Generation of Hybrid and Blended Images Based on Multi-scale Representation

Stephannie Jimenez Gacha
Universidad de los Andes
Cra 1 Este No 19A - 40 Bogotá, Colombia
s.jimenez16@uniandes.edu.co

Sergio Galindo León
Universidad de los Andes
Cra 1 Este No 19A - 40 Bogotá, Colombia
sa.galindo10@uniandes.edu.co

1. Introduction

Human vision, perception and recognition is highly determined by multiple factors including brightness, contrast and exposure time to images. However, scale remains as one of the most important factor influencing perceptual mechanisms[3]. Schyns and Oliva inquired in the time and spacial scale dependency of scene recognition, finding that coarse information, encoded into low frequencies is mainly interpreted in distant scenes, while fine information, found in high frequencies, is processed in slow recognition tasks where details of nearby scenes corresponds to the most important information retrieved [1]. This is why vision corresponds to a multi-scale experience and should be analyzed as it.

By reversing Schyns and Olivia experiments it is possible to achieve at least two different interpretations of the same image by varying the viewing distance. An image of this type, containing different interpretations for low and high frequencies it is said to be a *Hybrid* image, one application of the multi-scale representation [2]. Generally, this kind of images correspond to the superimposition of two different images at different spacial scales, where one of them contains the high frequency information and one of them the low frequency from different scenes.

In order to built a *hybrid* image it is necessary to obtain one high-pass filtered image and one low-pass filtered image. For this purpose, one of the the simplest way to obtain a low pass filtered image is by means of Gaussian filtering. This filtering, produces a smoothed image that can be perceptually seen as a spacial scale transform for humans. In specific, it looks like an image seen from a distant point of view.

On the other hand, a high-pass filtered image could be obtained by the subtraction of an image with is low-pass filtered image, leaving only fine information. This process can be achieved systematically by the construction of Gaussian and a Laplacian pyramids. The first one corresponds to a collection of images obtained by successive sub-sampling and Gaussian filtering of an original image. The second one, the Laplacian pyramid, corresponds to collection of images obtained by the subtraction of a Gaussian pyramid level with the up-sampling of the next level.

Nevertheless, the construction of a *hybrid* image should consider spacial characteristics too. The images used should posses the same dimensions, similar borders, and congruently aligned shape to generate an adequate perception of the frequency band selected, given by the observer distance, without perceiving interference from other bands. As a special feature, the cutoff frequencies of the filters should be taken into account, as they must be separate enough to avoid unambiguous interpretations from any distance[2].

In addition, another application of the multiscale representation of images is the construction of *blended* images. In these, a soft transition between two halves of different images is achieved by upsampling concatenated halves of low resolution, obtained by means of a Gaussian pyramid until the desired scale. For This, the images should be symmetric, with fitting borders and must be aligned previously.

As part of the study of the multi-scale representation of images, a methodology to construct *hybrid* and *blended* images based on Gaussian and Laplacian pyramids is presented. The main characteristics of the filters and images will be analyzed as well as the final output respecting the expected features of this kind of images.

2. Materials and Methods

The images used to create the hybrid and blended images belong to a personal collection of family and friends. We used two pictures. Both of them corresponds to portrait pictures of male subjects of approximately the same age. One of the pictures was taken in outdoors while the other was taken indoors, both with high luminosity. The first one, with a smiling man was crop from 640x640 to 340x416 1. The second one, in which a bearded man appears was cropped from 707x904 to 340x416 too2. Also, an alignment was performed using the second image as template. The chin, mouth and eyebrows were aligned by scaling and a 4° counterclockwise rotation, however, because of anatomical differences some little gap in between hair and eyes remain.

After that, cv2, matplotlib and numpy in Python 3 was used to obtain the low frequency components of the first image using convolutional filtering with a Gaussian kernel of 51x51 and a standard deviation of 8. On the other hand, the high frequency components of the second image were obtained by the subtraction of the original image with the filtered image using a Gaussian kernel of 25x25 with a standard deviation of 50. To create the hybrid image both processed images were added.

Finally, a low resolution representation of both images was obtained by the construction of a Gaussian pyramid for each image with the cv2 function PyrUp() (downsample factor of 2)5. Using the fifth level of each pyramid the left half of the first image and the right half of the second image were concatenated. Then, to create a blended image, the low resolution representation was upsampled the same five levels of downsampling while the laplacian pyramid level was added each upsampling.

3. Results

A hybrid image with the high frequency components of the bearded men and the low frequency components of the smiling men was constructed. Viewed from a short distance the image is interpreted as a serious man with a beard and black hair, with a little background noise of red patches. Viewed from a distance of approximately 5 meters, the image is interpreted as a smiling man in outdoors with little noise of a white silhouette in the upper part (belonging to high frequencies)3.

The effect was acquired due to the appropriate distance in the filters cut frequencies, given by the significant difference between the standard deviation of the kernels (5 Vs 50). Also, a coherent image was interpreted from each point of view because of the correct alignment. However, some interference between both frequency bands is identified in the face borders. The image was constructed using $H = G(I1, T1) + (I2 - G(I2, T2))$ where H is the hybrid image, G() a Gaussian filtering and I_i each image.

A blended image using both previous images was constructed by concatenation of two halves at low resolution. After that, up-sampling and the corresponding Laplacian pyramid level was added. We obtained a soft transition between both faces, with some blurring due to the information loss caused by the Gaussian filtering and up-sampling. The image contains many black spots corresponding to zeroed pixels when effectuating the subtraction of images during pyramid construction4.

The blending effect was generated due to the interpolation by up-sampling of the contrast differences from a low resolution level to a high resolution level. In this process a transition band is formed in which the contrast information, that has been partially lost in the upper pyramid levels smooths6.

4. Conclusions

Multi-scale representation of visual information allows different interpretations for the same scene. When an image is observed from a short distance the high frequency components is privileged while, if seen from a long distance, the low frequency information gets more importance. This characteristic of human perception, moderated by spacial and temporal features allow the construction of hybrid images.

To obtain low-pass and high-pass filtered images, the use of convolution with an appropiate Gaussian kernel is needed. The low pass filtered image corresponds to the convolution between the image and the kernel and the high pass filtered image corresponds to the subtraction between the original image and the low frequency components. To obtain a proper hybrid image, the cutoff frequencies of the kernels should be distant to avoid ambiguous interpretations to the viewer at some intermediate distances.

One tool to organize and represent the multi-scale representation of visual information are Gaussian and Laplacian pyramids. In this representations we can obtain successive filtering and sub-sampling of an image in order to obtain high and low frequency information and perception at different spacial scales.

Finally, to obtain a smooth transition between two images (blended image) it is possible to concatenate them at a low resolution level and upsample until the desired resolution. To avoid information loss from high frequencies the corresponding Laplacian level should be added in each iteration.

References

[1] A. Oliva and P. G. Schyns. Coarse Blobs or Fine Edges? Evidence That Information Diagnosticity Changes the Perception of Complex Visual Stimuli. *Cognitive Psychology*, 34(1):72–107, 10 1997.

[2] A. Oliva, A. Torralba, P. G. Schyns, A. Oliva, A. Torralba, and P. G. Schyns. Hybrid images. In *ACM SIGGRAPH 2006 Papers on - SIGGRAPH '06*, volume 25, page 527, New York, New York, USA, 2006. ACM Press.

[3] P. G. Schyns and A. Oliva. From Blobs to Boundary Edges: Evidence for Time- and Spatial-Scale-Dependent Scene Recognition. *Psychological Science*, 5(4):195–200, 7 1994.

5. Code and Images

The code is available at the team's repository at *https://github.com/steff456/IBIO4680/tree/master/04-Hybrid/Answers* The original images are shown in figures 1 and 2. Both are part of the personal collection of images from the team members.

Figure 1. Original image with a smiling man

Figure 3. Hybrid image obtained from the original images

Figure 2. Original image with a bearded man

Figure 4. Blended image obtained from the original images

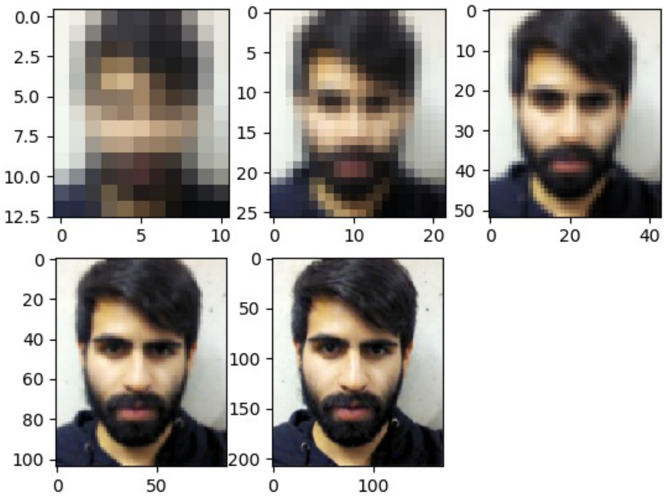

Figure 5. Gaussian pyramid for one of the images

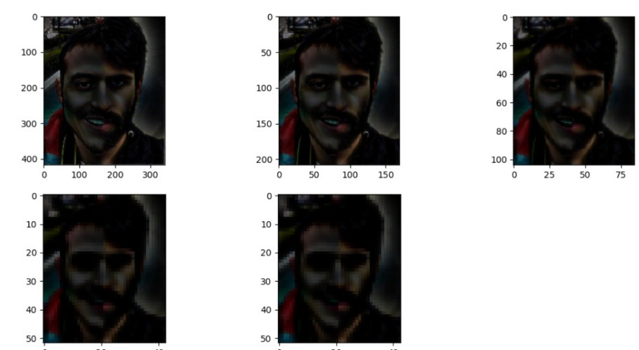

Figure 6. Blending pyramid