Hybrid Image Illusion

Aude Oliva (1) and Philippe G. Schyns (2)

(1) Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology

(2) Institute of Neuroscience and Psychology
University of Glasgow

Address:
Aude Oliva
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
77 Massachusetts avenue
Cambridge, MA 02139
Email: oliva@mit.edu

Philippe G. Schyns
Institute of Neuroscience and Psychology
University of Glasgow
58, Hillhead Street
Glasgow G128QB
United Kingdom
Email: Philippe.Schyns@glasgow.ac.uk

Word count: (with figure captions) ~ 1800
Number of figures: 3

In The Oxford Compendium of Visual Illusions, Ed. Shapiro and Todorovic, Oxford University Press. Chapter 111 (p 763-766).

Abstract

Artists, designers, photographers, and visual scientists are routinely looking for ways to create, out of a single image, the feeling that there is more to see than what meets the eyes. Many well-known visual illusions are dual in nature, causing the viewer to experience two different interpretations of the same image. Hybrid images illustrate a double image illusion, where different images are perceived depending on viewing distance, viewing duration or image size: one that appears when the image is viewed up-close (displaying high spatial frequencies), and another that appears from afar (showing low spatial frequencies). This method can be used to create compelling dual images in which the observer experiences different percepts when interacting with the image.

Keywords: hybrid image, illusion, percept, spatial frequency, image size, vision

The phenomenon: Perceiving two for the sight of one

A hybrid image is a single picture that combines the low spatial frequencies (LSF) of one image with the high spatial frequencies (HSF) of another image, producing a new image with an interpretation that changes with duration, viewing distance or size (Brady and Oliva, 2012; Oliva and Schyns, 1997; Oliva, Torralba and Schyns, 2007; Schyns and Oliva, 1994, 1999). Figure 1 shows the popularized Marilyn Monroe – Albert Einstein hybrid image, created in 2007 for *New Scientist* magazine. Very close up, you will perceive Albert Einstein’s portrait, with some seemingly randomly cast shadows. If you step back, squint your eyes, or reduce the size of the hybrid (such as by making the image smaller on your screen), you will gradually start seeing Marilyn Monroe’s face: at first you may perceive a blend of the two portraits (such as Marilyn with a mustache!), but if you go far enough, or look at a tiny image of the hybrid you should only see Marilyn Monroe, with Einstein's face completely gone from perception. The same principle applies in the real world: If you are standing far away from the object you are looking at, then all you can see are the low spatial frequencies. For example, if you are looking at the face of someone standing a few tens of meters away from you, you can tell if the face is male or female, but you might not be able to clearly distinguish the facial emotion or the age of the person (for a review, Sowden and Schyns, 2006).

Our visual system decomposes the information of an image into components termed spatial frequencies (from low to high, Robson, 1966). Figure 1 shows image thumbnails of the spatial frequency bands of the hybrid: the lowest spatial frequencies only depict the shape of Monroe’s head and bust. Low spatial frequencies are global luminance variations in the image and convey broad contours: they are literally a blurry, out of focus version of the original image. When adding medium spatial frequencies, Monroe’s eyes, mouth, nose, ears, and hairline start becoming clear. In contrast to the broad contours found in low spatial frequencies, the high spatial frequencies of an image represent sharp details and fine contours, such as the wrinkles of a face or the texture of a surface. The high spatial frequency bands of an image resemble a highly detailed line drawing. In

Figure 1, we can discern Albert Einstein's portrait in the right hand thumbnails: at these high spatial frequencies you can see the details of his moustache, the wisps of his hair and the sharp lines around his tie and collar.

Superimposing the low and the high spatial frequency components of two different images on each other creates *hybrid images*. In its simplest version, a hybrid overlaps one coarse, blurry picture with one fine, detailed picture. In practice, image-processing software is used to filter one image with a low-pass filter and the second image with a high-pass filter (Oliva, Torralba and Schyns, 2006).

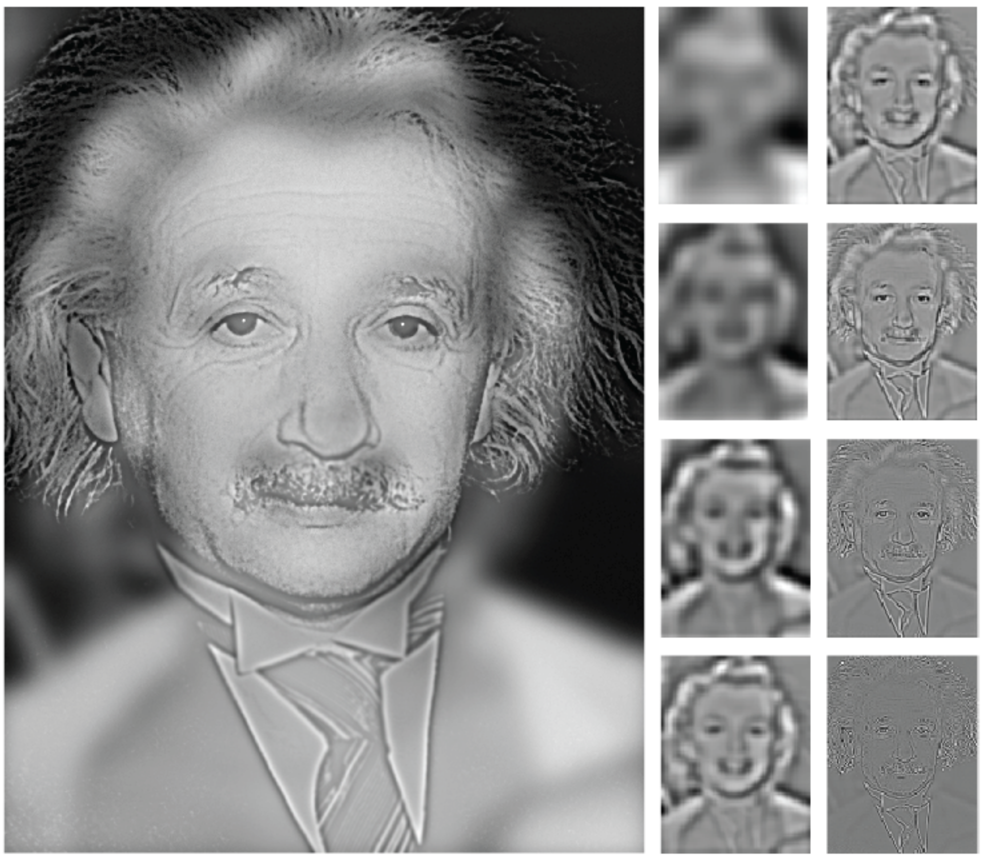

Fig. 1: Albert Einstein - Marilyn Monroe hybrid image. The thumbnailed images illustrate the band-pass filtered components of the hybrid, from Monroe (the blurred, or the low spatial frequency version) to Einstein (with sharp contours, or the high spatial frequency version).

The psychophysics of hybrid images

In the real world, all the spatial frequency bands of an image correspond to a single and coherent percept. Hybrid images change this rule, leading the visual system to switch from one percept to another. Hybrid illusions have been used in various behavioral and psychophysical studies to show that the order in which frequency bands are used in visual perception depends on the task, and it is not necessarily driven by a fixed schedule (e.g.

first mandatory perception of the Low Spatial Frequencies then perception of the High Spatial Frequencies).

Who do you see: Dr Angry or Ms Smile?

When we categorize a particular face as old or young, our visual system automatically knows the scale at which to find useful information. For example, an older face will tend to possess many fine-grained wrinkles that are typically represented in the high spatial frequencies of the image (see the wrinkles of Einstein in Figure 1). Which spatial frequencies the visual system uses for different face categorizations (e.g. identity or facial expressions) is thus an important question. In Schyns and Oliva (1999), we presented hybrid stimuli overlapping, for instance, an expressive male in LSF and a neutral female in HSF (and vice versa for the assignment of gender and expression to LSF and HSF). We briefly presented each hybrid stimulus on the screen (for 50 msec) and asked observers to categorize its gender, its identity, and whether it was expressive or not. We found that upon briefly seeing the same hybrid stimulus (e.g. a neutral female in LSF overlapped with an angry male in HSF, Figure 2-A) observers perceived an expressive face (on the basis of HSF) vs. "Mary" (on the basis of LSF), depending on the categorization task they were instructed to perform (i.e. "expressive or not?" or "which identity?"). Thus, the observers' brains used different spatial frequency information from the same hybrid image to categorize it in different ways, resulting in mutually exclusive perceptions of the same visual input.

Where do you look to see Local Jekyll or Global Hyde?

Another important question in vision concerns the local or global nature of the information responsible for everyday recognition. In typical viewing, information useful to identify a face (e.g. the blue color of Paul Newman's eyes) could simply vanish with increasing distance due to the physics of retinal image formation. Identification mechanisms invariant to viewing distances could flexibly adjust to use both local and global cues from the same visual input, depending on which are available on the retina. Here we created an "iHybrid" to study how the visual system uses local and global face identification strategies. An iHybrid represents one identity in LSF and another one in HSF (in Figure 2-B, Brad Pitt and William Macy, respectively). Importantly, the eye fixations of the observer *dynamically* determine the spectral composition of the image (see Figure 2-B): An eye-tracker records in real time the location of the eyes and a computer updates in real-time the visual display as a function of the observer's eye position to achieve the following effect: The observer sees local, full spectrum information from Brad Pitt, in their fovea (i.e., where their eyes fixate, indicated in the Figure with a dashed line) and global, complementary information from William Macy outside their parafovea. In Miellet, Caldara, and Schyns (2011), we showed that the visual system indeed identified faces with local and global cues. Consequently, on one trial a given hybrid could be perceived as "Brad Pitt" by integrating local information over several fixations, and on another trial as "William H. Macy," on the basis of the complementary global information accrued over fewer fixations.

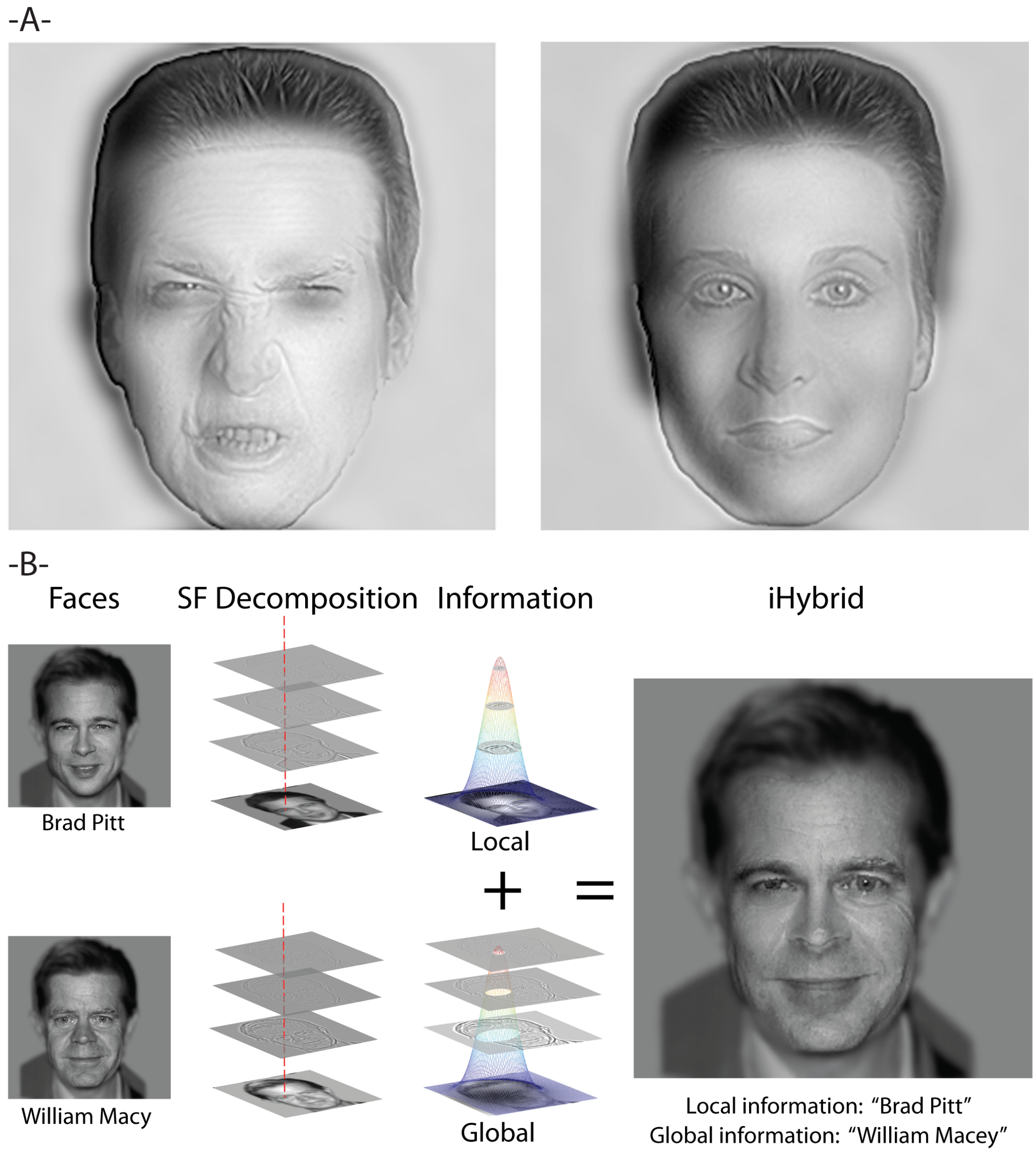

Fig. 2: -A- Two hybrid faces used in Schyns and Oliva (1999) combining each a male and a female, with a neutral or expressive face, either in LSF or HSF -B- Illustration of the iHybrid method of Miellet et al. (2011). An iHybrid simultaneously and dynamically represents two identities across local (here, Brad Pitt) and global (here, William Macy) spectral information, as a function of observer fixation. The local information at fixation (represented with a dotted line) represents a local cut across the SF spectrum of the first identity whereas the global information represents the complementary SF information from the other identity. Decoupling of local and global information in iHybrids enables understanding of the local and global information used for face identification.

The asymmetry of spatial frequency integration: seeing more than meets the eye

Perception is inherently a dynamic process: as we move through the world, the percept of an object we are looking at constantly changes in spatial frequency content (Bonnar, Gosselin, and Schyns, 2002; Pelli, 1999; Sowden and Schyns, 2006), creating an asymmetry in the availability of visual information about the objects we are looking at. When we approach an object or when an object comes towards us, our perceptual system is constantly gaining new spatial frequency information about it, adding details to the on-line representation of the object. But when an object is receding from us, our perceptual system loses information: the object becomes more ambiguous as its details are too high in spatial frequency to be perceived. In Brady and Oliva (2012), using hybrid images, we studied how the visual system deals with this asymmetry when integrating spatial frequency information from objects that approach or recede from the observer. Observers judged how similar a hybrid face or hybrid scene was to each of its original images while physically walking toward or away from it or having the stimulus virtually moved toward or away from them on a computer screen. Interestingly, when the observer or the stimulus are approaching each other, observers perceive the stimulus as if the image is simply gaining higher spatial frequency components as they become physically available to the eye. However, when the stimulus or the observer are receding from each other, observers show a *perceptual hysteresis* effect, holding onto fine details that are supposed to be imperceptible at that distance. They can see more high spatial frequency information than meets the eyes. This suggests that people naturally make optimal inferences when perceiving objects in the real world, by sticking with their previous interpretation when they lose information and constantly reinterpreting their input when gaining new information.

The art of hybrid images

Compelling hybrid images can be created with images other than faces: if the mind can reorganize from a distance the shapes of an image into a coherent form, switching between the percepts of two different stimuli can occur. For instance, forms seen as shadows close up will often work to the hybrids' advantage (Oliva et al., 2006). Figure 3 illustrates such an illusion: close up, we see a bicycle with shadows cast on the wall behind. The scene has a three-dimensional flavor to it. But if you step back from the image, what appear to be shadows will regroup to form the body of a motorcycle. At close perceptual range, the parts that made up the motorcycle are reinterpreted as the shadows of the bicycle, localized behind the attended object. This grouping allows for a seamless transformation from one object to another entirely different object.

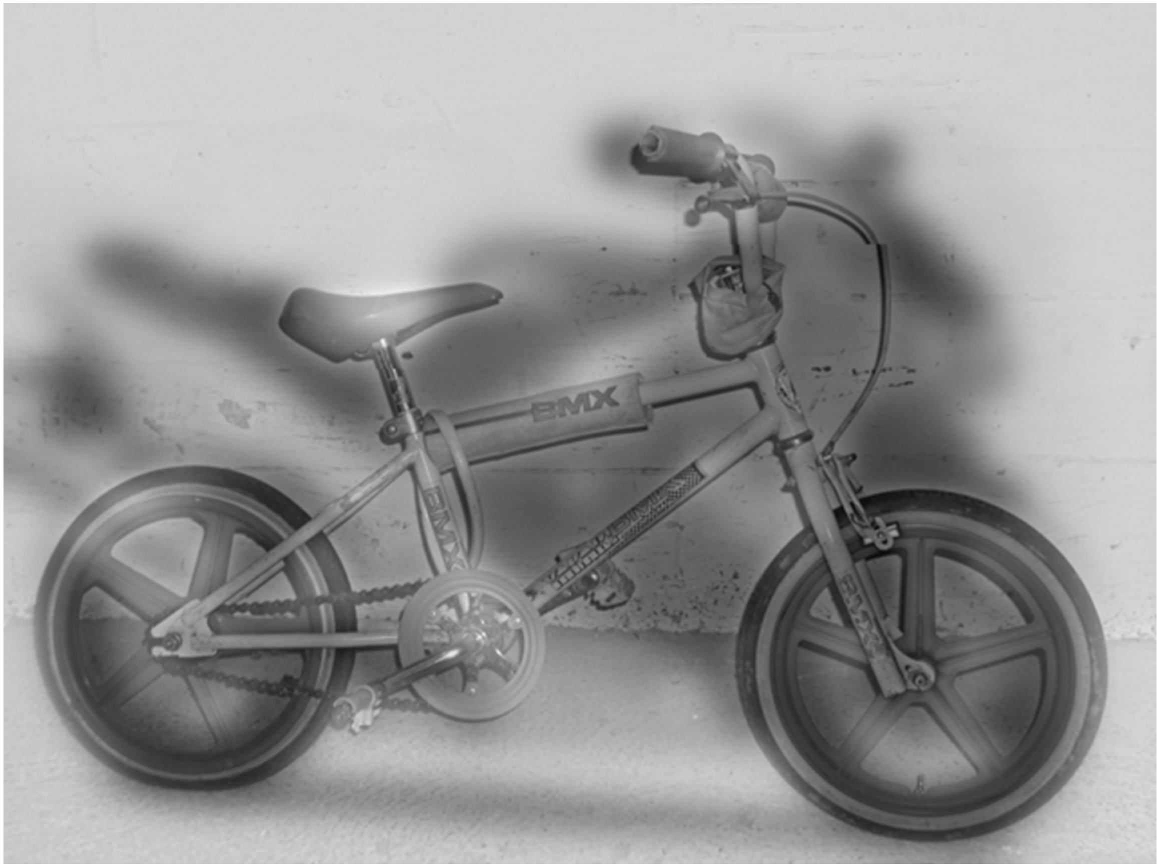

Fig. 3: A hybrid image made of a bicycle (seen up-close) and a motorcycle (seen from a distance). Parts of the body of the motorcycle are interpreted as cast shadows up-close, projected on the wall behind the bicycle (from Oliva et al., 2006).

Acknowledgements

The writing of this article was in part supported by a National Eye Institute grant EY020484 to A.O.

References

Bonnar, L., Gosselin, F., and Schyns, P.G. (2002). Understanding Dali's *Slave Market with the Disappearing Bust of Voltaire*: A case study in the scale information driving perception. *Perception* 31(6): 683-691.

Brady, T.F., and Oliva, A. (2012). Spatial scale integration during active perception: Perceptual hysteresis when an object recedes. *Frontiers in Perception Science* 3: 462.

Miellet, S., Caldara, R., and Schyns, P.G. (2011). Local Jekill and Global Hide: The Dual Identity of Face Identification. *Psychological Science* 22(12): 1518-1526.

Oliva, A., and Schyns, P.G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology* 34: 72-107.

Oliva, A., Torralba, A., and Schyns, P.G. (2006). Hybrid Images. *ACM Transactions on Graphics (SIGGRAPH)* 25(3): 527-532.

Pelli, D. G. (1999). Close encounters--An artist shows that size affects shape. *Science* 285: 844-846.

Robson, J.G. (1966). Spatial and temporal contrast sensitivity functions of the visual system. *Journal of the Optical Society of America* 56: 1141-1142.

Schyns, P.G., and Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science* 5: 195-200.

Schyns, P.G., and Oliva, A. (1999). Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition* 69: 243-265.

Sowden, P.T., and Schyns, P.G. (2006) Channel surfing in the visual brain. *Trends in Cognitive Sciences* 10: 538-545.