Hybrid images

Aude Oliva∗
MIT-BCS

Antonio Torralba†
MIT-CSAIL

Philippe. G. Schyns‡
University of Glasgow

Figure 1: A hybrid image is a picture that combines the low-spatial frequencies of one picture with the high spatial frequencies of another picture producing an image with an interpretation that changes with viewing distance. In this example, you might think that these people are sad. But ask someone to tell you what the expressions are while she sees the picture from a few meters away.

Abstract

We present Hybrid images, a technique that allows creating static images with two interpretations that change as a function of viewing distance. Hybrid images are based on the multiscale processing of images by the human visual system and are motivated by masking studies in visual perception. These images can be used to create compelling displays in which the image appears to change as the viewing distance changes. We show that by taking into account perceptual grouping mechanisms it is possible to build compelling Hybrid images with stable percepts at each distance. We show examples in which Hybrid images are used to create textures that only become visible when seen up-close, to create facial expressions whose interpretation changes with viewing distance, and to visualize changes over time within a single picture.

Keywords: Hybrid images, human perception, scale space

1 Introduction

Here we exploit 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. We use and extend the method originally proposed by Schyns and Oliva [1994; 1997; 1999]. Fig. 1 shows an example of a Hybrid image build from two images in which the faces display two different emotions. The high spatial frequencies correspond to faces with sad expressions. The low spatial frequencies correspond to the same faces but with happy and surprise emotions (the emotions are, from left to right: happy, surprise, happy and happy). To switch from one interpretation to the other you can step back from the picture a few meters.

Artists have effectively employed low spatial frequency manipulation to elicit a percept that changes when relying on peripheral vision (e.g., [Livingstone 2000; Dali 1996]). Inspired by this work, Setlur and Gooch [2004] propose a technique that creates facial images with conflicting emotional states at different spatial frequencies. The images produce subtle expression variations with gaze changes. In this paper, we demonstrate the effectiveness of Hybrid images to create images with two very different interpretations.

Hybrid images are generated by superimposing two different 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. Note that Hybrid images are a different technique than picture mosaics [Silvers 1997]. Picture mosaics have two interpretations: a local one (which is given by the content of each of the pictures that compose the mosaic) and a global one (which is best seen at some pre-defined distance). On the other hand, Hybrid images contain two coherent global image interpretations, one of them is in the low spatial frequencies and the other in the high spatial frequencies.

We illustrate this technique with several proof-of-concept examples. We show how this technique can be applied to create face pictures that change expression with viewing distance, to display in a single picture two configurations of a scene, and to create textures that disappear when viewed at a distance.

2 The design of 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 · G1 + I2 · (1 − G2), the operations are defined in the Fourier domain. Hybrid images
When we describe the real image resolution observed when the image has a global structure and the spatial relationships between components is based on a multi-scale, global to local analysis of the visual input scene edits in an action movie or in a music video. Research in hu-glance (100 msec [Potter 1975]). This phenomenal performance of Visual psychophysics research has shown that human observers are work for understanding the mechanisms involved in the perception of these double image percepts.

2.1 The perception of Hybrid images

We describe in this section the motivation behind Hybrid images as they relate to studies in human perception. This provides the framework for understanding the mechanisms involved in the perception of these double image percepts.

Visual psychophysics research has shown that human observers are able to comprehend the meaning of a novel image within a short glance (100 msec [Potter 1975]). 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 guide the analysis of local details [Schyns and Oliva 1994; Watt 1987]. The global precedence hypothesis of image analysis (“seeing the forest before the trees”, [Navon 1977]) potentially implies a coarse to fine frequency analysis of the image, where the low spatial frequency components, which are very contrasted and carried by the fast magnocellular pathway, dominate early visual processing [Hughes et al. 1996; Lindeberg 1993; Parker et al. 1992; Schyns and Oliva 1994; Sugase et al. 1999].

Using hybrid stimuli, Schyns and Oliva [1994] tested the role that different spatial frequency bands play for the interpretation of natural images. When the task consisted in identifying quickly a scene image, human observers interpreted low spatial frequency band (at a frequency cutoff of 8 cycles/image) before the high spatial frequency band (from 24 cycles/image): when showed hybrid images for 30 ms only, observers identified the low spatial scale (e.g., they would answer “cheetah” when presented with the image from Fig. 3) whereas for 150 ms duration, they identified the high spatial scale first (e.g., tiger in Fig. 3). Interestingly, participants were unaware that the visual stimuli had two interpretations. Additional experiments suggested that the spatial frequency band preferentially selected for interpreting an image depends on the task the viewer must solve. Using hybrid faces similar to the one in Fig. 5.b, Schyns and Oliva [1999] showed that when participants were asked to determine the emotion of an hybrid face image displayed for only 50 ms (happy, angry or neutral), they selected the low spatial frequency band (angry in Fig. 5.b), but when they had to determine the gender of the same image, they equally used the low or high spatial frequency components of the hybrid. Again, participants did not notice that the images had two emotions or two genders. These results demonstrated that the selection of frequency bands for fast image recognition is a flexible mechanism: the image analysis might still unfold according to a low to high spatial scale processing, but human observers are able to quickly select the frequency band, low or high, that conveyed the most information to solve a given task and interpret the image. Importantly, when selecting a spatial frequency, observers were not conscious of the information in the other spatial scale.

From the point of view of the study of human perception, Hybrid images allow characterizing the role of different frequency channels for image recognition, and evaluate the time course of spatial frequency processing. From the view point of computer graphics, Hybrid images are a new paradigm to create images whose interpretation can be modulated by playing with viewing distance or presentation time. For a given distance of viewing, or a given temporal

---

1 We use the units cycle/image for spatial frequencies as they are independent of the image resolution. The output of a gaussian filter with a cutoff frequency of 16 cycles/image will be the same independently of the resolution of the original image. The units cycle/degree of visual angle are used when we describe the real image resolution observed when the image has a fixed size and it is seen from a fixed distance.
Figure 3: Perceptual grouping between edges and blobs. The three images are perceived as a tiger when seen up-close and as a cheetah from far away. The difference between the three cases is the degree of alignment between the edges and blobs. a) Two images superimposed without alignment. b) Eyes are aligned. c) Head pose and the locations of eyes and mouth are aligned. Under proper alignment, the residual frequency band does not manage to build a percept. When seen up-close, it is difficult to see the cheetah’s face which is perfectly masked by the tiger’s face. From far away, the tiger’s edges are assimilated to the cheetah’s face.

Figure 4: Color at high spatial frequencies is used to enhance the bicycle when seeing up-close. From a distance, one can see a motorcycle. The shape of the motorcycle, is interpreted as shadows when seeing up-close.

frequency a particular band of spatial frequency dominates visual processing. Visual analysis of the hybrid image still unfolds from global to local perception, but within the selected frequency band: for a given viewing distance, the observer will perceive the global structure of the hybrid first (that the image in Fig. 3 represents a head), and take an additional hundred milliseconds to organize the local information into a coherent percept (organization of blobs if the image is viewed at a far distance, or organization of edges for close up viewing).

2.2 Rules of perceptual grouping and Hybrid images

In theory, one can combine any two images to create a hybrid picture. In practice, aesthetically pleasing Hybrid images require following some rules that we describe in this section. Successful Hybrid images are the ones that have a perfect blending of edges in both scales. In such cases, when one percept dominates it becomes almost impossible to consciously switch to the alternative interpretation. 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 looking at a short distance. By introducing accidental alignments it is possible to reduce the influence of one spatial channel over the other. For instance, in Fig. 2 the top of the elephant (low spatial frequencies) is aligned with the horizon line (both low and high spatial frequencies). Therefore, when seeing the image up close, the top edge of the elephant can be explained by some of the fine edges. This reduces the saliency of the elephant. Fig. 3 shows several examples of Hybrid images with different degrees of agreement between the low and high spatial frequencies.

Color provides a very strong grouping cue that can be used to create more compelling illusions. For instance, in Fig. 4 color is used only in the high spatial frequencies to enhance the bicycle and to reinforce the interpretation of the motorbike as shadows when the image is viewed up close.

Fig. 5 illustrates the importance of correctly choosing the cut-off frequencies for the filters. In Fig. 5.a, both filters have a strong overlap. As a consequence there is not a clean transition between the two faces. For the Hybrid image on Fig. 5.b, the two filters
Figure 5: An angry man or a thoughtful woman? Both hybrid images are produced by combining the faces of an angry man (low spatial frequencies) and a stern woman (high spatial frequencies). You can switch the percept by watching the picture from a few meters. a) Bad hybrid image. The image looks ambiguous from up close due to the filter overlap. b) Good Hybrid image.

Figure 6: Correlations across levels of a Laplacian pyramid for images following several manipulations. a) Natural image, b) two images added, c) blurry image with additive white noise, and d) hybrid image ($f_1 = 16 \text{ cycles/image}, f_2 = 48 \text{ cycles/image}$).

In summary, there are two main mechanisms that can be exploited to create compelling Hybrid images. The first is to maximize the correlation between edges in the two scales so that they blend. The second resides in the fact that the remaining edges that do not correlate with other edges across scales can be perceived as noise. This is the case in Fig. 5.b for which there is a very compelling blending of edges across scales, but, when seeing the image up close, there seems to be some low-spatial frequency noise.

2.3 Capacity of scale space

Up to now, hybrid images have been obtained by mixing two images. But would it be possible to combine more than two images and still have a coherent percept that transitions as we change viewing distance? In a study about text masking, Majaj et al. [2002] created a stimulus superimposing 4 letters with each containing energy at different spatial scales. As the observer moves away from the stimuli, they report the image switching from one letter to another. The results are interesting but the lack of good grouping cues between the multiple scales creates an image that looks distorted. Also, multiple letters are visible at any given time. Superposition of multiple images remains an open issue.

3 Applications

In this section we discuss some applications (see video complementing the paper for additional examples).
Private font: We can use the hybrid images to display text that is not visible for people standing at some distance from the screen. Commercial products for user privacy generally rely on head mounted displays or on polarized screens for which visibility decreases with viewing angle. Hybrid fonts comprises two components: the high spatial frequencies (which will contain the text) and the low spatial frequencies (which will contain the masking image).

For the high pass filter we use a gaussian filter with a width (\( \sigma \)) adjusted so that \( \sigma < n_p \), where \( n_p \) is the thickness of a letter’s stroke measured in pixels. The low-frequency channel (masking signal) contains a text-like texture [Portilla and Simoncelli 2000]. Solomon and Pelli [1994] have shown that letters are more effectively masked by a noise in the frequency band of 3 cycles per letter. Therefore we adjust the cut-off frequency of the low-pass filter to be \( 3 \times n \) with \( n \) being the number of letters in a text line. The goal is to reduce the interference of the noise with the text when we look up close while having an effective masking noise when looking from farther away.

In the example shown in Fig. 8 the text is only readable from a distance bellow one meter. From a distance of about two meters, the text becomes unreadable. Masking of the low spatial frequencies is very important to get this effect (Fig. 8). The text in the bottom has only been high-pass filtered and there is no masking at low spatial frequencies, therefore it remains easy to read at relatively long distances.

Hybrid textures: We can create textures that disappear with viewing distance. An example of this idea is shown in Fig. 9. This figure shows an example of a woman’s face that turns into a cat when looking close. Note that this effect can not be obtained by superimposing the woman’s face and the cat’s face using transparency. Using transparency (additive superposition) creates a face that will not change with distance.

Changing faces: Hybrid images are especially powerful to create images of faces that change expressions, identity, or pose as we vary the viewing distance. Fig. 1 shows a compelling example of changes of face expression. The edges at multiple scales blend producing images that look very natural at all distances. In the case of face images, correct alignment between facial features is very important in order to create pictures that seem unaltered. In case of misalignment, the best is to apply a distortion (affine warping) to the face that will be in the low spatial frequencies.

Time changes: Fig. 9 shows an example of using an hybrid image to show two states of a house by combining two picture taken at two different instants.

4 Conclusion

We have described the technique of Hybrid images that 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 provide also an interesting new visualization tool to present 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.

References


