IPSJ SIG Technical Report

Hybrid Image From Different Contents

Peeraya Sripian[†1] and Yasushi Yamaguchi[†1]

Hybrid image is the image that changes interpretation according to viewing distance. By simply extracting the low and high spatial frequency bands from two source images, the combination result image can be interpreted differently viewing distance. This research finds the way to allow construction of hybrid image regardless of source images shape. Without the need to carefully pick the two images to be superimposed, hybrid image can be extended to use with any kind of image contents. We present the idea that by enhancing detail, by compressing dynamic range and by intentionally inserting noise in the high frequency image, any alternative low frequency image can be perceived as meaningless noise in a close viewing distance.

1. Introduction

Hybrid image was introduced by Oliva et al in 1). It provides a new paradigm in which image's interpretation can be modulated by playing with viewing distance. For a given distance, a particular band of spatial frequency dominates human visual processing. Hybrid image (H) is obtained by combining two images (I_1 and I_2) at different spatial scales:

$$H = Lp(I_1) + Hp(I_2),$$

where Lp is a low pass filter and Hp is a high pass filter. According to contrast sensitivity function, human observer can see a sine wave grating g at 2 to 5 $[cycles/degree]$ at the lowest contrast. In this work, we use 2 $[cycles/degree]$ for high frequency image and 6 $[cycles/degree]$ for low frequency image.

The cutoff frequency C $[cycles/image]$ used to design high and low pass filter is determined according to distance. C is defined as $C = \theta g$, where θ is degree of visual angle per image and is calculated by $\theta = ((\arctan(h/2)/d)(180/\pi)2$, where d is image height, d is the distance from viewer to image (we use 40cm for high frequency image and 400 cm for low frequency image).

†1 the University of Tokyo

2. Previous work

The method of exploiting multiscale perceptual mechanisms of human vision is originally proposed by Schyns and Oliva (2), 3), 4)). Their researches attempted to study human ability in analyzing global and local structure of the visual input. With a hypothesis that based on Navon's work in 5), many researches proved that human visual perception is done in a coarse-to-fine frequency analysis. That is, the low spatial frequency components are carried by the fast magnocellular pathway, while dominate early visual processing. (6), 7), 8), 2), 9))

Oliva et al. described their motivation of creating hybrid image in 1). They tested the role of different frequency bands with hybrid visual stimuli, i.e. a hybrid image. They found that in their experiment, the participants were unaware that the visual stimuli had two interpretations in limited duration of viewing time. Their research also found that a hybrid face of different emotions is treated differently depending on task. From those works, they found that when visual system selects a spatial frequency, it often unconscious of the information in the other spatial scale.

1) proposed the method of creating hybrid image that allows an interesting new visualization tool to morph two complementary images into one. Unfortunately, due to rule of perceptual grouping, the method limits to images that contain similar shape or global characteristics only.

Konishi and Yamaguchi attempted to solve this limitation in 10) by inserting noise into high frequency image, in order to cover the entire low frequency image which contain totally different shape. Moreover, some visual artifacts may appearwere visible in the resulting image. Due to too much noise, high frequency image was difficult to perceive.

3. Background

3.1 Poisson Reconstruction

Given a 2D field of 2D vectors $\mathbf{v}$. An image I with a gradient ∇I can be built as close as possible to $\mathbf{v}$ in the least square sense. This can be achieved through

Poisson equation:

$$\frac{\partial I}{\partial t} = \Delta I - \mathrm{div}(\mathbf{v}).$$

A new image is obtained by solving a Poisson equation with the divergence of this vector field under Neumann boundary conditions specifying that the value of the gradient of the new image in the direction normal to the boundary is zero.

3.2 Bilateral Filtering

The bilateral filter[11] smooths the input image while preserving its main edges. Each pixel is a weighted mean of its neighbors where the weights decrease with the distance in space and with the intensity difference. Bilateral filtered image $BI(\mathbf{p})$ at pixel $\mathbf{p}$, with $\mathrm{g}_\sigma(x) = \exp(-x^2/\sigma^2)$ as a Gaussian function is given as below:

$$BI(\mathbf{p}) = \frac{\int_{n(\mathbf{p})} \mathrm{g}_{\sigma_s}(\|\mathbf{p}-\mathbf{q}\|)\mathrm{g}_{\sigma_r}(|I(\mathbf{p})-I(\mathbf{q})|)I(\mathbf{q})d\mathbf{q}}{\int_{n(\mathbf{p})} \mathrm{g}_{\sigma_s}(\|\mathbf{p}-\mathbf{q}\|)\mathrm{g}_{\sigma_r}(|I(\mathbf{p})-I(\mathbf{q})|)d\mathbf{q}},$$

where $n(\mathbf{p})$ stands for a neighborhood of $\mathbf{p}$, σ_s and σ_r is the spatial neighborhood and intensity difference accordingly. k is the normalization factor.

3.3 Histogram Equalization

Histogram manipulation can be used effectively for image enhancement. Histogram equalization is the method used to spans full spectrum of histogram for the given area, normally used for imagethat have low contrast, i.e. background and foreground are both bright or both dark. Let variable r stands for the gray level of the image to be enhanced and $T(r)$ is the transformation function as below:

$$T(r) = \int_0^r p_r(w)dw, \qquad 0 \le r \le 1, \quad 0 \le T(r) \le 1, \tag{1}$$

where $p_r(w)$ is a probability density of pixel value w. The equation represents probability function of r, i.e. cumulative distribution function of r. The new value is obtained by looking at probability value, or histogram, of r to get histogram equalized image.

4. Implementation

Our work focuses on constructing a hybrid image that does not require image alignment or similar shape or source image. To hide an alternative image, (in this case, low frequency image) with fewer artifacts to be seen at a close viewing distance, noise-like pattern needed to be produced as in 10). Noise is produced in high frequency image in a non-random manner. In addition, both source images contrast are manipulated before processing the extraction of high, low frequency images.

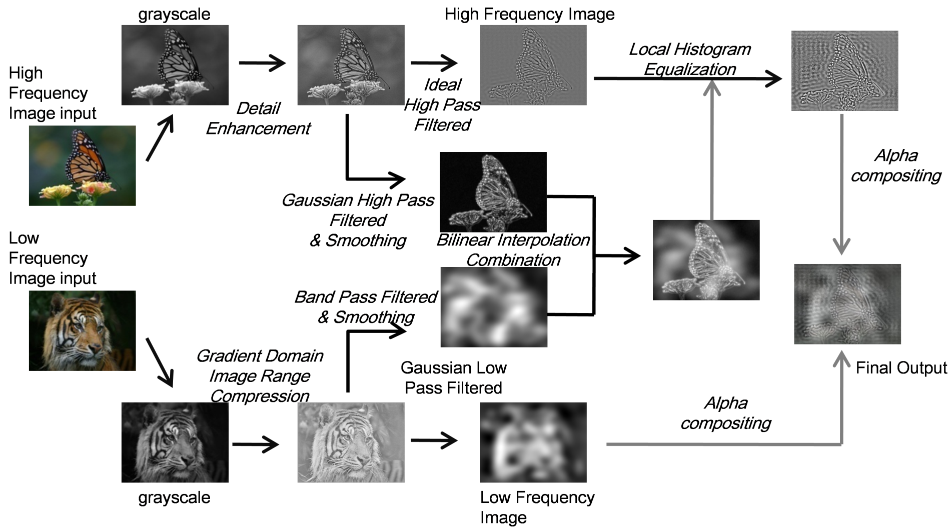

Fig. 1 Overview of our pipeline.

Our goal is to create noise in high frequency image that can cover most part of low frequency image. To prevent ambiguous result from inserting too much noise everywhere in the image, we consider original local frequency location of high frequency image and of low frequency image. Hence, local frequency map for each image is created.

In this section, the construction of low frequency image, high frequency image and each local frequency map is separately explained.

4.1 Low Frequency Image

Blob of low frequency image is perceived as meaningful interpretation in a far distance. Since we attempt to hide low frequency image by viewing in close distance, we expect low frequency to have make dynamic range small, i.e. low

Fig. 2 original grayscaled image.

Fig. 3 GDC image.

contrast.

(1) **Preprocessing: dynamic range compression**

With dynamic image range compression, we intend to brighten the dark area and darken the bright area. Of all methods of dynamic range compression, we experimented that gradient domain image range compression (GDC) provided satisfactory results.

GDC[12] is done by manipulating the gradient field of the luminance image and solve Poisson equation to obtain a result image. The method emphasized on detecting gradient of the image and shrinking gradients of large magnitude more than small ones. Originally this work is intended to compress dynamic range for High Dynamic Range image (HDR) to be displayed on a LDR display; however the work suggested that it can also be used to enhance LDR images as well. By attenuating strong gradients and reconstructing image back to the original range, small contrasts in dark regions become easier to see (**Fig. 3**). More detail about gradient attenuating function can be found in 12).

(2) **Extracting low frequency**

Low frequency is extracted through a Gaussian low pass filter in the frequency domain similar to previous work 1), 10).

(3) **Local frequency map for low frequency image**

Here we calculate high frequency position for low frequency image. This local frequency map (LF map) is used later in creating final local frequency map used in local histogram equalization for high frequency image. In order to know prospective location of high frequency, we need to isolate only selected frequency band. This operation is done by taking power of band pass filtered image. LF map (**Fig. 4**) is obtained by smoothing this image by Gaussian filter to avoid zero crossing positions and to compute local average.

4.2 High Frequency Image

High frequency image is used to hide shadow-like blobs of low frequency in composited image. Konishi and Yamaguchi[10] relied only on the ringing that is extracted from ideal high pass filter to create noise. However, we found that too much use of ringing led to the difficulty in recognition of the high frequency image when the viewer is close. As a solution, our work expects as less ringing as possible.

(1) **Preprocessing: detail enhancement**

With the manipulation of the detail in an image, non-visually noticeable detail such as small noise produced from high ISO value or variation of image value in dark region can be enhanced for visually perceptible. There are numerous methods in enhancing image detail; Bae et al.[13] gave an impressive result with simple algorithm. Their work is an image tone management that relies on two-scale decomposition based on the bilateral filter[15]. They also introduced gradient correction to prevent gradient reversals upon the re-combination of enhanced detail layer and base layer.

We begin by splitting input image I (in this case, an image to extract high frequency) into base B and detail D layer by bilateral filtering. The detail layer is enhanced by multiplying all pixels with scalar value, then is corrected by gradient reversal removal. Gradient reversal removal is achieved by directly constraining the gradient of the detail layer to have the same sign as the input derivatives and amplitude no greater than them. The corrected detail layer D' is obtained by solving Poisson equation. Finally, base layer can be updated by $B = I - D$. This approach keeps the original global gradient of the processed image to be the same as original image.

(2) **Extracting high frequency**

Similar to Konishi and Yamaguchi[10] we maintained the use of ideal high pass filter $G_H(u, v)$ in order to extract high frequency image with ringing at the edges.

We derived cutoff frequency as defined in Introduction section.

(3) **Local frequency map for high frequency image**

Local high frequency map (HF map) determines the location of high frequency in high frequency image. It can be calculated by filtering original image with a Gaussian high pass filter in a frequency domain.

To identify local average, we use Gaussian filtering similar to local low frequency map above.

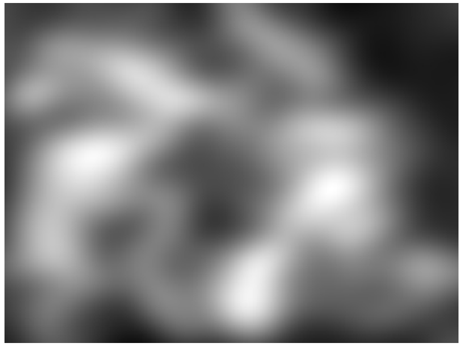

Fig. 4 LF map.

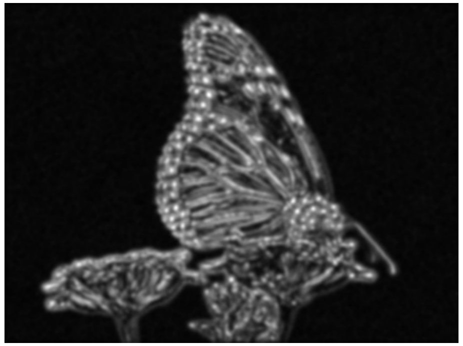

Fig. 5 HF map.

(4) **Map blending**

Before proceeding to local histogram equalization, we blend both frequency maps together with bilinear interpolation. Final local frequency map M is calculated as follows:

$$M(\mathbf{p}) = (k_L(1 - l(\mathbf{p})) + k_U l(\mathbf{p}))(1 - h(\mathbf{p})) + h(\mathbf{p}).$$

Here $l(\mathbf{p})$ and $h(\mathbf{p})$ is the pixel value of LF map and HF map accordingly. k_L and k_U give the lower and upper bound of the local frequency map when h is zero.

(5) **Local histogram equalization**

Although ringing obtained by ideal high pass filter can create some noises, its contrast is low and gradually fades off distance to original edges. Konishi and Yamaguchi[10] addressed this problem by manipulating the contrast locally according to high frequency map. Unfortunately, their contrast enhancement method and map construction led to visual artifact. Our work corrects contrast adjustment step by proposing local contrast enhancement with histogram equalization according to the new frequency map.

A histogram equalized image $E(\mathbf{p})$ at position $\mathbf{p}$ of the high pass filtered image $H_pI(\mathbf{p})$ is obtained by the following expression:

$$E(\mathbf{p}) = T_{\mathrm{w}}\left(H_pI(\mathbf{p})\right) c(\mathbf{p}) + 0.5(1 - c(\mathbf{p})).$$

Here $T_{\mathrm{w}}(r)$ is a transformation function of histogram equalization (1) within a window w around the pixel $\mathbf{p}$, and $c(\mathbf{p})$ represents a contrast defined by the map value $M(\mathbf{p})$ as below:

$$c(\mathbf{p}) = c_{\min} + (c_{\max} - c_{\min})M(\mathbf{p}),$$

where $c_{\min}$ and $c_{\max}$ are user-defined values standing for the minimum and maximum contrasts. We construct the locally equalized histogram without producing noise artifacts as seen in **Fig. 7**.

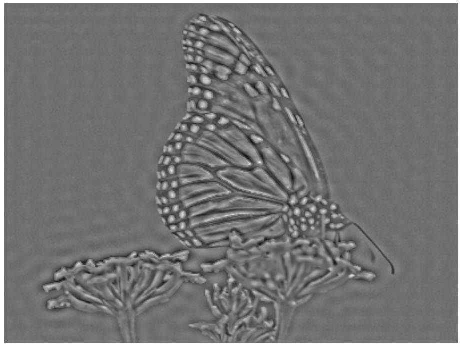

Fig. 6 High frequency image

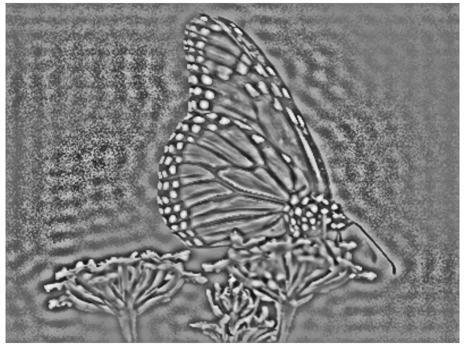

Fig. 7 After histogram equalization

5. Discussion

We found that our method can successfully create hybrid image from different contents. While looking closely, the user perceived low frequency blobs as meaningless shadow, when step a few meters away from the image, low frequency image is presented. However, the more visibility high frequency image is, when seen from far distance, it becomes artifacts to low frequency image and perceived unnatural. High frequency image is usually preferred when it is colorized; therefore choosing image with originally solid and bright colors is recommended.

Furthermore, choosing image to be low frequency image is still limited to image that contain simple shape structure, or content that easily understand. In our work, we choose images of animal's face to be low frequency side. To extend hybrid image to use with any type of images combination is still an open problem.

6. Conclusion

We have presented an approach to allow construction of hybrid image with different image contents, i.e., image shapes. By enhancing and by intentionally inserting noise in the high frequency image, different content of alternative low frequency image can be perceived as meaningless noise in a close viewing distance. The creation of image noise uses the special characteristics that naturally comes when extracting frequency bands from the image itself. Both image contrasts are carefully adjust prior to the frequency extraction to reduce visual artifacts. Finally, high frequency image noise is enhanced by locally histogram equalization using frequency maps introduced in our work.

This work still opens for several areas for future research. It provides a novel method in art, advertising, and human visual perception. More experiments can be done regarding visual perception to image frequency. Finally, to create hybrid image from any type of images with zero user intervention is the big challenge.

References

1) Oliva, A. et al.: Hybrid Images, *ACM Transactions on Graphics*, Proc. of ACM SIGGRAPH conf., Vol.25, No.3,pp.527–530 (2006).
2) Schyns, P. and Oliva, A.:From blobs to boundary edges: Evidence for time- and spatial-scale-dependen scene recognition, *Psychological Science*,Vol. 5,pp. 195–200 (1994).
3) Oliva, A. and Schyns, P.: Coarse blobs or fine edges? evidences that information diagnosticity changes the perception of complex visual stimuli, *Cognitive psychology*,Vol.34, No.1,pp.72–107 (1997).
4) Schyns, P. and Oliva, A.: Dr.angry and mr.smile: when categorization flexibly modifies the perception of faces in rapid visual presentations, *Cognition*,Vol.69,pp. 243–265 (1999).
5) Navon, D.: Forest before trees: the precedence of global features in visual perception, *Cognitive Psychology*,Vol.9 ,pp.353–383 (1977).
6) Hughes, H.C. et al.: Global precedence, spatial frequency channels, and the statistics of natural images, *Journal of Cognitive Neuroscience*,Vol.8, pp.197–230 (1996).
7) Lindeberg, T.: Detective salient blob-like images structures and their spatial scales with a scale-space primal sketch: a method for focus of attention, *Int. J. Comp. Vis* ,Vol.11, pp.283–318 (1993).
8) Parker, D. et al.: Temporal integration of spatially filtered visual images, *Perception*, Vol.21, pp.147–160 (1992).
9) Sugase, Y. et al.: Global and fine information coded by single neurons in the temporal visual cortex, *Nature*, Vol.400, pp.869–873 (1999).
10) Konishi, M. and Yamaguchi, Y.: Hybrid Images by Local Frequency Analysis , *Proceedings of Visual Computing, Graphics and CAD Symposium*, DVD, Information Processing Society of Japan, 36:1-36:6 (1990). (in Japanese).
11) Tomasi, C., and Manduchi, R.: Bilateral filtering for gray and color images, *Proc. IEEE Int. Conf. on Comp. Vision.*, (1998).
12) Fattal, R. et al.: Gradient domain high dynamic range compression, *ACM Trans. on Graphics*, Proc. of ACM SIGGRAPH conf., Vol.21, No.3 (1998).
13) Bae, S. et al.: Two-scale Tone Management for Photographic Look, *ACM Trans. on Graphics*, Proc. of ACM SIGGRAPH conf., Vol.25, No.3 (2006).
14) Gonzales, R. C.: *Digital Image Processing*, Prentice Hall (1990).
15) Durand, F. and Dorsey, J.: Fast bilateral filtering for the display of high-dynamic-range images, *ACM Trans. on Graphics*, Proc. of ACM SIGGRAPH conf., Vol.21, No.3 (2002).

Appendix

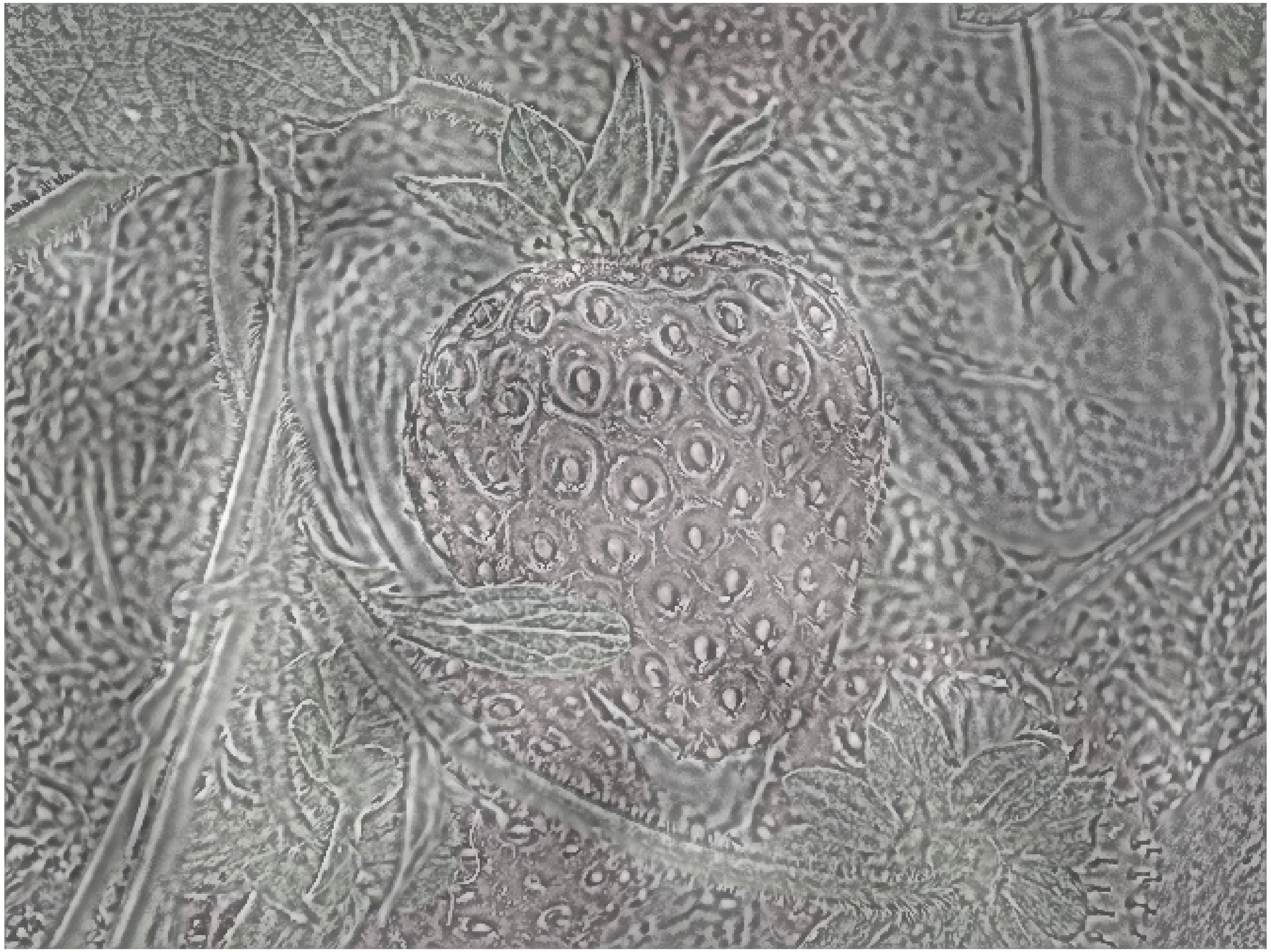

Fig. 8 Hybrid image from our method. When seen up-close (around 40 cm from image), we can see a strawberry with its leaves and the shadow in the background is percept as meaningless blobs. Step a few metres away (around 4 m from image), a face of tiger becomes visible.