Far and Yet Close: Multiple Viewpoints for the Perfect Portrait

Pietro Perona *

Computation and Neural Systems Program, Division of Engineering and Applied Science, California Institute of Technology, 1200 E. California Blvd, Pasadena, CA 91125, USA

Received 31 January 2013; accepted 9 April 2013

Abstract
Realistic pictures, whether paintings, photographs or rendered images of models, are traditionally obtained using classical perspective projection. Here I focus on pictures of people and discuss evidence suggesting that, for best results, full-length realistic-looking portraits should be produced by combining multiple viewpoints. I test this prediction experimentally and find that full-length portraits look more compelling if appropriate multiple viewpoints are chosen. I notice that multi-viewpoint portraiture has likely been used since the Renaissance, although no explicit mention of it is found in the literature.

Keywords
Portrait, portraiture, perspective, joiner, photography, emotion, projection

1. Introduction
People are the most important component of our environment, and depictions of people are the most valuable images we can produce. Rendering effectively the human form, clothes, skin and hair, as well as pose, actions, and emotion, has been perhaps the most pressing concern of painters since the dawn of art. Addressing these challenges is also at the focus of much computer graphics research (Chang et al., 2002; Daldegan et al., 1993; Igarashi et al., 2007; Thalmann et al., 1996; Weil, 1986).

Renaissance artists realized that perspective projection onto a flat surface would produce an image that, when viewed from a privileged viewpoint, would be identical to the original scene viewed from the painter’s viewpoint.

* E-mail: perona@caltech.edu
Figure 1. A painter using a simple contraption to draw a model onto a flat surface using perspective (A. Dürer, c. 1525). The tip of the peg helps the painter maintain the same viewpoint (the center of projection). The grid helps obtain an undistorted mapping of the image onto the drawing surface. There is evidence that painters used optical devices from as early as 1450 (Hockney, 2001) to obtain faithful projections.

(Kubovy, 1986; Vishwanath et al., 2005). With this insight, the geometry of producing pictures of humans appears to be straightforward: if a painter, or a photographer, has access to a model posing on an appropriate stage, all that needs to be done is choose a viewpoint (the center of projection) and apply perspective (see Fig. 1). Similarly, one might think that, once computer graphics experts perfect modeling the form of the human body and rendering its materials, good pictures of synthetic people will be obtained by a single perspective projection.

I will argue, instead, that the geometry of portraiture, specifically of full-length portraiture, presents interesting challenges and opportunities. Conventional perspective is not sufficient if one wishes to obtain a ‘good’ portrait; rather, a number of different viewpoints should be combined.

2. What Makes a Good Portrait?

The advent of photography in the mid-1800s forced painters to re-think their role as picture makers (Hockney, 2001). Realistic pictures have since become the almost exclusive province of the photographer, and the vast majority of portraits are captured by a camera. Yet, some still argue that only the painter will be able to realize that ‘perfect’ portrait. Is there any reason, apart from tradition and fashion, in favor of painting a portrait?

The conventional explanation is ‘time’. Maurice Grosser puts it clearly and succinctly this way: “The painter, by observing the sitter over a certain length of time, is able to extract out of the flux of personality an average, and can present this average as a convincing enough resemblance. […] The still camera’s shutter […] does not represent nature, it immobilizes it” (Grosser, 1956). David Hockney makes a more general statement: “The experience of art is
more real in painting than in photography. The moment is longer and we can feel that moment. […] Perhaps this is why there are so few good photographs.” (Joyce, 2002). Is this a good argument? Although intuitively clear, this insight is not actionable. How is this ‘average’ computed, and how is it used in producing the end result? How can one translate the ‘flux of personality’ into a single picture?

Here I will explore instead a different line of explanation: ‘space’.

Grosser gives a refreshingly precise definition of ‘portrait’: “A portrait can be defined as a picture painted at a distance of four to eight feet of a person who is not paid to sit.” (Grosser, 1956). He motivates the choice of distance with the need to create a state of sympathy between the painter and the sitter.

I believe that this intuition may be studied quantitatively and objectively. Here is my hypothesis: distance changes the picture systematically, and this change affects the emotional valence of the picture. Thus, painters and photographers may produce portraits with different emotional effects by placing themselves at the appropriate distance from the sitter.

Consider the human head: it is a three-dimensional object. Two frontal perspective pictures of the head of the same individual will look different if one is taken from close and one from far (see Fig. 2). I am, of course, considering two pictures that have been scaled to the same size. There are three main effects: relative size of the features, occlusion and foreshortening. The nose

Figure 2. The same face viewed from 0.6 m (left) and from 2.4 m (right). The focal length of the lens was adjusted so that the head would fill the frame and the resolution held constant as in the ‘dolly zoom’ used in cinematography. The pictures were corrected for lens distortion to obtain exact perspective (Bouguet, 1999). Notice that the nose appears larger on the left. The ears and hair are relatively larger on the right.
juts out in front of the face and is 12–18 cm closer to the center of projection than the ears; therefore, the nose looks relatively larger in the picture when the camera is closer. Furthermore, when the head is seen from up-close, the ears are partially occluded by the cheeks, and the top of the head is not visible. Thus, an up-close picture of a head shows a larger nose and larger face within the head. Conversely, when the picture is taken from a distance (e.g., 3 m or more) the relief of the face is small when compared to the total distance between the camera and the head, and perspective projection approximates a parallel projection; thus, in the image, the nose is relatively smaller, the ears are more visible, the eyes will appear closer to each other, the hair and the underside of the chin are more visible.

Why does this matter? These differences are rather small, and they are typically not noticed explicitly by a casual observer viewing a single picture. However, these differences are available to the visual system and have a significant emotional impact; this was revealed by a number of psychophysical experiments that my collaborators and I carried out in the laboratory (Bryan et al., 2012; Perona, 2007). A closer (0.6–1.2 m) viewing distance makes the sitter seem more approachable, peaceful and friendly, while longer viewing distances (2.4 m and more) give the impression of aloofness and, relatively speaking, unfriendliness and hostility (see Fig. 3).

An explanation of this phenomenon is suggested by the study of proxemics (Hall, 1966): humans interact at close quarters when on intimate or friendly terms, while they maintain a distance when they have a formal relationship. It is therefore possible that a certain ‘look’ of faces (bigger nose, hidden ears) is associated to the experience of intimacy, while the opposite look (smaller nose, eyes closer together, more visible ears) is associated with formal, emotionally distant and sometimes intimidating interactions.

Thus, when portraying the head the painter and the photographer will be able to use the distance from their subject to convey different emotions.

3. Multiple Viewpoints for Portraiture

When portraying the head both the painter and the photographer are at liberty to place their viewpoint closer or further from the subject. However, when a full-length portrait is commissioned, the painter has an advantage. The photographer (and the computer artist) will be forced to step back 4–5 m or more in order to capture the entire length of the body in their picture. Closer distances may be kept if a panoramic lens is used; however, perspective will produce an apparent distortion which is usually considered unacceptable (see Note 1). The painter, instead, has more liberty: they may, for example, sketch the entire figure from far, and then come closer to fill in important details such as the face and the hands. This will happen naturally, since the painter will need a detailed
Figure 3. Relative emotional response (arbitrary y units) to the same face viewed from 60 cm ($x = 1$), 120 cm ($x = 2$) and 360 cm ($x = 3$). The plots show average responses from ten subjects who viewed fifteen faces (Perona, 2007). More complete experiments carried out on a larger set of faces may be found in (Bryan et al., 2012).

By combining different views of the body, the painter may control the emotion conveyed by a portrait without distorting the overall figure, while the photographer is forced to choose between producing a formal-looking portrait and a distorted one.

Consider the posthumous portrait of Dante Alighieri in Fig. 4. I believe that this picture was produced by combining at least three viewpoints. The viewpoint used for the lower portion of the body is a touch below the plane on which Dante is standing: both the front and the back of the feet are visible and one may peek up inside Dante’s cloak. If we assume that the cloak was cut so that the hem hung at a constant distance from the ground, the viewpoint must be rather close to Dante’s feet. Consider now the head: the viewpoint is much higher, approximately level with Dante’s eyes: one cannot see the underside of Dante’s chin, while one sees well the lower lip. Again, the viewpoint is probably rather close because one looks ‘down’ onto the lower lip and the
Figure 4. Multiple viewpoints in 1450 portrait of Dante Alighieri (Andrea del Castagno, Uffizi, Florence).

...chin, and ‘up’ into the orbits of the eyes. The right hand holding the book is viewed from a third viewpoint, roughly level with the middle of the body: the underside of the fingers is not visible and the finger pads are barely visible. There are thus at least three viewpoints for this picture; it is possible that a fourth viewpoint was used for the left hand, but I do not know how to estimate its location. It is difficult to attribute the viewpoint inconsistencies to lack of skill or inattention: in the mid-1450s in Florence, interest in perspective was at its height. Furthermore, it is known that Andrea del Castagno used skillfully multiple viewpoints in the equestrian portrait of Nicolò da Tolentino (Kubovy, 1986). I suggest instead that the painter used multiple viewpoints in order to produce a more compelling portrait of Dante.
It is perhaps unsurprising that photographic portraits predominantly depict either the sole head or the head and bust, while painted portraits throughout history rarely depict less than the bust, often include the full standing figure and sometimes a much larger scene including multiple people, horses, buildings and assorted furnishings (see Note 3). It is also telling that David Hockney felt compelled to collage photographs taken from multiple viewpoints when producing standing portraits of visitors to his studio in the '80s and '90s (Hockney, 1984; Joyce, 2002). One might say that this is the exception that proves the rule.

4. Experiments

Image editing software, such as Photoshop, levels the ground between the photographer and the painter. By combining photographs taken from different viewpoints, and comparing such collages with single-photograph portraits, I explored the difference between the ‘single-’ and the ‘multi-viewpoint’ standing portrait.

After experimenting with different arrangements, I decided to collect the images for my multi-view portraits by scanning the body lengthwise (see Note 4 and Fig. 6). This ‘slice’ method is inspired by David Hockney’s ‘L.A. visitors’ series (Joyce, 2002, p. 186) and may be thought of as a coarse discretization of a pushbroom camera (Gupta and Hartley, 1997; Seitz, 2001). The ‘slices’ were then manually composited into one picture (see Note 5). Figures 5, 7, 8 and 9 show some of the results (left multi-perspective and right single-perspective for comparison — please scale to full-screen for proper viewing). Observe in Fig. 5(left) that, while the person in the foreground looks almost seamless, the background appears to be cut up in a number of repeating scenes. This is inessential to this study — it is due to the geometry of imaging which is similar to that of a multi-camera stereo system where distant objects appear at the same location in each image. A seamless figure superimposed on a continuous background may be obtained by segmenting the picture of the sitter and overlaying it on top of the picture of the original background (not shown).

A second method, the ‘zoom’ method, consists of taking a full-length picture from far (5–6 m) followed by a sequence of four pictures each taken closer so that smaller regions of the body fill the viewfinder of the camera: knees-to-head, crotch-to-head, chest-to-head, head only. These pictures are easily aligned in Photoshop in 1–2 min. I find the results less compelling and will not discuss this method further here.

Seven experts (2 artists, 2 photographers, 1 visual psychologist, 2 computer vision/graphics researchers) evaluated the pictures; I collected opinions from most of the sitters as well (see Note 6). The experts unanimously found that
the multi-perspective portraits have a distinctively different ‘feel’. Most indicated that they appear closer — at times too close for comfort and slightly
‘jarring’. One commented that in the multi-perspective picture the subject is ‘talking to me’, while in the single-perspective picture the subject is ‘looking at me’. Curiously, most do not realize that the picture is a composite of multiple views, until they are shown the single-viewpoint picture for comparison, at which point they are surprised by the difference. When the face is photographed too close the effect is slightly caricatural (Fig. 8). More often than not, the multi-perspective portraits were preferred. There was one notable exception: the sitters mostly preferred a conventional, rather than a multi-perspective, portrait of themselves.

5. Related Work

The idea of using multiple projections in the same picture has been explored in other contexts. Hockney (1984) used multiple viewpoints in photographic portraiture to capture time and space. Wood et al. (1997) developed multi-perspective panoramas for simulating motion through 3D scenes in animated movies. The idea that using multiple viewpoints in scenes is a useful degree of freedom for the artist is discussed by Agrawala et al. (2000), who built algorithms for rendering such scenes. Balog (2004) collaged photographs taken from multiple viewpoints to portray remarkable trees in North America. Pericoli (2003) combined multiple viewpoints to produce continuous drawings of large sections of Manhattan. The same technique is used by Google engineers to render continuous views in StreetView (Anguelov et al., 2010).

Pirenne (1970) and Kubovy (1986) noticed that Renaissance artists minimized distortion by using separate and different perspective projections for architecture and for each person in the scene: this idea was used by Zelnik-Manor et al. (2005) to build wide-field image mosaics. Hockney (2001) noticed perspective inconsistencies in Renaissance paintings and suspected that artists had changed viewpoint while painting the scene. His hypothesis is that this may be due to the constraints imposed on the painter by the use of lenses and mirrors to project pieces of the scene onto the canvas.

British artist Ewan Uglow in 1967 painted a standing woman (‘Nude, 12 Vertical Positions from the Eye’) combining twelve vantage points that were...
Figure 7. Multi-viewpoint (left) and single-viewpoint (right). Photographs not adjusted for boundary continuation and not equalized for brightness.
Figure 8. (Left) Multi-viewpoint. Face photographed from 60 cm, body from 100 cm. (Right) Single viewpoint taken from about 5 m.

equally spaced along a line parallel to the standing subject’s axis. Conveniently, he decided to leave a trace of his device in the painting. Uglow’s motivates his technique by the urge to be close to his subject and yet avoid the distortion described by Urbino: “My painting of a girl standing against a wall was painted in an effort to cancel the distortion (that of an elongated lemon) which had occurred in a previous painting. This I had painted from a
very close position to a model who was standing in a room. I think it is necessary for me to be close to the model. To overcome this distortion I painted from twelve vertical positions of the eye and at a constant distance from the flat vertical wall surface” (see Note 7) (Lampert and Kendall, 2007, pp. 96–97).

**Figure 9.** Relatively seamless (left) vs ‘blocky’ figure (right). Both pictures incorporate multiple viewpoints.
6. Discussion and Conclusions

Based on the observation that pictures of heads taken from different distances have different emotional valence (Bryan et al., 2012; Perona, 2007), I observed that artists face a challenge when producing full-length realistic portraits of people. I hypothesized that combining multiple viewpoints into a single picture may give the artist useful degrees of freedom in constructing full-length portraits that produce a desired emotional effect.

This is a novel explanation for ‘perspective errors’ found in classical painted portraits (Hockney, 2001; Kubovy, 1986; Pirenne, 1970), as well as an explanation for the poignancy of Hockney’s ‘joiners’ of people (Hockney, 1984; Joyce, 2002). A systematic study of portraiture through the ages might reveal the geometrical choices and compromises that artists made in order to produce the desired emotional effect. Estimating the distance and viewpoint from which the face was painted is possible, given the fact that faces have easily recognizable landmarks whose position is known in 3D (Della Francesca, c. 1480; Kemp, 1990). The same appears more difficult, although not impossible, for other parts of the body.

A comparison of multi-viewpoint and single-viewpoint portraits constructed by collaging photographs suggests that multi-viewpoint portraits can appear overall more compelling than equivalent single-viewpoint ones. Crucially, viewers do not notice that the viewpoint is inconsistent across the body of the sitter.

The availability of photo editing software will enable photographers to produce expressive full-length portraits of people by combining multiple viewpoints. The ‘slice’ method I used is just one possibility; ad-hoc viewpoints could be used for the full figure, the head, hands and other significant parts of the body, to give each a specific projection that was designed to obtain a desired visual and emotional effect. In the limit, the photographer might wish to capture the entire light field (Levoy and Hanrahan, 1996) around the person (e.g., using a multi-camera array) and then choose the best combination of viewpoints — or perhaps the continuous viewpoint of a pushbroom camera moving along the body (Gupta and Hartley, 1997; Seitz, 2001).

Acknowledgements

This research was supported by the California Institute of Technology. Some of the initial work was carried out while I was a guest of Carlos Brody, Zach Meinen and Tony Zador at Cold Spring Harbor Laboratory. I am grateful to Stephen Shore, Michael Kubovy, Paolo Perona, Mark Stevenson, Marco Ponti, Kayley Vernallis, Ronnie Bryan, Tamara Knutsen, Michael DeWeese, Marty Banks, Ralph Adolphs and David Freedberg for inspiring conversations, help
with the experiments, encouragement and suggestions. I am grateful to an anonymous reviewer for making me aware of the work by Uglow.

Notes

1. This was clear to painters as early as the 1550s. For instance, Carlo Urbino studied the apparent distortion of the standing human body when viewed in perspective views from different distances (Codex Huygens fol. 101r) (Kemp, 1990). Paul Gauguin uses this distortion to great effect when portraying his meek friend Schuffenecker in his studio.

2. The painter will typically disguise the inconsistencies that appear where pictures taken from different viewpoints join together. This is usually not too challenging (Hockney, 2001; see Fig. 5, left, and Fig. 9, right). Disguising inconsistencies is not always necessary: modern viewers often find the ‘cubist’ blockiness produced by collaging the views on top of each other somewhat attractive, as in Hockney’s visitor series and in Fig. 9(right).

3. Just to give a few examples, culled from different periods in history, of large painted scenes containing portraits: The Gonzaga Family (Mantegna, 1475), The Equestrian Portrait of Charles V (Titian, 1548), The Burial of the Count of Orgaz (Theotokopoulos, 1586), The Night Watch (Rembrandt, 1642), The Coronation of Napoleon (David, 1807), Ceremonial Session of the State Council (Repin, 1900), Mr and Mrs Clark and Percy (Hockney, 1971).

4. Here is a recipe that works well: five ‘slice’ images are taken while translating a forward-facing camera along the vertical axis of a standing subject at a distance of 0.8–0.9 m using a 50 mm lens. The camera is held horizontally and aimed at landmarks: the bridge of the nose, the sternum, the crotch, the point in between the knees and the point in between the heels. The last photograph is taken with the camera aiming 20°–30° downwards. During the same session a ‘conventional’ single-viewpoint picture may be taken from a distance of approximately 4.5 m, such that the entire figure fits in one frame. An array of cameras at fixed positions could also be used to take the picture in one shot. However, the result would probably be less compelling as the pictures would not be centered on anatomically interesting features (see comments on Uglow’s 12-viewpoint painting below).

5. First the slices were manually scaled and translated to match the overall proportions of the body (the single-perspective image may be used as a reference) (see Fig. 7, left). Small affine transformations were then applied to align the outer edges of the body, and color correction used to even the
tones of the slices. It takes 10–15 s to shoot the pictures, and 2–5 min to align them and perfect them using elementary operations in Photoshop.

6. The pictures were either viewed on hardcopies printed using a 7-cartridge Epson Stylus 2200 printer on large format (19″ × 13″) sheets of enhanced matte paper, or on a 23″ high-quality LCD computer screen, which allows quick ‘flipping’ back-and-forth between two versions of the same portrait.

7. In the end Uglow was disappointed by his experiment: “In trying to do this a completely different form of distortion took place — not unlike some of the interpretations of the globe we see in an atlas. The result seems no better or worse than what had occurred in the picture before — just different” (Lampert and Kendall, 2007, pp. 96–97). His disappointment may be a result of his process being too mechanical; rather than tailoring his viewpoints to the important parts of the body, he treated the body as an undifferentiated pillar and chose the viewpoints following a blind geometrical process. The head is only sketched, as in many of Uglow’s figure paintings, suggesting that Uglow may have had little interest in the expression of the face. Another (later) painting by Uglow, Large Nude Walking towards You (1969–71) suffers from the apparent distortion produced by a single-viewpoint perspective (Lampert and Kendall, 2007, p. 111).

References


Della Francesca, P. De prospectiva pingendi. c. 1480.


