed by adding these two filtered images. Hybrid images, unlike picture mosaics, contain two coherent global image interpretations, one of which is of the low spatial frequencies, the other of high spatial frequencies. This project is based on the work by Schyns and Olivia for effectively analyzing hybrid Images.[1]

2) Motivation and Brief Background

Visual psychophysics research has shown that human observers are able to comprehend the meaning of a novel image within a short glance (100 msec). This phenomenal performance of rapid image understanding can be experienced while watching fast scene edits in an action movie or in a music video. Research in human perception has suggested that image understanding efficiency is based on a multi-scale, global to local analysis of the visual input [Burt and Adelson 1983; Majaj et al. 2002]: an initial analysis of the global structure and the spatial relationships between components guides the analysis of local details.The global precedence hypothesis of image analysis ("seeing the forest before the trees".[Navon 1977])[2] implies a coarse-to-fine frequency analysis of an image.

In successful Hybrid images, when one percept dominates, consciously switching to the alternative interpretation becomes almost impossible. Only when the viewing distance changes can we switch to the alternative interpretation. In a hybrid image it is important that the alternative image is perceived as noise (lacking internal organization) or that it blends with the dominant subband. Rules of perceptual grouping modulate the effectiveness of hybrid images. Low spatial frequencies (blobs) lack a precise definition of object shapes and region boundaries, which require the visual system to group the blobs together to form a meaningful interpretation of the coarse scale. When observers are presented with ambiguous forms they interpret the elements in the simplest way. Observers prefer an arrangement

having fewer rather than more elements, having a symmetrical rather than an asymmetrical composition and generally respecting other Gestalt rules of perception. Symmetry and repetitiveness of a pattern in the low spatial frequencies are bad: they form a strong percept that it is difficult to eliminate perceptually. If the image in the high spatial frequencies lacks the same strong grouping cues, the image interpretation corresponding to the low spatial frequencies will always be available, even when viewing from a short distance. By introducing accidental alignments it is possible to reduce the influence of one spatial channel over the other. For instance, the dog-eagle example discussed in the results and analysis section.

3) Work Done -

3.1) Image set used -

The goal of the project was to generate efficient hybrid images having two global perceptions superimposed in such a way that the perception changes from fine to coarse as the distance of viewing of the images is increased. At a closer distance the high frequency components are more refined but as the distance increases the low frequency components which are generally present as blobs become significant.

In this project, we have used images available publicly which look like they will produce good hybrid images. Images which show hybridization of different moods were clicked using a laptop webcam of our project team member. We have created our own images containing text and random patterns to create hybrid images which can be used as private fonts.

3.2) Producing Hybrid Images-

A hybrid image (H) is obtained by combining two images (I1 and I2), one filtered with a low-pass filter (G1) and the second one filtered with a high-pass filter (1 − G2): $H = I1 \cdot G1 + I2 \cdot (1 - G2)$, the operations are defined in the Fourier domain. Hybrid images are defined by two parameters: the frequency cut of the low resolution image (the one to be seen at a far distance), and the frequency cut of the high resolution image (the one to be seen up close). An additional parameter can be added by introducing a different gain for each frequency channel. For the hybrids shown in this paper we have set the gain to 1 for both spatial channels. We use **gaussian** filters (G1 and G2) for the low-pass and the high-pass filters. We define the cut-off frequency of each filter as the frequency for with the amplitude gain of the filter is 1/2.

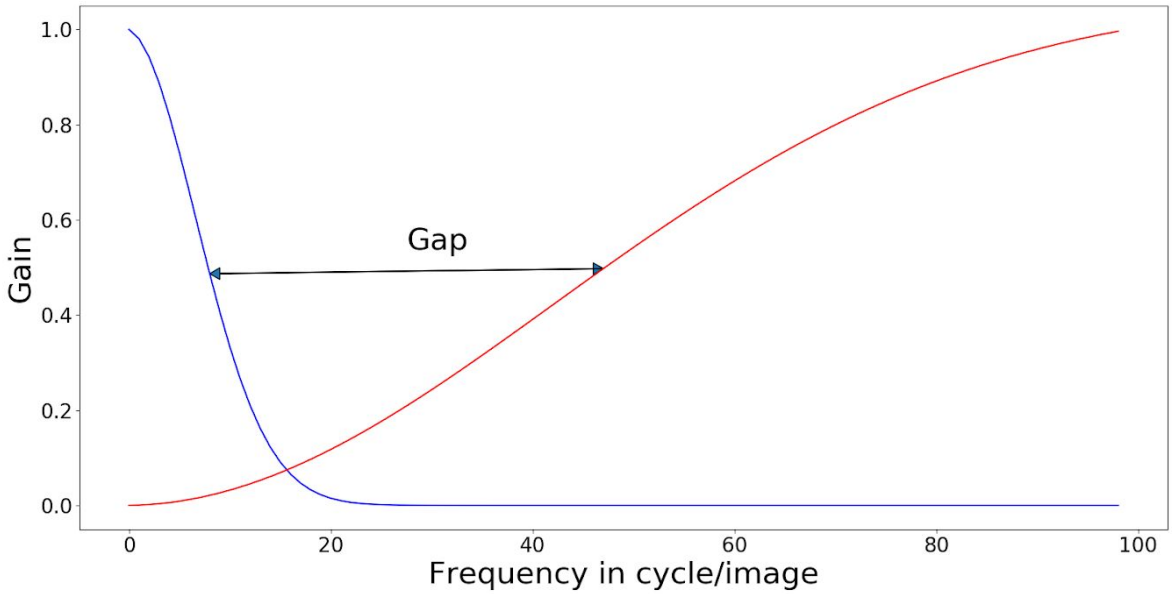
In [1] , the author takes the cutoff frequencies where the gain of both of the filters becomes ½ of the maximum. Since we have normalized filter outputs, the maximum gain of any filter is 1.
But in our work, we have taken the cutoff frequency to be the sigma(σ) of the Gaussian filter. This is because, at the input value of σ, the gain of the filter becomes 0.6 times of the maximum gain, which is pretty close to 0.5.

For our work the filter used were -

G1 - gaussian filter with $\boldsymbol{\sigma} = 10$

1- G2 - where G2 is also a Gaussian filter with $\boldsymbol{\sigma} = 40$

The below **figure** shows the gap which is created for the cutoff frequencies used.
Note, that the figure only shows frequency spectrum for one dimension. As a Gaussian filter over 2-D can be divided into two 1-D Gaussian filters, the results also hold for 2-D.



For producing successful hybrid images two properties must be satisfied :

1. Non dominant components are perceived as noise for both close and far viewing distances.
2. Maximum correlation between the edges so that they blend.

4) Results and Analysis

The below results show the hybrid images which were obtained as results for the cutoff frequencies explained above.

Near-Lion Far-Tiger

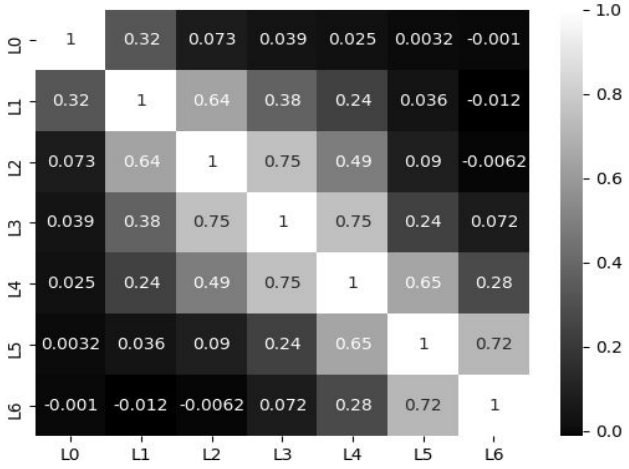
Near-Angry Far-Smiling

Near-Sad Far-Smiling

Effect of Alignment:

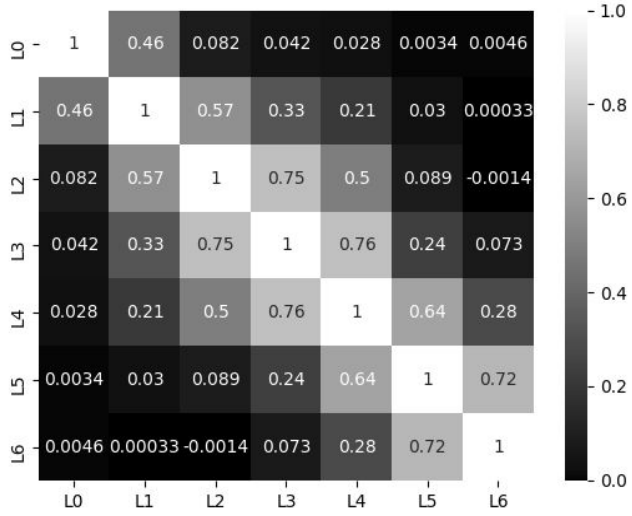
Alignment is also an important factor which affects the quality of hybrids produced. It is generally used for breaking any grouping cues which might be present for one channel. If not properly aligned, one channel will always dominate irrespective of distance of viewing.
Thus, by introducing accidental alignments it is possible to reduce the influence of one spatial channel over the other. This effect is depicted by the below hybrid images - The image is a hybrid of a dog(low frequency) and eagle (high frequency).

Aligned (Good)



Less Overlapping of two groups (Correlation of Laplacian Pyramids)

Non - Aligned (Bad)



Relatively More Overlapping of two groups

Effect of Color:

Color provides a very strong grouping cue that can be used to create more compelling illusions. For instance, in the following results shown, color is used only in the high spatial frequencies to enhance the bicycle and to reinforce the interpretation of the motorcycle as shadows when the image is viewed up close.

Colored and Mis-aligned(Worst)

Near - Cycle Far - Bike

Non-colored and mis-aligned(Better)

Near - Bike Far-Cycle

Non-colored and aligned**(Best)**

Near - Bike Far-Cycle

Effect of filter overlap:

The importance of correctly choosing the cut-off frequencies for the filters is illustrated in the following results . In the first figure, both filters have a strong overlap, and consequently, there is not a clean transition between the two faces.

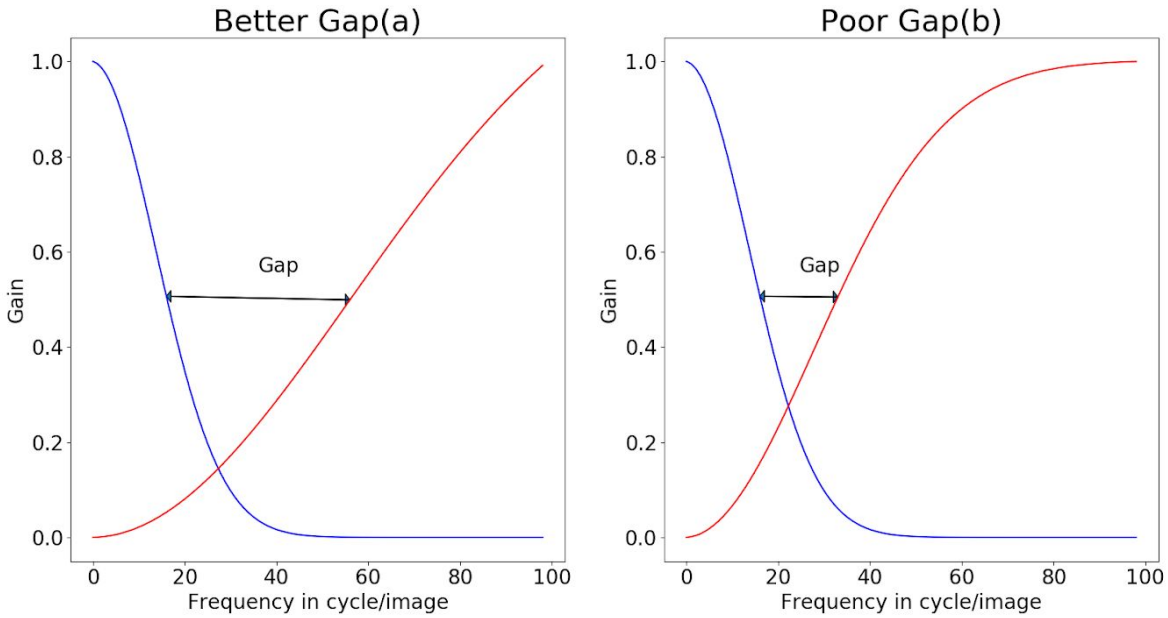
For the hybrid image on the next figure , the two filters have little overlap. The result is a cleaner image that produces an unambiguous interpretation (*Albert Einstein from up close and Marilyn Monroe from far away*). This is especially important when the images are not perfectly aligned.

Example of choosing proper thresholds: (In first Image - You can see Marilyn Monroe better at far distances).

Good Thresh (10, 40) (a)

Bad Thresh (10, 20) (b)

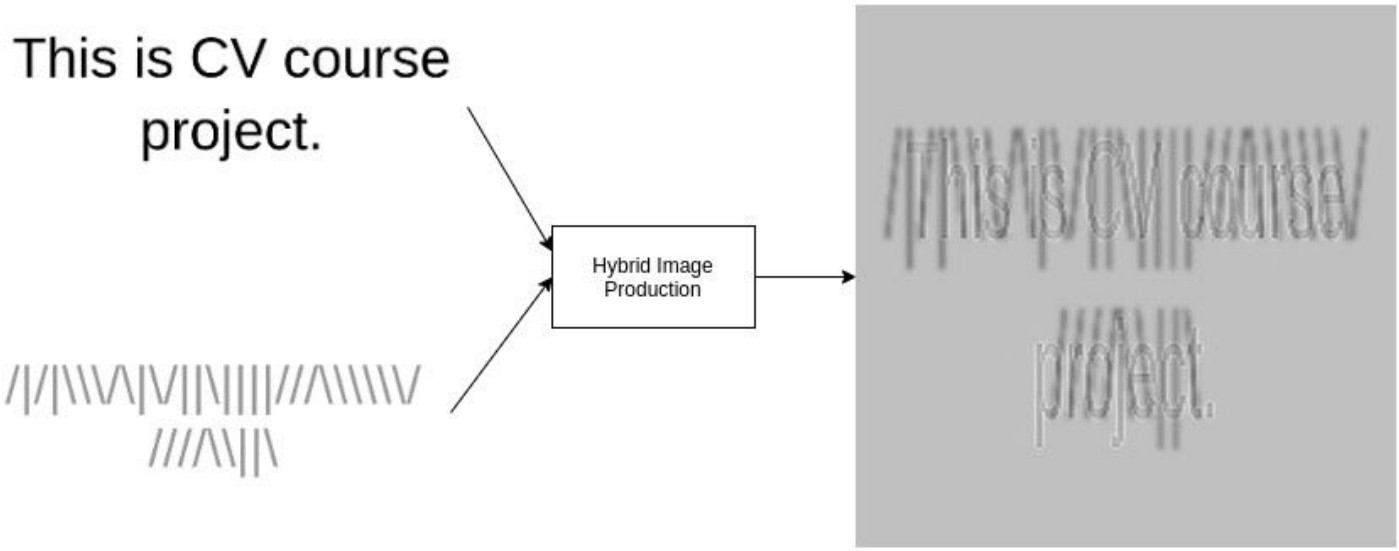
The below graph, shows the effect of thresholding. As shown, the figure (b) has not well formed gap between the cutoffs and hence the image shown above is also not of much good quality.



One interesting observation is that when the images are properly constructed, the observer seems to perceive the masked image as noise. Hybrid images break one important statistical property of real-world natural images i.e., the correlations between outputs of pass-band filters at consecutive spatial scales.

Private Font:

Hybrid images can be used to display text which can only be read when standing near to the text and couldn't be read from a distance. In the image shown, the text "This is CV course project." can be read when the hybrid image is close from the eye. From a distance, this text can't be read. The other image is used to mask the main text.



5) **Conclusion**

We have analysed the technique of generating good hybrid images which permits creating images with two interpretations that change as a function of viewing distance. Despite the simplicity of the technique, the images produce very compelling surprise effects on naive observers. They also provide an interesting new visualization tool to morph two complementary images into one. Creating compelling hybrid images is an open and challenging problem, as it relies on perceptual grouping mechanisms that interact across different spatial scales.

6) References

[1] **https://stanford.edu/class/ee367/reading/OlivaTorralb_Hybrid_Siggraph06.pdf** - By Aude Olivia, Antonio Torralba and Philippe. G. Schyns

[2] NAVON, D. 1977. Forest before trees: the precedence of global features in visual perception. Cognitive psychology 9, 353–383.