Countering Racial Bias in Computer Graphics Research

Theodore Kim and Holly Rushmeier and Julie Dorsey, Yale University Raqi Syed, Victoria University of Wellington Wojciech Jarosz, Dartmouth College Derek Nowrouzezahrai, McGill University A.M. Darke, University of California, Santa Cruz

SIGGRAPH Talk, August 8, 2022

Good afternoon everybody

We're going to switch gears from what we usually talk about at SIGGRAPH

Problems of racial bias have been getting increasing attention over the last few years

We're going to switch gears back to what Haven was looking at at the top of this session.



Especially after the murder of George Floyd in 2020



and all of the worldwide protests that followed, including right here in Vancouver, which involved thousands of people

To my knowledge, this is the first year that these issues have appeared in the technical program of SIGGRAPH, I believe this is the first in-person technical talk that openly discusses it.

So let's get right to it.

Computer graphics research has a race problem

Computer graphics research has a race problem.

Computer graphics research has a race problem

Systemic Racism

You've probably heard the term systemic racism, and we're not exempt from it.

Racism is baked into our everyday institutions, whether we like it or not, and if we don't try to deal with it directly, we will just end up reproducing existing inequalities.



This has actually already happened in computer graphics. If you look at old classic manuals on movie lighting, like John Alton's *Painting with Light*

LIGHT AND THE HUMAN FACE

The first step in illumination is the study of light and the **human face.** If you watch carefully, under varying circumstances in subways, buses, streetcars, in the open, on beaches, at home, or elsewhere, you will find that faces look absolutely different when silent and expressionless than when they laugh, smile, talk, or cry. Hitherto unsuspected hidden wrinkles and harsh lines suddenly appear

when he talks about lighting a "human face", he's talking almost exclusively about lighting a white face.



So of course he did, it was the 1940s, this was pre-Civil Rights in the US.

I wouldn't have been allowed to use the same bathroom as a white person in the US back then.

Things were different back then.



Fast forward to today, when we talk about rendering a human face, we're still talking the same way about the same thing.



You do a sampling of rendering papers, especially the foundational papers on this topic, like the Jensen et al. 2001 paper, and the language is the same.

Let's render a human face, and by human face, we're going to show exclusively white faces.

Here's a whole bunch of rendering papers over the last 20 years that proport to be rendering "human faces", "human skin" or most insidiously "skin"

They only show white faces. If you look in these papers, there are no black or brown people.

Just like Alton in the 1940s, when we say "human face" we mean "white face".



In graphics when we talk about "skin" we're talking about white skin.



We're going to codify the light transport phenomena that's most important to white skin, subsurface scattering, and make that synonymous with skin



Now unfortunately, once you cast things as a double integral, it becomes essentially impossible for non-specialists to dispute.

$$L_o(\mathbf{x}, oldsymbol{\omega}) = \int_{S^2} \int_{\partial\Omega} S(\mathbf{x}_i, oldsymbol{\omega}_i, \mathbf{x}, oldsymbol{\omega}) L_i(\mathbf{x}_i, oldsymbol{\omega}) doldsymbol{\omega}_i d\mathbf{x}_i$$

So the racial assumption is now baked into the math.



We look at the faces of Black people, and it's not the ethereal glow of subsurface scattering that dominates, it's these really intricate and expressive specular highlights that dominate. Subsurface scattering does not characterize all skin, but because we live in a world freighted with a legacy of systemic racism, this is what our formulations look like.



There are papers that try to diversify, but they are in vast minority. And, they're hobbled from the start by the fact that they have to start from mathematical models that were custom-made for white skin.

Where's the customized GGX or Burley distribution that capture the highlights on Delroy Lindo's face? There is none. We haven't done the research. That is systemic racism in action.

I'm only picking on the rendering folks because it's the easiest to cram into 10 minutes. This has happened in other places in graphics.



When we talk about hair, we explicitly avoid talking about this kind of hair. Professional stylists call this "Type 4" hair, otherwise known as Afro-textured, or kinky hair.



This doesn't show up in graphics. You look at papers over the last 20 years, and it's all straight or wavy hair.

Again, the problems are systemic, and all this history predates us, and goes back to the analog era.



Look at the hair that Alton was lighting, this is what counted as human hair in the 1940s, but we have carried this bias forward into the present day



I gave a talk on this last year at the DEI summit at SIGGRAPH

So, this topic appeared at SIGGRAPH last year

This is the first year we're discussing it in the technical program.



That talk is actually on YouTube, here's the QR code for it.

You can watch that talk if you want more details on everything I've just gone over.

How do we *measure* human features?

I want to talk about something a little different today though. Let's start with yet another seemingly basic question.

How do we measure human features in computer graphics?

How do we *measure* human skin?

To start with, how to we measure human skin?

We can take a closer look at some of the good faith efforts that graphics researchers have put into trying to capture the span between white and black skin.

Analysis of Human Faces using a Measurement-Based Skin Reflectance Model

 Tim Weyrich *
 Wojciech Matusik †
 Hanspeter Pfister †
 Bernd Bickel *
 Craig Donner ‡
 Chien Tu †

 Janet McAndless †
 Jinho Lee †
 Addy Ngan §
 Henrik Wann Jensen ‡
 Markus Gross *



 Photograph
 Rendering
 Original Model
 Appearance Change

 Figure 1: Photograph compared to a keeper dered using our skin reflectance model. The rendered image was composited on top of the photograph. Kight: Changing the albedo and BRDF using statistics of measured model parameters from a sample population.
 Sample population.

One of the largest scale studies done in graphics on skin diversity is this one from 2006, Weyrich et al.

| Skin | Skin Color | Sun Exposure Reaction | Subjects |
|------|----------------|-----------------------|----------|
| Туре | | | (M/F) |
| Ι | Very white | Always burn | N/A |
| II | White | Usually burn | 8/6 |
| III | White to olive | Sometimes burn | 49/18 |
| IV | Brown | Rarely burn | 40 / 8 |
| V | Dark brown | Very rarely burn | 13/2 |
| VI | Black | Never burn | 4/1 |

Table 2: The *Fitzpatrick* skin type system and the number of subjects per skin type.

They use something called the Fitzpatrick system to characterize skin color. You can see the types on the left, and the colors they correspond to in the second column.



Other papers have used this too: here's one from Donner in 2008. It actually has a bunch of the same authors as that 2006 paper.

(Seems like a contradiction that I showed this paper as a white-skin-only paper before, but it's not. They claim to measure different skin types, but in the end only render an image of a white person. This sadly happens in multiple papers. You got the diverse measurements! Show us the renderings!) **Chromophore variation:** For further analysis, we eliminate outlier pixels within each patch reconstruction by discarding pixels where one of the coefficients falls outside the 5- and 95-percentiles; all numbers refer to the full set of 39 scans. First, we concentrate on skin type. Of subjects that cover the Fitzpatrick skin types I through V [Fitzpatrick 1988], we measured patches on the lower arm, both on the inside and on the more strongly pigmented outside. This provides us with the skin-type dependent average concentrations shown in Table 2. Unfortunately, most of the statistics are with respect to a single subject each (column n). Differences between skin types are quite noticeable and, as to be expected, mainly established in the melanin concentration.

They use it here

A Biophysically-Based Model of the Optical Properties of Skin Aging

Jose A. Iglesias-Guitian Carlos Aliaga Adrian Jarabo Diego Gutierrez

Universidad de Zaragoza



Here's Iglesias 2015, this is Diego Gutierrez's group





Again, the Fitzpatrick system.

There's more, but I hope you get the point. A bunch of people in graphics use this thing called the Fitzpatrick scale.

Editorial

The Validity and Practicality of Sun-Reactive Skin Types I Through VI

The concept of sun-reactive "skin typing" was created in 1975' for a specific need: to be able to classify persons with white skin in order to select the correct initial doses of ultraviolet A (UVA) (in joules per cubic centimeter) in the application of the then newly developed technique for the treatment of psoriasis—oral methoxsalen photochemotherapy (PUVA).² The need arose as a result of experience with several nations who were a "dark" phenotype and a light tan at seven days." This group is skin type II. These are fair-skinned individuals with blond, red, or brown hair, green or hazel eyes, and skin that burns and peels easily. These individuals tan slightly only after repeated exposures. Also, a subgroup of skin type IV will respond: "A slightly tender burn at 24 hours and a moderate tan at seven days." This is skin type III and is the largest group in the United States.

> Thomas B. Fitzpatrick, MD Department of Dermatology Harvard Medical School Boston, MA 02114

What's the Fitzpatrick scale? It goes back to Thomas Fitzpatrick, a Harvard dermatologist who proposed a skin typing system back in the 1970s.

Here's one of his papers from the 1980s that gets cited a lot.



We can look at one of the illustrations from one of his papers, here's one from the 1980s. It's a little odd once you examine this actually.



On the left he's talking about sun exposure



in the middle, how much you burn after 3 hours in the sun



and then how tan you are after a week

The system is actually based on a questionnaire. How much do you burn after 3 hours in the sun, etc. This is weird, right? The speed at which you burn in the sun is a proxy for how much melanin is in your skin, but it's not actually a direct measurement, right?

Editorial

The Validity and Practicality of Sun-Reactive Skin Types I Through VI

Fitzpatrick himself complained about it in his own lifetime. If we go back to his 1988 paper that everybody cites, he says

Editorial

The Validity and Practicality of Sun-Reactive Skin Types I Through VI

> What is needed for an objective method of sunreactive skin typing is a quantitative means of measuring the degree of the tanning response—none (ie, 0) for E persons (skin type I) and some number (eg, 1 to 10) for P persons (skin types II through IV,

"what is needed for an objective method of sun-reactive skin typing is a quantitative means of measuring the degree of tanning response."

This is from the end of his paper. He's talking about what his system *doesn't* do.

So we have it from Fitzpatrick himself: stop treating this like some sort of scientific measurement. It's just a questionnaire. It's not a lab measurement.


What's going on? How come even Fitzpatrick himself was dumping on this own scale in his lifetime, but decades years later we have graphics people using it?

It's not just graphics. If you look at computer vision and medical physics, they all use this scale. It's everywhere.



Why not do something like this: have a palette of colors, and if you want to classify somebody's skin color, you just hold this palette up against their skin, and read off the number.

People did this in the past. This isn't perfect: we're not using a calibrated sensor of some kind. But, it's more direct than a questionnaire asking about how fast you burn in the noon-day sun.

The system you see here is called the von Luschan scale, and it was a really popular system up until the 1950s

Why isn't it used anymore? It seems like a fine system – more direct than Fitzpatrick anyway.



I told you, it gets really ugly.

These images of the von Luschan scale?



got them from the US Holocaust Museum.

This was the scale used by Nazis in World War 2 to sort people into the master race and the inferior races.

This scale was used for genocide.



I want to be fair to Felix von Luschan himself here – he passed away in the 1920s before all this happened. It is not clear that he ever intended his work to be used this way.

But during WW2, this is what his scale was used for.

I am speculating here, but it's not hard to see. After WW2, when the full scale of the Nazi eugenics program became clear, nobody wanted to use this anymore.

Again, because it had been used for genocide.

Editorial

The Validity and Practicality of Sun-Reactive Skin Types I Through VI

The concept of sun-reactive "skin typing" was created in 1975' for a specific need: to be able to classify persons with white skin in order to select the correct initial doses of ultraviolet A (UVA) (in joules per cubic centimeter) in the application of the then newly developed technique for the treatment of psoriasis—oral methoxsalen photochemotherapy (PUVA).² The need arose as a result of experience with several nations who were a "dark" phenotype and a light tan at seven days." This group is skin type II. These are fair-skinned individuals with blond, red, or brown hair, green or hazel eyes, and skin that burns and peels easily. These individuals tan slightly only after repeated exposures. Also, a subgroup of skin type IV will respond: "A slightly tender burn at 24 hours and a moderate tan at seven days." This is skin type III and is the largest group in the United States.

> Thomas B. Fitzpatrick, MD Department of Dermatology Harvard Medical School Boston, MA 02114

When this other Fitzpatrick scale stuff came along in dermatology, it was something else, anything else, something not associated with Nazis, people just glommed onto it.

This is not hard to imagine. I want to emphasize again that I'm only speculating here. There is probably lots of scholarship and books written on this topic.



I'll be honest, when I hit this point in my historical digging, I had to stop. The history of this stuff was getting too painful to look at.

This is not what I had signed up for when I came to SIGGRAPH. I suspect some of you are probably feeling the same way right now.

Why are we talking about the Holocaust?

How do we *measure* human skin?

In graphics, this is how we measure human skin. It's a weird tanning-based scale that people glommed onto because people were trying to back away from the Nazis.

How do we *measure* human hair?

Let's run it again with human hair.

In graphics, how do we classify hair? There's actually not as much on this – nothing as established as the Fitzpatrick scale.



What we saw before was that when we say "hair" we mean "straight hair".



The most popular hair typing system out there is called the Walker system.



It's named after this guy Andre Walker -- he developed this system back in the 1990s when he started his own line of hair care products.



His main claim to fame is that he was Oprah Winfrey's hair stylist for years and years. Here you can see him working with his main client.

NO NAZIS. OPRAH. It's the story of her hair stylist becoming a successful entrepreneur! Way more fun. I'll take that any day.



Now we've seen three separate scales for measuring human features. Fitzpatrick, von Luschan, and Walker.



Notice a pattern here. Type 1 is always the white, European feature.

Pale skin is Type 1.



Pale skin is Type 1.



Straight hair is Type 1.



The Black features, like dark skin and kinky hair, this is always some big number. It's not even consistent



In Fitzpatrick, black skin is Type 6



In von Luschan, it's type 36



In Walker it's type 4.



There's more systems. Loreal has a system for hair classification, I've added to the bottom here. It follows the pattern



Google just released a new skin typing system back in May called the Monk system.

Same problem.

All of these systems give the European feature a special status. In all cases, whiteness is #1.

Literally, whiteness is #1. This is white supremacy.



This is not idle word play. This is a direct reference to

the fact that the Nazis used a system that put them at #1

when establishing the master race

during World War 2.

How do we *measure* human features?

So again, we're coming back to the basic question of how we measure human features in computer graphics.

Why do the hegemonic features always gravitate towards 1?

$1 \times x = x$

We can think about this more quantaitatively: 1 has a very special meaning.

It is the multiplicative identity. Multiply it by a number, you get the number back. No other number but 1 has this property.

Literally, "the identity." A very special position within a number system.

$$\begin{aligned} 1 \times x &= x \\ \frac{\mathbf{n}}{\|\mathbf{n}\|} &= \hat{\mathbf{n}} \end{aligned}$$

If you work in graphics, you do this a billion times a day. You normalize a vector.

$$\begin{aligned} 1 \times x &= x \\ \frac{\mathbf{n}}{\|\mathbf{n}\|} &= \hat{\mathbf{n}} \\ \|\hat{\mathbf{n}}\| &= 1 \end{aligned}$$

Normalize to what? Normalize to 1. Once you peg the magnitude of the vector to 1, all sorts of special stuff happens.

One is literally considered "normal". It is the baseline. It has a special status in our minds.

It is the identity. It what is normal.



What effect do all these number 1s have on our research? How does it influence our research priorities?

Just from a day-to-day perspective, if your code doesn't work for the identity case, if it doesn't work on unit vectors, or on the identity matrix, you know it's broken, right?

You make extra sure that the identity case is fully debugged before moving on to anything else, right? That's how research practice works.

And that's what we've done, but across decades of research, writ large.



Get the identity case working.



Get the identity case working.

There's no technical reason that this should be the identity case. These specific features do not have a special status.

There is no reason to believe that getting this case right is some technical precondition for getting the other features right!

So what do we do now?

A measurement system that *lacks* idempotence?

$$1 \times x = x$$

I am going to pose a question to you, the audience.

Do you have an idea for a better measurement system?

There's actually a pretty clean problem statement here. Can we devise a measurement system that lacks idempotence? No number has special status like 1, so no measurements can then have a special status?

The answer here might be no. Maybe some of you number theorists out there can say "oh, no, you can't have that. that's number systems 101." I'm not a number theorist, so that would be an unfortunate answer, but it would be an answer.

If you can think of a clean solution, I heartily invite you to submit it as a short Talk to SIGGRAPH next year. Really. I would love to see it.

Can math fix this?

There is a broader question here. Can math even fix this problem? We should consider that the answer to this question might be no.

It might be that we come up with a whole new system, but whatever value gets assigned to white skin will eventually just become shorthand for "good".

My colleague Wojchiech Jarosz pointed this out. We see this happen with arbitrary number systems all the time.



If you went to college in the USA, when I say "I have a 4.0" or "I scored a 1600", you know what I mean, right?

There are numbers that have a halo of "good" around them, even though they're not 1. Even if we come up with a different way of measuring, it can be systematically distorted as well.

This might happen, it might not. I don't know. Again, if you out in the audience have an idea for this, it's worth a try. What we're doing right now is not working.



The good news here is that if you think it's important to combat these issues, you're not alone.

I said this is the first year that these issues are being discussed at SIGGRAPH, but they are being discussed WIDELY.

Here are SIX OTHER events at SIGGRAPH this year discussing bias at a technical level.

It's not just me that cares about this - lots of other people in the community do too.

Sex and Gender in the Computer Graphics Research Literature

Ana Dodik, MIT Silvia Sellán, University of Toronto Theodore Kim, Yale University Amanda Phillips, Georgetown University



Wednesday 2:22pm - 2:29pm East Building Room 1-3

There will be a few other chances to talk about these and related issues this year

One is the related issue of gender representation in our research practices. Are our research practices being trans-exclusionary?

There will be an in-person roundtable this Wednesday in East Building Room 1-3.
Countering Bias in Computer Graphics Research (The BOF!): One Year Later

Theodore Kim and Holly Rushmeier, Yale University Raqi Syed, Victoria University of Wellington Wojciech Jarosz, Dartmouth College Derek Nowrouzezahrai, McGill University James Malazita, Rensselaer Polytechnic Institute



TONIGHT, 6:30 PM

The other is that this will be discussed in a Birds of a Feather, on Zoom, in about TWO HOURs.

Here's the name of the BoF, and here's a QR code for where it is on Hubb. I'll post this to the Discord channel as well.

Countering Bias in Computer Graphics Research (The BOF!): One Year Later

Theodore Kim and Holly Rushmeier, Yale University Raqi Syed, Victoria University of Wellington Wojciech Jarosz, Dartmouth College Derek Nowrouzezahrai, McGill University James Malazita, Rensselaer Polytechnic Institute



TONIGHT, 6:30 PM

Thank You

And with that, thanks for your attention. I'd be happy to take questions now.