The Art of Hybrid Images: Two for the View of One

Aude Oliva [*]

Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, MIT 32-D432, 77 Massachusetts Avenue, Cambridge, MA, 02139, USA

Abstract

Hybrid images are static pictures with two interpretations that change depending on the image's viewing distance or size. The phenomenon of hybrid images arises from the multiscale processing of images in the human visual system. By taking into account perceptual grouping mechanisms, one can build compelling hybrid images with two different stable interpretations: one that appears when the image is viewed up-close, and the other that appears from afar. Hybrid images can be used to create compelling prints and photographs in which the observer experiences different percepts when interacting with the image.

1. Introduction

Artists, designers, and visual scientists have long been searching for ways to make multiples meanings out of a single image. The *Mona Lisa*'s smile seems both present and absent at the same time (see Livingstone, 2000, for a scientific article on the painting). *Salvador Dali*'s 'double image' paintings contain hidden figures and meanings. Many well-known visual illusions cause the viewer to experience different interpretations of the exact same image, such as 'Rubin's vase' which can be viewed as a pair of faces. This article presents examples of one way to make multiple meanings out of a single image: namely, double and hybrid images, which are images that look different depending on viewing distance or size.

---

[*] E-mail: oliva@mit.edu

1.1. Duality in Seeing: The Art of Double Images

Figure 1 illustrates three famous pieces of 'double image' artwork. Each piece can be interpreted differently depending on the distance of viewing, or the size of the picture (e.g., a large image *versus* a thumbnail). In 1892, Charles Allan Gilbert drew an impressive piece named *All is Vanity* to illustrate the fate of life. From close up, the drawing depicts a young woman looking at her reflection in a boudoir mirror. From a distance, however, the drawing becomes a skull. This dual concept was used to advertise Dior's advertisement for the 'Poison' fragrance in 2007.

The most effective contemporary 'double image' painting may be Salvador Dali's 1940 *Slave Market with the Disappearing Bust of Voltaire*. In this painting, what looks like Voltaire's head from afar is actually a group of characters and an arch opening in the wall. When viewed from a middle distance, the viewer's perception seems to switch back and forth naturally between seeing the overall picture and the details: one second the viewer sees the 'global' image of Voltaire's face, and the next second the viewer sees the 'local' image of two nuns at a market.

Using a method from the study of psychophysics (*Bubbles*, Gosselin and Schyns, 2001), Bonnar *et al.* (2002) have shown that the two different perceptions are actually captured in different parts of the visual information that reaches our eyes (specifically, different 'local spatial filters'). The principle of capturing different information in the global and local aspects of an image has been used successfully in advertising campaigns by photographer Ian Butterworth. In his creative pieces, the viewer decides to assign form and shape to a figure or a ground, like a butterfly composed by two shoes or Walkman mp3 players interpreted as footsteps on the ground.
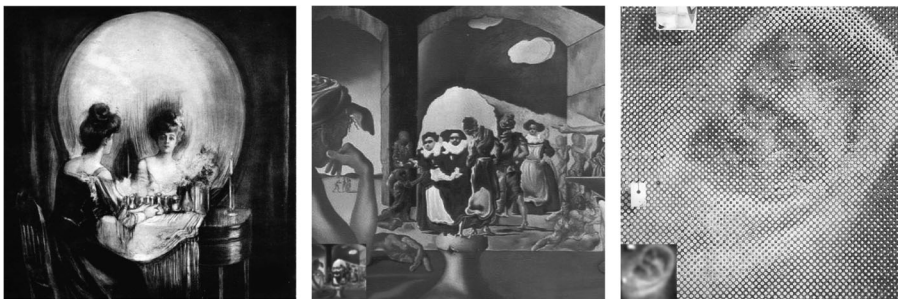

Figure 1. Illustration of double images: (left) A section of Charles Allan Gilbert's drawing, *All is Vanity*, 1892; (middle) A grey-scale version of Salvador Dali's painting entitled *Slave Market with the Disappearing Bust of Voltaire*, 1940; (right) A grey-scale version of Salvador Dali's oil canvas *Madonna*, 1958. For each painting, a low-resolution image is shown on the bottom left corner, simulating perception of the painting at a distance.

A third double image type — illustrated here by a grayscale excerpt from *Madonna* (Salvador Dali, 1958) — is based on tiny dots of different colors and brightness. Different shapes emerge from these dots depending on the viewer's distance from the image. In this example, within the global shape of a large ear one can perceive a Madonna and a child forming the ear's ridged interior surface.

The technique of manipulating the size of elements to affect shape perception is prominent in the artwork of Chuck Close, whose huge portraits collapse into dots when viewed from up close (see Pelli, 1999, for a scientific account of the phenomenon). This phenomenon is also used in the construction of 'picture mosaics' (see Silvers, 1997).

1.2. What is a Hybrid Image?

Hybrid images represent a fourth approach to making double images, directly from photographs or drawings. Hybrid images can be made from existing images using standard image manipulation software, such as Photoshop. The phenomenon and method were first published in *Psychological Science*, Schyns and Oliva (1994).

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 viewing distance or size (Brady and Oliva, 2012; Oliva and Schyns, 1997; Oliva *et al.*, 2006; Schyns and Oliva, 1994, 1999).

Every image is made up of different components at different spatial frequencies, which capture different aspects of that image. Low spatial frequencies (LSF) depict global luminance variations in the image and broad contours. When viewed alone, the low spatial frequencies of an image are literally a blurry, out of focus version of that image. LSF images reveal the outer shape of objects; at the lowest spatial frequencies only very large objects can be seen, with increasing spatial frequencies depicting more detail. For example, see the images of Marilyn Monroe in the left column of thumbnails in Fig. 2: at the lowest spatial frequencies you can only see the outer shape of her head and bust, with increasing spatial frequencies enabling you to see her eyes and mouth, and then her nose, ears and hairline. If you are standing far away from the thing 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, see Brady and Oliva, 2012; Sowden and Schyns, 2006).

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
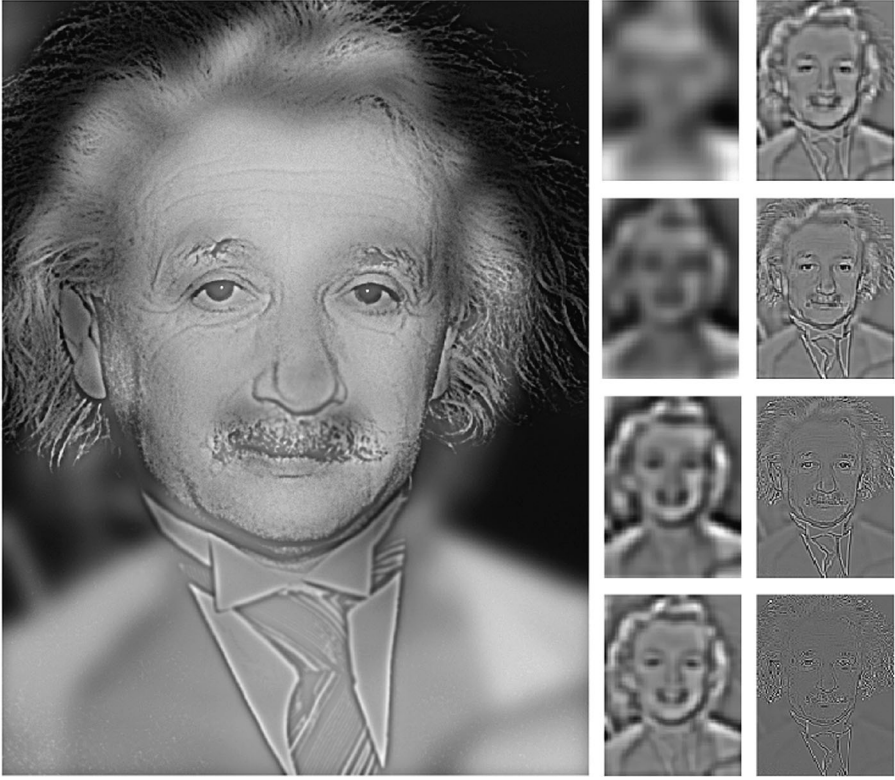
Figure 2. A hybrid image made from the high spatial frequency components of a picture of Albert Einstein and the low spatial frequency components of a picture of Marilyn Monroe. From close up, Albert Einstein is seen. At a greater distance or smaller size, Marilyn Monroe is seen. The thumbnails images show the band-pass filtered spatial frequency components of the hybrid, evolving from Monroe in the low spatial frequencies (blurry thumbnails starting in the top right), to Einstein in the high spatial frequencies (sharp thumbnails ending in the lower left). The Marilyn/Einstein hybrid image was created originally for the March 31st 2007 issue of *New Scientist* magazine.

as the wrinkles of a face or the texture of a surface. When viewed alone, the high spatial frequencies of an image resemble a highly detailed line drawing. For example, see the images of Albert Einstein in the right column of thumbnails in Fig. 2: at these high spatial frequencies you can see the sharp lines around his tie and collar, the details of his moustache, and the wisps of his hair. The fine lines represented in the HSF component are the complement of the blurry and coarse shapes in the LSF component.

Superimposing the coarse and the fine parts of two different images creates hybrid images. In its simplest version, a hybrid can be made by overlapping 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 *et al.*, 2006). Figure 2 shows the popularized Marilyn Monroe–Albert Einstein hybrid image, created by Aude Oliva for the March 31st, 2007 issue of *New Scientist* magazine. Very close up, you only see Albert Einstein's facial features, expression and hair, and you cannot even perceive Monroe; on Albert's face, Monroe's features are dissolved to look like shadows. However, 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 Monroe's face: at first you may perceive a mesh of the two (such as Monroe with a moustache), but if you go far enough, or look at a tiny image of the hybrid you should only see Marilyn, with Einstein's face completely gone from perception.

Figure 3 illustrates two novel hybrids made with the same Albert Einstein used in Fig. 2. Nearby, both pictures look like Einstein, with different shadows being projected on his face. Now step back and squint your eyes until you can start making sense of the shadows on his picture into a different face. Who do you see on the left and on the right? The faces of two different famous characters (a male and a female) should emerge as your visual system let the contours go.
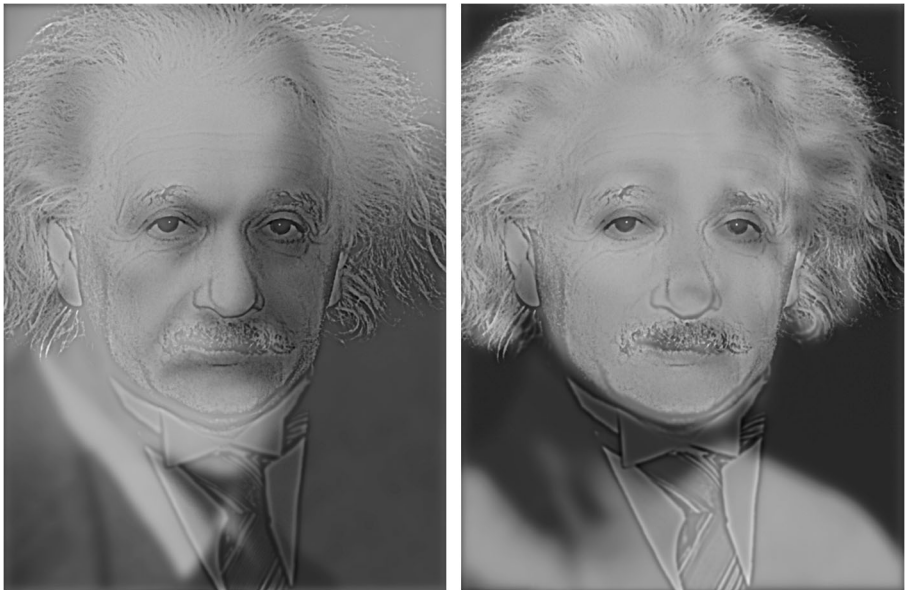

Figure 3. Two other versions of Einstein's hybrid: close-up, the high spatial frequencies represent Albert Einstein in both images, whereas at a distance, the low spatial frequency components feature two different characters. Can you figure out who they are? (See answers in Note 1).

2. Perceptual Principles Behind Hybrid Images

2.1. Role of the Contrast Sensitivity Function for Near and Far Viewing

It is well known that the human visual system sees the world through the contrast sensitivity function (Robson, 1966; for review, see Palmer, 1999), which acts as a band-pass filter, effectively making low-contrast information at particular spatial frequencies invisible to the naked eye. In fact, our visual system is on average most sensitive to contrast differences that occur at a particular range of spatial frequencies (roughly between 4 and 20 cycles/visual degree (see Note 2). If you look at the thumbnails band-pass images of Fig. 2 from a distance, the contours of Einstein in the right column images become all gray; you do not see the contours anymore.

Figure 4 (see Brady and Oliva, 2012) is a cartoon illustration of the type of information contained in the left hybrid image of Fig. 3. From any distance, the hybrid is a mixture of LSF and HSF components. The question is what shall we see if we filter the hybrid image with an average human contrast sensitivity function (CSF, depicted as a gray dashed line). Figure 4(B) shows the part of the hybrid seen, from a distance above 3 m: because the high spatial
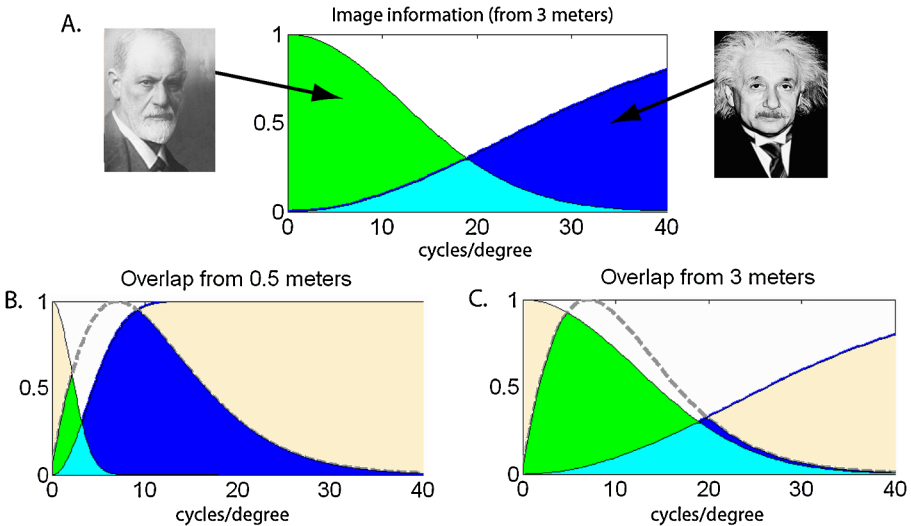

Figure 4. (A) The information contained in a hybrid image like Fig. 3 is a combination of the information originating from the HSF (in blue, Albert Einstein) and the LSF component of another image (in green, Sigmund Freud). (B) From a distance greater than 3 meters, the information visible to human observers (the information available from a standard contrast sensitivity function, depicted as the dashed-gray line) is predominantly driven by the low spatial frequency component: observers should preferentially perceive Freud. (C) From a distance inferior to 0.5 m, the high spatial frequency component dominates perception: observers should see Einstein. (Figure adapted from Brady and Oliva, 2012).

frequency image (blue) is too high frequency to be perceived, and is filtered away by the contrast sensitivity function, the observer will almost only see the low spatial frequency component. By contrast, panel C illustrates the same hybrid filtered by the contrast sensitivity function, but for a viewing distance below 0.5 m. Here, the hybrid information is shifted into almost only HSF, and so in addition to the LSF component information that exists in the image, there is now a large amount of information coming from the HSF component. Importantly, the LSF component of a hybrid is almost always available to the observer, whatever the distance of viewing, whereas the HSF component can only be seen when the observer is close to the image (given a fixed size). Very close, the LSF image can dissolve from perception, because the human mind interprets the detailed HSF image, which masks the still-visible but blurry image. By carefully choosing the amount of LSF and HSF in each image, the size of the image, the contrast and color components, and by applying the contrast sensitivity function to these images, we can predict what most observers will see from a given distance (Brady and Oliva, 2012).

2.2. Perceptual Grouping Effects

The most compelling hybrid stimuli are those where the shape and forms of the two original images are aligned, permitting the mind to switch interpretations by reorganizing the visual information. How to efficiently use this perceptual grouping phenomenon and perceptual organization principles (Kubovy and Pomerantz, 1981; Palmer, 1999) to create compelling hybrid images is more an art and an intuition than an algorithmic recipe, and the success depends on finding two images that work well together.

Figure 5 illustrates a compelling hybrid, based on alignment of features and the use of color. The center image is a hybrid 'kitten-puppy', flanked by the two original images used to create it. At standard reading distance, the center image should appear as a kitten, but if you squint your eyes and step back, the puppy face will appear. In this hybrid, two original images have been aligned beforehand so that the nose of the dog becomes a shaded patch on the kitten, its tongue overlaps with the kitten's nose, and the eyes of the two animals are superposed. Here, we kept the color of the puppy picture to enhance the perceived quality of the hybrid. Although there is a successful and definite switch that occurs with altered distance, this hybrid picture appears up-close as if it is a new kitten altogether, and it even looks like a normal photograph. The bottom row of Fig. 5 shows a set of band-pass images, the 'filtered images' that together form the hybrid, demonstrating that the LSF component only contains the puppy image, whereas the HSF component only contains the kitten image.

Compelling hybrid images can be created with images other than faces: if the mind can reorganize from a distance the forms and shapes into a coherent
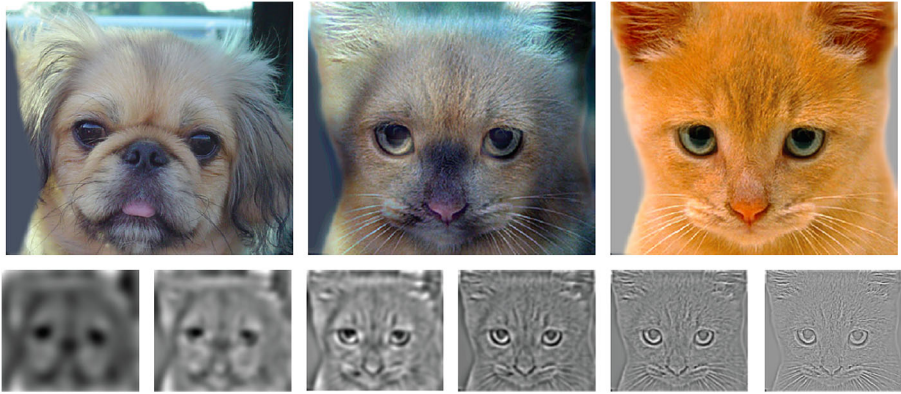
Figure 5. The top row illustrates a hybrid image in the center (kitten close up, to puppy further away), alongside the two original pictures used to make the hybrid. The bottom row shows the band-pass filtered versions of the hybrid, illustrating the information contained in the different spatial frequency bands from LSF (only the puppy face) to HSF (only the kitten face).

Figure 6. A hybrid image made of a bicycle (see up-close) and a motorcycle from the 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).

form, switching between the percepts of two different objects or scenes can occur.

Forms seen as shadows close up will often work to the hybrids' advantage (Oliva *et al.*, 2006). Figure 6 shows a bicycle, with shadows cast on the wall

behind. The scene has a three-dimensional flavor to it. Once the viewer steps back from the image, what appears to be cast shadows from the bicycle will regroup to form a motorcycle. At close perceptual range, the parts that made 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.

3. Conclusion

In principle, any two images can be combined to create a hybrid image. With careful attention to alignment and perceptual grouping mechanisms, one can build compelling hybrids with a stable and unique percept at near and far distances. Such images have many artistic and practical applications, including showing textures or text only visible up-close, displaying two different objects or events within the same image, generating facial expressions whose interpretation changes with viewing distance or image size, or visualizing changes over time within a single picture.

Notes

1. Spatial frequencies are measured as cycles/visual degree in a psychophysics experiment. One cycle represents one alternation of a black and a white component, two cycles are two alternations of black and white (like black, white, black, white), etc. If you look at your thumb at arm distance, your thumb covers about 2 degrees of visual angle. The peak of spatial frequencies you perceive the best is about 6 cycles per visual degree, which would be twelve stripes if you were to paint black stripes on your thumb. Higher spatial frequencies are harder to perceive, which is why some of the fine wrinkles on your finger cannot be seen very well any more at an arm distance.

2. The LSF images of the Einstein's hybrids are Sigmund Freud (left, hybrid made for the MIT Museum of Science, 2007, as part of the hybrid collection *Eight Einsteins*) and Madonna (right).

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**, 683–691.

Brady, T. F. and Oliva, A. (2012). Spatial frequency integration during active perception: perceptual hysteresis when an object recedes, *Front Percept. Sci.* **3**, 462, 1–8.

Gosselin, F. and Schyns, P. G. (2001). Bubbles: a new technique to reveal the use of visual information in recognition tasks, *Vision Research* **41**, 2261–2271.

Kubovy, M. and Pomerantz, J. R. (1981). *Perceptual Organization*. Lawrence Erlbaum, USA.

Livingston, M. S. (2000). Is it warm? Is it real? Or just low spatial frequency? *Science* **290**(5495), 1299.

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

Oliva, A., Torralba, A. and Schyns, P. G. (2006). Hybrid images, *ACM Trans. Graph. (SIGGRAPH)* **25**, 527–532.

Palmer, S. E. (1999). *Vision Science: Photons to Phenomenology*. MIT Press, Cambridge, MA, USA.

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, *J. Opt. Soc. Am.* **56**, 1141–1142.

Schyns, P. G. and Oliva, A. (1994). From blobs to boundary edges: evidence for time- and spatial-scale-dependent scene recognition, *Psychol. Sci.* **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.

Silvers, R. (1997). *Photomosacis*. Henry Holt and Co., New York, USA.

Sowden, P. T. and Schyns, P. G. (2006). Channel surfing in the visual brain, *Trends Cogn. Sci.* **10**, 538–545.