Computational Aesthetics in Graphics, Visualization, and Imaging
Angus Forbes and Lyn Bartram (Editors)

EXPRESSIVE 2016

Composition and Perception beyond Photorealism

Li Ji*[1] Brian Wyvill[1] Lynda Gammon[2] and Amy Gooch[3]

[1] Department of Computer Science
[2] Department of Visual Arts, University of Victoria
[3] Scientific Computing and Imaging, University of Utah

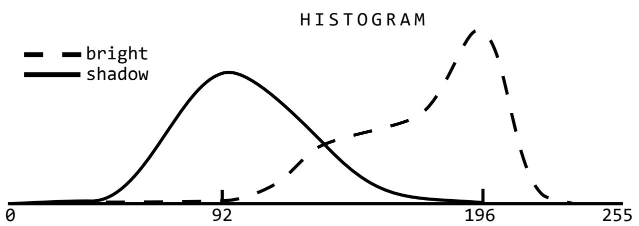
Abstract

Composition is an important aestheic aspect of celebrated works of art, and we examine a few common compositional techniques in the context of computer graphics rendering and perception. In computer graphics, photorealistic rendering simulates a camera, which defines an image of a scene in a single instant after the shutter is released. In contrast, a human observer looks at one part of a scene at a time and stitches a series of visual memories together to form a complete impression of the scene. This perception process is related to visual composition, in which an artist selectively articulates and suppresses details to direct the viewers' eyes. In non-photorealistic rendering research, painting is an important source of examples for stylized rendering. We discuss the importance of painting's specific presentation conditions, and how painting composition takes effect through a viewer's attentive looking. Based on these analysis, we demonstrate how to apply knowledge of composition to digital image synthesis with an interpolative material model and a staged photography art project.

1. Introduction

Around 1890, artist Claude Monet started a series of paintings depicting haystacks under different lighting conditions, at various times of day and at various seasons across years. This series became a systemic study of colours and shades in outdoor scenes revealed by transient natural lights [CS60, Spa92]. During this project, the painter took many canvases with him to the field, working on each version only when a particular lighting effect appeared. One painting from this series, *Haystack at Sunset near Giverny*, is shown in Figure 1.

From a photometric point of view, a clear daytime sky exhibits a luminance around the magnitude of $10^6 cd/m^2$, and shadows behind an outdoor object are several hundred times darker. In comparison, paintings are usually shown under indoor gallery lighting with luminance around $10^3 cd/m^2$, three magnitude lower than



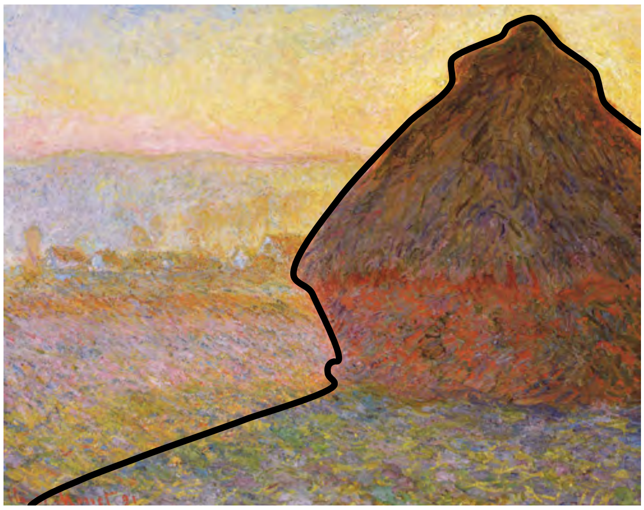


Figure 1: A histogram analysis of Claude Monet, *Haystack at Sunset near Giverny*, 1891, Oil on canvas, Museum of Fine Arts, Boston. We added a black curve to separate the painting into two parts, with their histograms shown on the left. The bright part on top of the curve is represented by the dashed plot, which peaks at 196 on a 0-255 pixel value range. The solid plot shows the shadowed part peaks at pixel value 92.

*ilucky@uvic.ca

the daytime sky. The darkest oil pigment reflects about $\frac{1}{20}$ of the light of the whitest pigment. To depict a bright landscape on the canvas, an artist must contend with this huge difference in both the absolute luminance level and contrast range. Considering the Weber-Fechner law, which states that perception is proportional to the contrast of a stimuli [Wan95], we may assume that the absolute luminance level does not matter much, as long as the contrast ratio between visual elements are maintained. Still, the contrast alone provided in painting pigments seems to be quite small compared to the depicted scene. We may therefore expect a painter to perform an artistic tone mapping process, reducing the vast contrast range to his available palette, and using every possible colour, from the whitest to the blackest, in order to minimize the loss of information. Under this circumstance, restricting one's palette to a few middle-luminance colours and abandoning most brighter or darker tones seems to be an unreasonable choice; yet this is what we have observed in Monet's haystack paintings.

We divide the haystack image into two parts and perform histogram analysis on them. In Figure 1, on top of the black curve is the part of the scene under bright lighting: the sky and the far-away fields bathed in sunshine. This part is represented by the dashed plot in the histogram. The part below the curve is under shadow, which contains the back-lit haystack and nearby grassy field, and is denoted by the solid plot. From the histogram, we can see that most colours used by the painter are fairly close to the centre, with two peaks sitting at about one-third and two-thirds of the total possible contrast range. Both ends of the histogram contain very little colour. Especially on the dark end, there are scarcely any shades darker than one fifth of the maximum luminance. The histograms indicate that the painter used little of his brightest or darkest pigments. Nevertheless, when we look at the painting, the golden sunset appears stunningly bright behind the heap of hay, and almost forces us to move our gaze away from the contour of the haystack. (The black curve we added in Figure 1 undermines Monet's lighting effect considerably. In the original painting the sunshine appears much brighter.) The deep blue and red shades on the haystack unmistakably depict shadow, showing vivid details of the individual stalks of hay and the way they were piled together. With colours around the medium luminance range, the painter constructed an image that depicts a scene with vast range of contrast. One may spend time contemplating such a painting and discover the sublime in a landscape with humble haystacks.

Perception theories about human visual systems (HVS) may account for part of the painting's visual effects, and we will briefly discuss the luminance adaptation with bleaching and regeneration in section 2. Beyond perception theories, the composition of paintings operate in complicated ways that depend not only on the image in the painting but also on the presentation conditions. We shall examine the topic of painting composition in section 3.

Mainstream research in computer graphics deals with photorealistic rendering, and pursues synthetic images indistinguishable to photographs. The research of non-photorealistic rendering (NPR) usually takes well-known paintings as examples to synthesize images that are different from photographs, but the relationship between these two art media are rarely investigated. Instead, photos and paintings are simply treated as two contrasting end points on a spectrum of possible image styles, with photography being more "physical", "accurate", therefore more objective; and painting being more "artistic" and "expressive", which means subjective creations of personal artistry. In contrast, we show that both forms are artificial styles of visual expression, and their compositional techniques are closely related to each other. We examine this idea in the following order:

1. In Section 2, we discuss the difference between photography and *seeing*. We start by examining photorealistic rendering within its historical context. Then, we explore the process of *seeing*, in which we only look at one part of the scene at a time, and the visual impression of each part is locally adapted. This process is then related to the composition of the haystack painting.
2. In Section 3, we focus on painting, and explore a few compositional techniques that enable artists to create images both believably realistic and visually attractive. By examining two of Vermeer's paintings and related art work, we investigate how artists articulate and suppress details to direct the viewers' eyes.
3. In Section 4, we present experiments in rendering and digital photography. To further illustrate our idea, we blended multiple texture images onto one object's surface, and combined photographs from different lighting conditions into a single image guided by compositional rules.

2. Photorealism and Perception

> *Computer graphics (CG) technology has advanced to the point where models and effects can now be rendered with near-perfect photorealism.*
>
> Kristy Barkan,
> *ACM SIGGRAPH NEWS*, January 2015 [Bar15].

Photorealistic rendering is the central topic of contemporary computer graphics rendering. As its name suggests, photographs of real world scenes are usually defined as the ground truth standard for digitally synthesized images; with the development of digital cameras, easily captured digital images can also be used as the source of data for rendering [WLL*08, AWL15]. The choice of photographs as the source and the goal for graphics rendering is related to the social context of its development. At the time when computer graphics was invented, photographic images dominated the mass media such as newspapers and TV, delivering stories from foreign affairs to community events. The affordable film cameras at that time enabled everybody to take snapshots and fill albums with photos from daily lives. In this photographic era, we rely heavily on photographic images to perceive the world around us. As observed by art historian Linda Chase:

> The photograph, for all its ubiquity, offered the Photorealist a realm that had never really been fully explored by art. ... When Degas employed photographic perspective and distortion, it was considered a bizarre way of seeing and depicting the world, an aberration. Today these photographic aberrations are so commonplace that we can look at a painting that employs extreme distortion and out-of-focus areas and comment on how real it looks. In the nineteenth century, arguments raged over what was real — the photographic depiction, artistic convention, or the

Figure 2: Tom Blackwell, *Triumph Trumpet*, 1977. Oil on Canvas, 180 x 180 cm.

> way the eye perceives — and the photograph's truthfulness was often questioned. Nevertheless, the visual language of the photographic process soon gained an aura of validity that took precedence over all other ways of seeing. "We accept the photograph as real," observed Richard Estes in a 1972 interview, "Media has to affect the way you see things." And Tom Blackwell took the idea even further: "Photographic images, movies, TV, newspapers are as important as actual phenomena. They affect our perception of actual phenomena." [Cha02]

The applications of computer graphics in the consumer market produce imagery to be shown on media that are already dominated by photographic images. By aligning itself with the established photographic style of visual representation, digital rendering acquired broad attention from the visual effects and video game industries, which facilitated its rapid development in the last few decades. The computer graphics literature borrowed the term 'photorealistic' from the art world; in its original context it means a genre of painting that 'appears like a real (untouched) photograph' [Mei02, Let13]. One well-known painting in this genre is shown in Figure 2. In the graphics rendering context, we tend to use this term to mean 'as real as a photograph'. This notion assumes that photographs faithfully represent reality, and reflects the previous discussion about the ubiquity and validity of photography in the contemporary world. For some applications the rendering result must be photographic, such as injecting virtual objects into a live-action film footage, where the synthesized images have to be consistent with the captured images from a camera. In addition, research on non-photorealistic rendering has demonstrated possibilities of communicating information with imagery beyond the photographic style. Traditional art media such as painting mostly doesn't appear like photographs, and in contemporary visual arts, photographers are leveraging advanced digital cameras and computers to compose images that deviate from what is conventionally considered a straightforward exposure. We like this non-photographic imagery and accept it as understandable yet attractive depictions; this is because we see the world differently from a camera, which defines a picture in single moment exposures to record slices of light transportation [Dur02].

Standing in front of a landscape lit by bright sunshine, we can see details all around us, including clouds in the bright sky and the shadowed wall of a house. We achieve this by staring at each part of the scene for a short time before shifting our line of sight onto the next region, and after a while we have seen the entire landscape. Our line of sight needs to be constantly rotated, because the viewing angle within which our eyes can see the maximum spatial and colour resolution is quite small. This viewing angle is less than 10 degrees, which is about the same size as the thumbnail as it is seen when the arm is fully stretched to the front. The narrow viewing angle corresponds to the small fovea centralis on the retina, where the colour-sensitive photoreceptor cells (cones) are packed with the highest possible density (Figure 3). Moving away from the fovea centralis, the density of the colour-sensitive photoreceptor decreases rapidly, and the colour-insensitive photoreceptor used for night vision (rods) begins to appear [Wan95]. To see a large scene, we look at one part of the scene at a time, and acquire the appearance of the entire scene afterwards, by stitching a sequence of visual memories together.

During each short gaze, the HVS adapts itself to the image projected on the fovea to best distinguish details. The adaptation can be categorized into two kinds: those we can feel and those we cannot. For example, we clearly feel it if we change the focal point of our eyes to look at objects with a different distance to us. This action involves muscle movement to morph the shape of the lens, and it rapidly changes the retina image. The contraction and dilation of the pupil and strong bleaching after-images can also be clearly felt when the surrounding lighting changes abruptly. Since these adaptation behaviours generate a direct sensation, we can clearly tell their absence when we look at planar, low contrast visual media, such as photographs, paintings or a computer screen, even if a similar visual effect is being simulated. On the other hand, more subtle visual adaptations that generate weak or no direct sensations have granted artists opportunities for adding less noticeable touches to their work. Since a viewer can not reliably tell whether an adaptation behaviour is present or absent, artists can simulate an adaptation effect to suggest a specific viewing context. Research of perception based tone mapping has explored various methods for simulating visual adaptation in compressing HDR images [KMS05, PTYG00, AG06], but they generally do not distinguish between those adaptation behaviours that can be sensed and those cannot, nor do they relate these behaviours to compositional techniques. For an example of visual adaptation, we sketch the bleaching-regeneration process here to prepare our discussion of painting composition in the next section. Interested readers are referred to the textbooks and researches on vision, perception and tone mapping for the details [Boy79, Cor70, EJGAC*15, RE12].

In our photoreceptor cells, photosensitive molecules change their chemical configurations when they absorb photons within their sen-

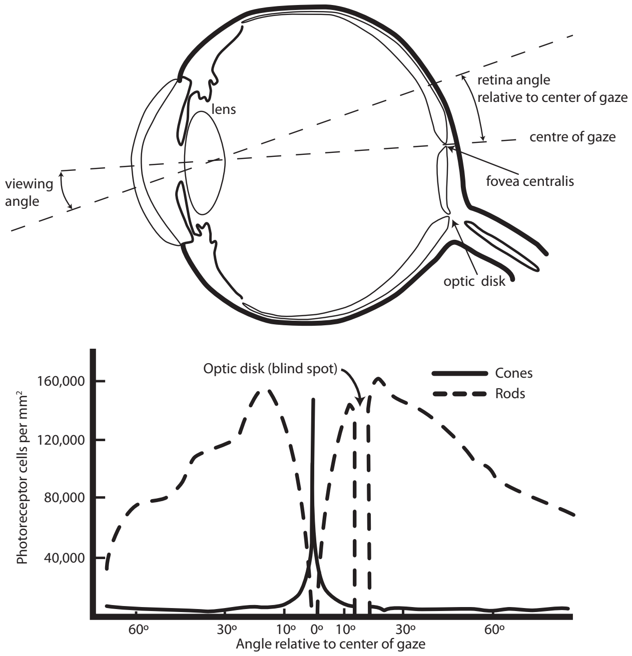


Figure 3: The fovea centralis is a small dent on the retina, where the colour photoreceptor cells (cones) are most densely packed. It corresponds to a small viewing angle of approximately 6 degrees, in which we can see the highest spatial and colour resolution. Moving away from the fovea centralis, the density of cones drops rapidly, and the low light photoreceptor (rods) starts to appear [Boy79, Cor70].

sitive frequency range. This change results in a series of biochemical reactions, and in the end triggers neural signals on the subsequent visual pathway. A photosensitive molecule is said to be *bleached* after it absorbs a photon, and will be no longer capable of absorbing photons. This terminology came from the fact that the photosensitive pigment molecules extracted from animal eyes have a colour. When the pigments are placed in a test tube and exposed to strong light for a while, the solution becomes transparent, indicating that all pigment molecules have changed their configuration, and no photon is absorbed. Obviously, our photoreceptor cells have a way of constantly replenishing the photosensitive pigment molecules after they are bleached, otherwise we would not be able to see anything after an initial exposure to light.

The mostly accepted theory about how the photoreceptor cells replace the bleached pigment molecules with photosensitive ones is the *regeneration* theory. It proposes that photoreceptor cells rarely make new pigment molecules; instead, they keep reverting the bleached molecules back to their photosensitive status, such that there is always a constant number of total pigment molecules in a photoreceptor cell. In a healthy eye, photoreceptor cells maintain a stable biochemical environment for pigment regeneration, and the regeneration speed is proportional to the concentration of bleached pigment molecules in that particular cell. Assuming the concentration (proportion) of photosensitive molecules is $p, p \in [0,\ 1]$, the bleaching-regeneration kinetics can be modelled as:

$$\frac{dp}{dt} = \frac{1-p}{T_0} - \frac{Ip}{T_0 I_0}.$$

In this equation, T_0 and P_0 are constants for each particular kind of photoreceptor cells. The first term on the right hand side represents the regeneration process, and the second term represents the bleaching process with a given light intensity I. The strength of the neural signal emitted from a photoreceptor cell depends on how many photosensitive pigment molecules are being bleached, which is approximately proportional to the second term. When luminance I increases from a steady state, this term increases proportionally, meaning we see a brighter light. But the increase in the second term immediately gives $\frac{dp}{dt}$ a negative value, which causes p to decrease. In this way, the strength of the emitted neural signal is decreased, meaning now we have a diminished sensation of the incoming light, and the bleaching-regeneration process moves to a new steady state with a smaller p. More importantly, this process does not happen in an instant, and is not carried out at a constant speed. At the moment of a change in light intensity I, $\frac{dp}{dt}$ reaches the largest absolute value and the luminance adaptation is the fastest. After a while, when the bleach-regeneration process is close to its next steady state, $\frac{dp}{dt}$ becomes smaller and it takes longer to fully adapt to the new luminance condition. For the red and green cones used for daylight vision, it only takes a few seconds to perform the major amount of the bleaching-regeneration adaptation, while it may take up to half an hour to completely stabilize p with a given luminance in a vision lab. When luminance environments gently change, the bleaching-regeneration kinetics moves between stable states though a mild biochemical process, and is mostly unnoticeable.

The bleaching-regeneration process, combined with other adaptation mechanisms in the photoreceptor and the subsequent visual pathway, creates a constantly shifting visual impression of the physical retina image. Every time we shift our gaze, our eyes always attempt to adapt to the local lighting to best distinguish details. How well our vision can adapt to a particular region depends on how long we look at that region and what has been viewed just before. Particularly for artists, who need to carefully observe the scene to depict it, they usually stare at each part of the scene for a long time and have their eyes accurately adapted during each gaze. In Monet's Haystack painting (Figure 1), the painter needed to attentively observe the shade on the shadowed side of the haystack before he could paint it. This observation gave his eyes ample time for luminance adaptation, which in turn placed the visual neural signal around a moderate strength, and generated a visual impression of a well lit side of the haystack. Similarly, a prolonged gaze adapted the painter's eyes to the high light intensity from the sky, and the dazzling vision of the sunset recedes to mildly bright. Although the painting appears rather non-photorealistic, the painter faithfully reproduced his visual impression of every part of the scene. The important point to reiterate here, as Monet complained in his letter, is that such a visual impression can be only formed *over time*:

> ...for in October (1890) he (Monet) wrote to Geffroy:
>
> "I'm working terribly hard, I'm struggling stubbornly with a series of different effects (stacks), but at this time of year the sun sinks so fast that *I can't keep up with it. I'm beginning to work so*

slowly that I despair, but the longer I go on, the more I see that it is necessary to work a great deal in order to succeed in rendering what I seek - *'instantaneity'*, above all the envelope, the same light spreading everywhere - and *more than ever I'm disgusted with things that come easily in one go.* I am more and more obsessed by the need to render what I experience..." [Spa92]. (italics marked by the author)

As a result of such careful observations, the painting not only demonstrates an effective perception based tone mapping method, but also strongly suggests a particular viewing process that achieves such a perception. The painting does not conform to a visual perception of a quick glance into a sunset landscape. Rather, a conventional photograph with over exposed sky and dark shadows will better conform to a hasty glance, since the output of our photoreceptor can easily be saturated by the bright sunset, given insufficient time for adaptation. By confining his palette to the medium luminance range, the painter instructs his viewers that *in order to see the scene like this*, one must have taken time and carefully observed it in a meditating manner. This calm, motionless viewing experience is further strengthened by the blurry depiction of horse-drawn carts in the distance, which suggests a motion-blur effect from a static, long exposure camera.

The invention of photography has greatly changed the way images are produced. It obviated the need for representational painters, while pushed leading artists like Claude Monet to produce images clearly dissimilar from photographs. This led to the impressionism genre of paintings in the 19$^{\text{th}}$ century. To differentiate themselves from what is supposed to be done by the camera machine, the impressionists use large paint-strokes and vibrant colours, and intentionally leave out small details. One distinct feature of the impressionism genre from previous representational paintings is that impressionists use continuous shade variation to define spatial geometry, instead of discrete gestures such as line drawings and contours. Ironically, this is also the essential property of photographs. Before the invention and popularization of cameras and camera-like devices such as camera-obscuras, this visual effect can not be observed from previously existing forms of art. We could say the technology of photography both undermined the established practise of representational art, and at the same time facilitated the creation of new forms of visual images. Likewise, computer graphics rendering is approaching its consummate ability to synthesize photographic images. Pursuing better aesthetic quality and deviating from an exact photographic depiction will be a natural development, a next step similar to those that the impressionist artists took after photographs dominated the market of representational images. At this turning moment, we believe an in-depth study of visual composition using established paintings as examples will be helpful and instructive.

3. The Artist's Shade

It's perhaps slightly less obvious that photorealistic imagery might not be the best choice for communicating information about 3D scenes...

Adam Finkelstein and Lee Markosian,
IEEE Computer Graphics and Applications,
July 2003 [FM03].

Figure 4: Johannes Vermeer, Woman in Blue Reading a Letter, c. 1662-64, oil on canvas, Rijksmuseum, Amsterdam. 46.6x39.1 cm. The background wall is painted with little colour variation in the blocks marked as (a), (b) and (c) [Whe95].

Research of non-photorealistic rendering usually refers to painting and drawing as examples [GG01, SS02, KLK*00]; and tone mapping takes traditional photography as the guide in addition to knowledge of perception and signal processing [RSSF02, YS12]. Except for a few user-assisted methods [LFUS06, KKL10], research on both topics seeks to best communicate the information contained in the input three-dimensional scenes or high dynamic range imagery [DD02, MKMS04, SMDA12]. The automatic methods do not, in general, attempt to inject additional information on top of the given scenery to be depicted. In contrast, the discipline of creative art despises the idea of straightforwardly mimicking the subject of depiction. From one of the earliest art theory of painting [vH78] to a recent summary on photography [Edw06], it has been insisted that the artistry in creating images lays not in how well the depicted subject appears but in how to construct and deliver a novel idea through the depiction. In short, the aesthetic value is correlated to inserting and communicating additional information beyond objective depiction. Techniques for this purpose have acquired the name 'composition'; and particularly for representational art, 'composition' means to convey aesthetic effects to its viewers without breaking the realistic visual appearance of the depicted scene.

In this section, we examine two of Vermeer's paintings, "Woman in Blue Reading a Letter" and "The Music Lesson" to further explore the relationship between visual adaptation and composition. We choose Vermeer's paintings as examples because his depiction of domestic scenes are praised as accurately photorealistic yet visually attractive. Unlike the impressionists who intentionally differentiate their work from photographs, Vermeer only deviates his representational paintings from an exact photographic image in a subtle but powerful manner.

In "Woman in Blue Reading a Letter"(Figure 4), part (b) of the background wall is tinted towards blue. This blue shade reminds us of the global illumination effects [Shi03], as the blue coat on the foreground figure will reflect blue light onto the wall. Nevertheless, Vermeer's execution of this visual effect is different from an exact physically based solution. The blue light fills up part (b) in a uniform and exaggerated way, and cuts off sharply on the boundary between part (a) and part (b) of the background wall. In contrast, in part (c) which is to the right of the figure, the wall is tinted to brown-yellow, instead of blue. From a physically based point of view, one would expect the blue shade on the wall to fall off gradually around the figure in both directions. In the painting, however, the three blocks of the background wall are constructed as three monochromatic blocks with little gradient [Whe95, Lie12].

The boundary of the blue shade in part (b) is delineated by its visual context in the image space, such as the chair, the map and the main figure. Since blue is an ostensible shade in oil painting, the artist suppressed its expression on the background wall to avoid distracting his viewers from the main character in the front. In contrast, the blue dress on the character is painted with bright and intense (high saturation) blue colours. The silky reflections on the dress, rendered with sophisticated colour variation and indiscernible fine paint strokes, show the artist's exceptional ability to construct accurate, complex realistic surface shadings. Similarly, the yellow-brown shade on part (c) of the wall can be related to its lower right position in the painting. This earthy tint serves to stabilize the visual structure of the painting and echoes the brown map on the top. The manipulation of details and colour intensity does not break the impression of a realistic scene, but instead draws our eyesight unmistakeably onto the foreground character. Equipped with eye-tracking devices, contemporary perception researchers have demonstrated that variations in colour hue and value

Figure 5: Jeff Wall, A View from an Apartment, 2004-05, transparency on light-box, 1670 x 2440 mm

do attract the viewer's gaze; while monochromic shade blocks do not catch as much attention [Liv15].

An artist can also force his viewers to carefully examine every part of an image even if there appears to be some central characters, by artificially enhancing details in the seemingly peripheral parts. A well-known example in contemporary visual art is Jeff Wall's "A View from an Apartment" (Figure 5). The artist, who came from a background of traditional painting, used computers to edit and combine many photographs together and created a hyper-realistic image that has sharp details and intense colours everywhere. The reflection on the foreground TV, the two characters in the middle ground and the port outside the window are all in perfect focus; and every object in the scene seems to be of roughly equal brightness. In addition, the artist presented this image on a transparent layer on top of a huge (2.4 by 1.6 meters) light box. The sheer size of the work prevents a viewer from looking at the whole image in one gaze [Fri08]. The sharply defined details with the equalized luminance encourage a viewer to spend time and closely examine the photograph part by part. In comparison, we remember that the size of the "Women in Blue Reading a Letter" painting is only 50 by 40 centimetres.

"The Music Lesson" (sometimes titled *"A Woman at a Virginal with a Gentleman"*) is another important work in Vermeer's oeuvre. Because of its apparent photorealistic quality, researchers have experimented with reconstructing the painting using both film photography and computer graphics rendering [Ste01, Hil02]. When we place the original painting side by side with its photographic and rendering reconstructions, the differences between them become obvious (Figure 6). Apart from the different overall tonality and perspective, the difference on the shadowed part of the Turkish carpet (white square in Figure 6. a) is evidently clear. In the original painting, this part of the carpet contains vibrant details of the fabric in a shimmering blue colour, while in both the photographic and computer graphics reconstruction this part is completely black. But this blacked-out region is not due to technical mistakes of the reconstruction experiments. In Vermeer's studio setting, bright sunshine penetrates the studio window and is reflected directly towards the viewpoint by the white wall and the brilliant, silky clothes on the characters. Because of the limited dynamic range of the film negative, one has to stop down (shrink the aperture) or to shorten the exposure to properly portray the main characters at the end of the room. This will make details on the shadowed side of the carpet becomes invisible. The computer graphics reconstruction with physically based rendering resulted in a similar image.

If we were to look at Vermeer's studio in reality and to fix our gaze on the shadowed carpet, we would see the exotic fabrics on it. Then, we may lift our eyesight to look at the warm sunshine on the characters farther into the studio. After a short moment, our eyes adapt to the bright light, and the characters reveal themselves to us with full detail. The painter's depiction mimics this experience, and therefore suggests this process to its viewers. On a canvas approximately twice as big as "Woman in Blue Reading a Letter", we can not see the entire "Music Lesson" painting in one gaze but must shift our eyesight. The luxurious Turkish carpet, which is a symbol of wealth during Vermeer's time, is the nearest, largest object directly facing the viewer. The bright blue shades on the shad-

(a) (b) (c)

Figure 6: (a). Johannes Vermeer, The Music Lesson, 1662-65, oil on canvas. 74 x 64.5 cm. The Royal Collection, Her Majesty Queen Elizabeth II. (b) A photography reconstruction of this painting. (c) A computer graphics reconstruction of the painting. The shadowed part of the Turkish carpet on the table is marked by the white square. In the original painting it exhibits vivid blue fabrics. In both the photography and rendering reconstruction, where the main characters are properly captured, the carpet under shadow is completely black.

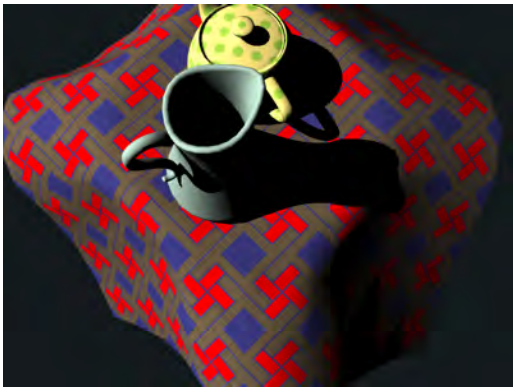
owed part traps a viewer's gaze on this foreground object before the viewer moves attention to the smaller figures further into the room. In this way, the composition creates an articulated distance between the viewer and the characters in the painting, and establishes an intimate atmosphere around the characters that is undisturbed by the crowd of viewers in front of the frame [Whe95, Ste01]. In contrast, the blacked-out carpets in the photograph and computer graphics reconstructions do not contain any interesting visual details; and a viewer would immediately focus on the characters in the centre of the scene. If an artist does not want viewers to spend too much time on a piece of front-facing fabric in the foreground, it may indeed be painted black, as the example in Figure 7. Here, the fabric dripping in the foreground is painted only with a hint of dark blue, and the major part of it is lost in the deep black. In this way, the painter focusses our attention on the key objects on top of the table, such as the skull, the book and the flute.

The compositional techniques we examined above include suppressing and articulating details on different parts of the image, and stitching together visual impressions from different adaptation contexts. These techniques can be considered as a reverse-engineering of the visual adaptation process, such that the parts that an artist wants to attract attention are painted with more details, and vice versa. The composition is an additional layer of information constructed by the artist on top of the objects in the depicted scenery. In addition to what is in the image, the artist instructs viewers *how the image should be seen*. Furthermore, the size and the presentation of paintings enhances their compositional effects. In galleries and museums where paintings are shown, the white gallery wall, the elaborated frames and the careful lighting all encourage a viewer to pay close attention to the images in the paintings. Because composition implies a complicated order of seeing, it requires a viewer to *attentively look at a static image over a period of time*.

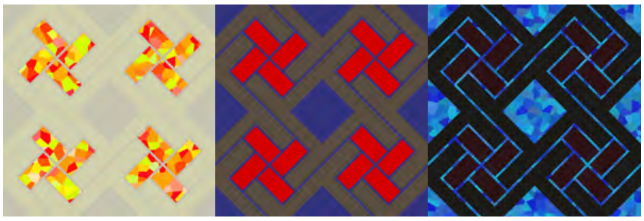
Figure 7: Harmen Steenwijck, An Allegory of the Vanities of Human Life, ca. 1640, oil on canvas, 39.2 x 50.7 cm, National Gallery, London.

4. Experiments in Rendering and Photography

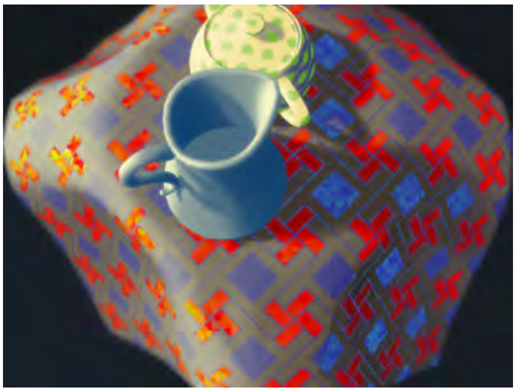
In this section, we summarize two experiments of rendering and digital photography to further illustrate the idea of image composition. Our first rendering experiment approaches the effect of the Turkish carpet in "The Music Lesson", which seems to be a difficult scene for strictly physically based rendering methods. One rendering result is shown in Figure 8. In this experiment, we manually created multiple texture images for a table-cloth mesh. When rendering this mesh, we first define a 'brightness value' for each point



(a) A table-cloth scene using diffuse shading.



(b) The three texture images for the table-cloth mesh.



(c) Result of our interpolative material model.

Figure 8: The interpolative material model. Figure 8a is rendered with only diffuse shading. In comparison, Figure 8c is rendered with our material model taking the textures in Figure 8b, and exhibits vivid details in the shadows.

on the mesh using the Lambertian diffuse term $\mathbf{n}\cdot\mathbf{l}$, taking into account the shadows from other objects [SM09]. The 'brightness value' is stored on the object surface in a barycentric coordinate per-face texture, similar to the 'Mesh Colour' method proposed by Yuksel et al. [YKH10]. Then, we interpolate the final pixel colour by sampling all texture images and taking the 'brightness value' as the interpolation weight. The interpolation does not need to be linear, and a customized mapping curve can be used. In the shown example with three texture images, our shading equation can be written as:

$$s_{\mathbf{p}} = f_s(\mathbf{n}\cdot\mathbf{l}), s_{\mathbf{p}} \in [0,\ 1] \text{and}$$

$$c'_{\mathbf{p}} = \begin{cases} 2(c_d(0.5 - s_{\mathbf{p}}) + c_a s_{\mathbf{p}}), & 0 \leq s_{\mathbf{p}} < 0.5 \\ 2(c_a(1 - s_{\mathbf{p}}) + c_h(s_{\mathbf{p}} - 0.5)), & 0.5 \leq s_{\mathbf{p}} \leq 1, \end{cases}$$

In the above equation, $s_{\mathbf{p}}$ stands for the 'brightness value', which is mapped from the diffuse term $\mathbf{n}\cdot\mathbf{l}$ by a customized function f_s. The terms c_h, c_a, and c_d stand for the colour sample on point $\mathbf{p}$ from the texture images representing the appearance of the object in highlight, moderate and and dim luminance conditions. The final shading colour for surface point $\mathbf{p}$ will be $c'_{\mathbf{p}}$. Technically speaking, our method can be related to non-photorealistic rendering methods that mix a palette of colours or textures to create stylized shadings [BTM06, TABI07, PHWF01]. Rendering tools such as Autodesk Maya sometime come with a 'ramp shader' that allows blending of multiple input texture images into one material. The goal and implementation of our material model is different from these techniques.

The goal of our material model is to separate lighting and surface shading. In the commonly used material models, usually only one colour texture is provided for each mesh in a virtual scene, representing the reflectance of the surface. Then, the surface colour of the mesh under different lighting conditions can be *extrapolated* from this single image with a lighting model and surface geometry. In this way, the variation of the appearance is closely coupled to the light source, whose parameters are globally constant. This model conforms to the photographic manner of depiction, which defines every object in an image at a single instant, therefore lighting conditions cannot change from object to object. In contrast, our rendering method *interpolates* the pixel colour from the pre-defined multiple texture images. These texture images resemble the intended appearance of the target mesh under various lighting and visual conditions; but since they are edited in external software tools, artistic touch can be freely added onto them.

In our method, the light sources do not emit colour information at all, but only send out an abstract 'brightness value' $s_{\mathbf{p}}$ on the objects' surfaces. The entire possible set of appearances of an object is confined in its texture images c_h, c_a, and c_d. Since the brightness value is calculated via a globally consistent light source, the weight of the shading interpolation still behaves consistently across objects, but the representation of brightness is now a separately defined parameter for each object. Our rendering results show rich details in the shadowed part of the table-cloth mesh, at the same time clearly defining the structure of light and shadows on the object's surfaces (Figure 8c).

With more sophisticated rendering systems than a single material model, there will be a large amount of parameters that affect the quality of the synthesized image. It then becomes difficult to clarify whether a visual feature comes from a particular parameter setting or a compositional technique. For a succinct example of image composition, we briefly discuss an art project based on digital photographs, which are by definition perfectly photorealistic. Lessons learnt from such a project can be directly applied to digital rendering as multi-pass photorealistic rendering methods.

Figure 9: The final image of the photography art project.

In this art project, we follow the style of historical flower still-life paintings [Tay95], with modern motifs such as a computer box (Figure 9). We take multiple shots towards a table-top arrangement under different lighting conditions, including direct sunshine from the side window, and different indoor studio lighting (Figure 10). For each shot, we keep the camera's position, orientation and aperture stop fixed. Afterwards, the photos are combined with manually painted alpha layer masks. The alpha masks specify which exposure will be visible on each pixel position. In the final image, the red lily in the centre, the red daisy on the upper-right and the violet orchid come from the photo taken under direct sunshine (Figure 10a). Other flowers are lit by indoor studio lighting (Figure 10b). The background fabrics, the part of the computer box behind the bouquet and the blackberry vein come from a photo with a dimmed indoor lighting (Figure 10c). The result is a picture in which the red lily appears very bright and attractive, and other surrounding flowers gradually come into view following the central flowers. In the actual project approximately 20 exposures of the same scene are used.

Combining multiple photographs with different lighting conditions has been explored in the topic of image space relighting and perception enhancement [FAR07, MKVR09]. When this method is applied for composition, the goal is not to enhance every detail in the source imagery but to selectively enhance and *suppress* details according to how a viewer's sight should be directed. From a viewpoint of image perception enhancement, Figure 10b might be closest to an optimal image, since the structural details in both background and foreground objects are clearly defined; and every object is illuminated with moderate strength diffuse lighting without strong specular reflection or hard shadows. Precisely because of this, the picture is not successful as an art project since it does not contain information about how it should be seen by a viewer. (Unless the compositional goal is to intentionally overwhelm viewers with sharp details as in Jeff Wall's example, in that case the details need to be enhanced much further.)

We printed this image on a large (1 meter by 2 meters) piece of photographic paper and presented it in a gallery setting. During our discussion in the visual arts department, a few professional photographers noticed the particular lighting effect, and asked about the setting of the lights in our studio, and whether the photograph was edited. Most viewers remarked that it is simply an attractive photo with "good lighting". Before we introduced our workflow, the audiences did not recognize that the image may not be a single exposure but a combination of a few photographs under dramatically different lighting.

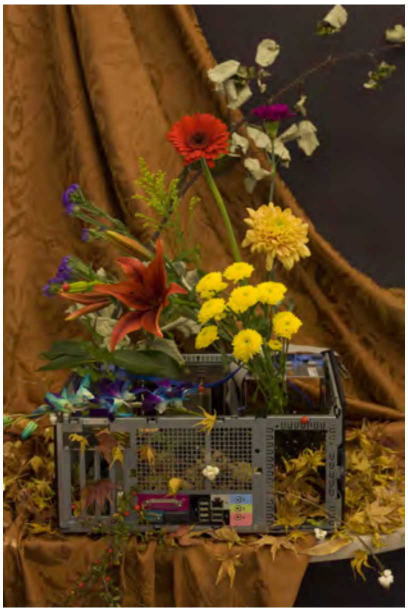
For the future work, to automate our art project as an image space relighting algorithm may require the incorporation of computer vision methods. If we render from the object space, the compositional information may be attached to the models. In either case, the most urgent future work is to develop a language that can describe compositional intentions about the synthesized image, and to insert this information into the rendering process [Dur02]. To our knowledge no popular technical frameworks on this goal have been published.

5. Conclusion

Through our examples, we have demonstrated that an artist can intentionally elaborate or suppress details in an image to achieve compositional effects. Based on knowledge of visual perception, we discussed how these compositional decisions direct a viewer's eyes. One difference between creative visual art and image enhancing methods is that artists frequently conceal information in the depicted scene, while perception based image space relighting or tone mapping methods tend to bring out details everywhere in the image. For painting, which is a popular topic in non-photorealistic and expressive rendering research, we shall be reminded that their presentation conditions are important, such as the size of the image, the framing and the gallery environment. Painting implies that viewers look at a static image attentively and over a period of time, in order to let its composition take effect. Extending painting techniques such as paint-strokes and composition rules onto animated and interactive rendering should be carefully considered for their possible aesthetic effects.

Our visual impressions of reality are closely related to the visual context of seeing. The order in which we look at objects, the amount of time we fix our gaze and the surrounding environments all contribute to the formation of the visual impression of

(a) Direct sunshine photo.



(b) Indoor studio lighting photo.

(c) Dimmed indoor lighting photo.

Figure 10: Three source photographs for composing the final image shown in Figure 9. Figure 10a provides shading definition for the central objects, while the other two supplies pixels for the peripheral objects.

a real scene. Because these visual contexts are highly subjective, and are dependent on particular locations and moments, to define or evaluate an objective 'realism' with human viewers will be difficult [AG06]. At the very least, we can be confident that realism can not be deduced merely from the physical properties of objects in the depicted scene, without considering the perception of a human viewer. We show that artists seek to better represent realistic visual impression by creating compositions based on the manner in which the depicted objects should be seen. As suggested by Aaron Hertzmann [Her10], realistic visual impression results from an interactive process between the physical world and its observers. The composition serves as a summary and guideline of such a process.

References

[AG06] ASHIKHMIN M., GOYAL J.: A reality check for tone-mapping operators. *ACM Trans. Appl. Percept. 3*, 4 (Oct. 2006), 399–411. 3, 10

[AWL15] AITTALA M., WEYRICH T., LEHTINEN J.: Two-shot svbrdf capture for stationary materials. *ACM Trans. Graph. 34*, 4 (July 2015), 110:1–110:13. 2

[Bar15] BARKAN K.: Visual effects: Are computer graphics always the answer? ACM News Blog, 2015. http://www.siggraph.org/discover/news/visual-effects-are-computer-graphics-always-answer, retrived Januray, 2016. 2

[Boy79] BOYNTON R. M.: *Human Color Vision*. Holt, Rinehart and Winston, 1979. 3, 4

[BTM06] BARLA P., THOLLOT J., MARKOSIAN L.: X-toon: an extended toon shader. In *Proceedings of the 4th international symposium on Non-photorealistic animation and rendering* (New York, NY, USA, 2006), NPAR '06, ACM, pp. 127–132. 8

[Cha02] CHASE L.: The not-so-innocent eye: Photorealism in context. In *Photorealism at the Millennium* (2002), Harry N.Abrams, Inc, pp. 11–22. 3

[Cor70] CORNSWEET T.: *Visual Perception*. Academic Press, 1970. 3, 4

[CS60] C. SEITZ W.: *Claude Monet*. Harry N.Abrams, Incorporated, 1960. 1

[DD02] DURAND F., DORSEY J.: Fast bilateral filtering for the display of high-dynamic-range images. *ACM Trans. Graph. 21*, 3 (July 2002), 257–266. 5

[Dur02] DURAND F.: An invitation to discuss computer depiction. In *Proceedings of the 2Nd International Symposium on Non-photorealistic Animation and Rendering* (New York, NY, USA, 2002), NPAR '02, ACM, pp. 111–124. 3, 9

[Edw06] EDWARDS S.: *Photography: A Very Short Introduction*. Oxford University Press, 2006. 5

[EJGAC*15] E. JACOBS D., GALLO O., A. COOPER E., PULLI K., LEVOY M.: Simulating the visual experience of very bright and very dark scenes. *ACM Trans. Graph. 34*, 3 (May 2015), 25:1–25:15. 3

[FAR07] FATTAL R., AGRAWALA M., RUSINKIEWICZ S.: Multiscale shape and detail enhancement from multi-light image collections. *ACM Trans. Graph. 26*, 3 (July 2007). 9

[FM03] FINKELSTEIN A., MARKOSIAN L.: Guest editors' introduction: Nonphotorealistic rendering. *IEEE Comput. Graph. Appl. 23*, 4 (July 2003), 26–27. 5

[Fri08] FRIED M.: *Why Photography Matters as Art as Never Before*. Yale University Press, 2008. 6

[GG01] GOOCH B., GOOCH A.: *Non-Photorealistic Rendering*. A. K. Peters, Ltd., Natick, MA, USA, 2001. 5

[Her10] HERTZMANN A.: Non-photorealistic rendering and the science of art. In *Proceedings of the 8th International Symposium on Non-Photorealistic Animation and Rendering* (New York, NY, USA, 2010), NPAR '10, ACM, pp. 147–157. 10

[Hil02] HILLIARD C.: Vermeer, master of light. Siggraph 2002 CAF, Electronic Theater. 6

[KKL10] KANG S. B., KAPOOR A., LISCHINSKI D.: Personalization of image enhancement. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on* (June 2010), pp. 1799–1806. 5

[KLK*00] KLEIN A. W., LI W., KAZHDAN M. M., CORRÊA W. T., FINKELSTEIN A., FUNKHOUSER T. A.: Non-photorealistic virtual environments. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 2000), SIGGRAPH '00, ACM Press/Addison-Wesley Publishing Co., pp. 527–534. 5

[KMS05] KRAWCZYK G., MYSZKOWSKI K., SEIDEL H.-P.: Lightness perception in tone reproduction for high dynamic range images. *Computer Graphics Forum 24*, 3 (2005), 635–645. 3

[Let13] LETZE O.: *Photorealism: 50 Years of Hyperrealistic Painting*. Hatje Cantz Verlag, 2013. 3

[LFUS06] LISCHINSKI D., FARBMAN Z., UYTTENDAELE M., SZELISKI R.: Interactive local adjustment of tonal values. In *ACM SIGGRAPH 2006 Papers* (New York, NY, USA, 2006), SIGGRAPH '06, ACM, pp. 646–653. 5

[Lie12] LIEDTKE W.: *Vermeer: The Complete Paintings*. Harry N. Abrams, 2012. 6

[Liv15] LIVINGSTONE M. S.: *Vision and Art (Updated and Expanded Edition)*. Harry N. Abrams, 2015. 6

[Mei02] MEISEL L. K.: *Photorealism at the Millennium*. Harry N.Abrams, Inc, 2002. 3

[MKMS04] MANTIUK R., KRAWCZYK G., MYSZKOWSKI K., SEIDEL H.-P.: Perception-motivated high dynamic range video encoding. *ACM Trans. Graph. 23*, 3 (Aug. 2004), 733–741. 5

[MKVR09] MERTENS T., KAUTZ J., VAN REETH F.: Exposure fusion: A simple and practical alternative to high dynamic range photography. *Computer Graphics Forum 28*, 1 (2009), 161–171. 9

[PHWF01] PRAUN E., HOPPE H., WEBB M., FINKELSTEIN A.: Real-time hatching. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 2001), SIGGRAPH '01, ACM, pp. 581–. 8

[PTYG00] PATTANAIK S. N., TUMBLIN J., YEE H., GREENBERG D. P.: Time-dependent visual adaptation for fast realistic image display. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 2000), SIGGRAPH '00, ACM Press/Addison-Wesley Publishing Co., pp. 47–54. 3

[RE12] RITSCHEL T., EISEMANN E.: A computational model of afterimages. *Computer Graphics Forum 31*, 2pt3 (2012), 529–534. 3

[RSSF02] REINHARD E., STARK M., SHIRLEY P., FERWERDA J.: Photographic tone reproduction for digital images. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques* (New York, NY, USA, 2002), SIGGRAPH '02, ACM, pp. 267–276. 5

[Shi03] SHIRLEY P.: *Realistic Ray Tracing, Second Edition*. A K Peters, Natick, MA, 2003. 6

[SM09] SHIRLEY P., MARSCHNER S.: *Fundamentals of Computer Graphics, Third Edition*. A. K. Peters, Ltd, Natick, MA, 2009. 8

[SMDA12] SHAN Q., MEYER M., DEROSE T., ANDERSON J.: Tone mapping high dynamic range videos using wavelets. *Pixar Technical Memo* (2012). 5

[Spa92] SPATE V.: *Claude Monet: Life and Work, with over 300 illustrations, 135 in colour*. Thames and Hudson Ltd., 1992. 1, 5

[SS02] STROTHOTTE T., STEFAN S.: *Non-Photorealistic Computer Graphics: Modeling, Rendering, and Animation*. Morgan Kaufmann, 2002. 5

[Ste01] STEADMAN P.: *Vermeer's Camera: Uncovering the Truth Behind the Masterpieces*. Oxford University Press, 2001. 6, 7

[TABI07] TODO H., ANJYO K.-I., BAXTER W., IGARASHI T.: Locally controllable stylized shading. *ACM Trans. Graph. 26*, 3 (July 2007). 8

[Tay95] TAYLOR P.: *Dutch Flower Painting, 1600-1720*. Yale University Press, 1995. 9

[vH78] VAN HOOGSTRATEN S.: *Inleyding tot de hooge schoole der schilderkonst: anders de zichtbaere werelt (Introduction to the High School of Painting, or the Visible World)*. 1678. "Want een volmackte Schildery is als een spiegel van de Natuer, die de dingen, die niet en zijn, doet schijnen te zijn, en op een geoorlofde vermakelijke en prijslijke wijze bedriegt." (A perfect painting is like a mirror of Nature, in which things that are not there appear to be there, and which deceives in an acceptable, amusing, and praiseworthy fashion.). 5

[Wan95] WANDELL B. A.: *Foundations of Vision*. Sinauer Associates Inc., 1995. 2, 3

[Whe95] WHEELOCK A. K.: *Vermeer & the Art of Painting*. Yale University Press, New Heaven and London, 1995. 5, 6, 7

[WLL*08] WEYRICH T., LAWRENCE J., LENSCH H., RUSINKIEWICZ S., ZICKLER T.: Principles of appearance acquisition and representation. In *ACM SIGGRAPH 2008 Classes* (New York, NY, USA, 2008), SIGGRAPH '08, ACM, pp. 80:1–80:119. 2

[YKH10] YUKSEL C., KEYSER J., HOUSE D. H.: Mesh colors. *ACM Trans. Graph. 29* (April 2010), 15:1–15:11. 8

[YS12] YUAN L., SUN J.: *Computer Vision – ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part IV*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, ch. Automatic Exposure Correction of Consumer Photographs, pp. 771–785. 5