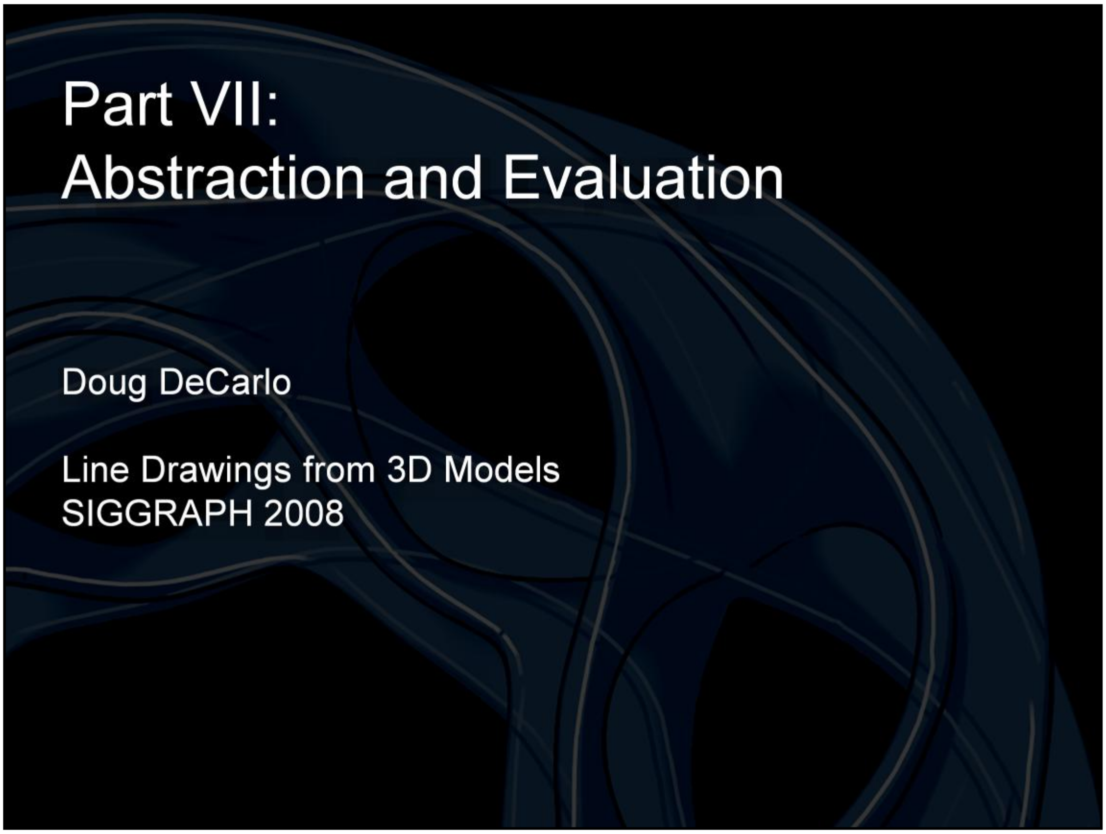


Now that you have an idea of how an artist can make a drawing,
we can talk about how artists control
what content you get out of their drawings.

I’ll now talk about methods
for achieving meaningful abstraction,
and ways to assess
how effective a drawing actually is.

Abstraction

from “Drawing with Pen and Ink”, Arthur Leighton Guptill

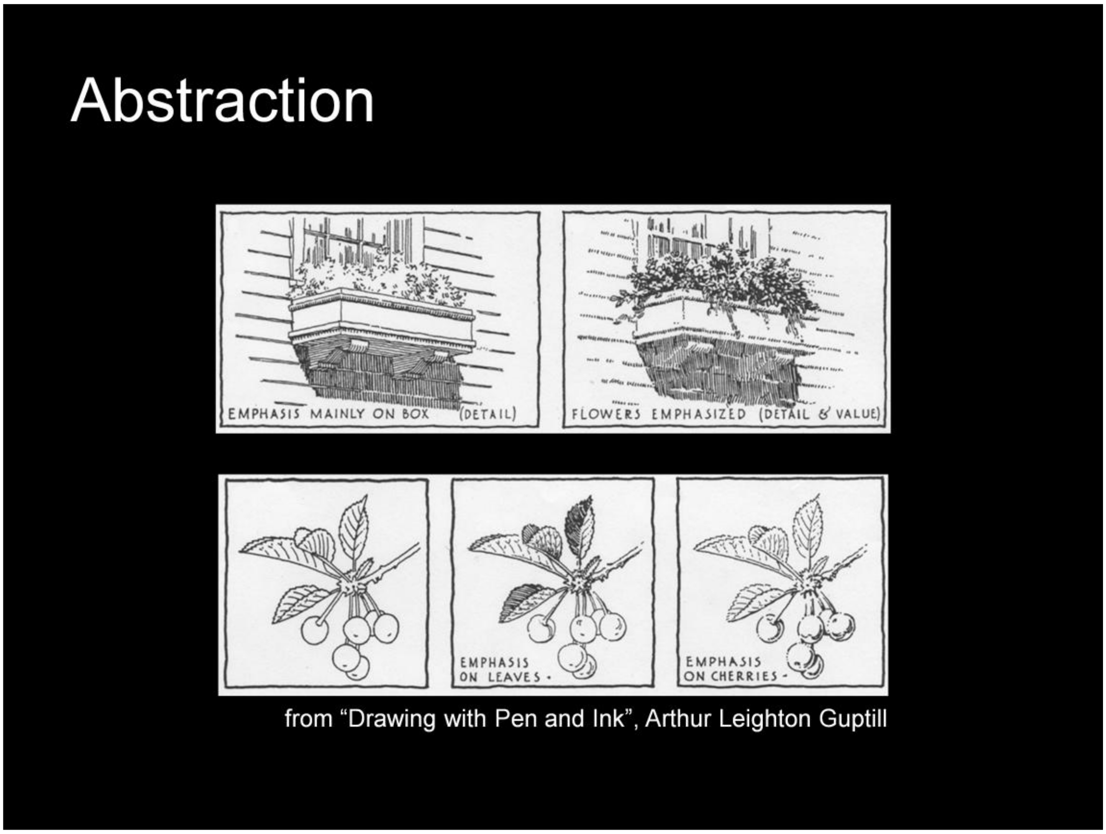
Artists can design effective imagery
by changing or leaving out specific visual content.

The result of this process
encourages particular interpretations for the viewer
and enhances the viewer’s understanding of the scene or situation.

This is the process of abstraction.

It’s a tool for effective visual communication.

Here, Guptill adapts the shading in this drawing
to guide your attention to different parts of the scene.



Artists often omit content,
such as the detail on the flowers on the top left.

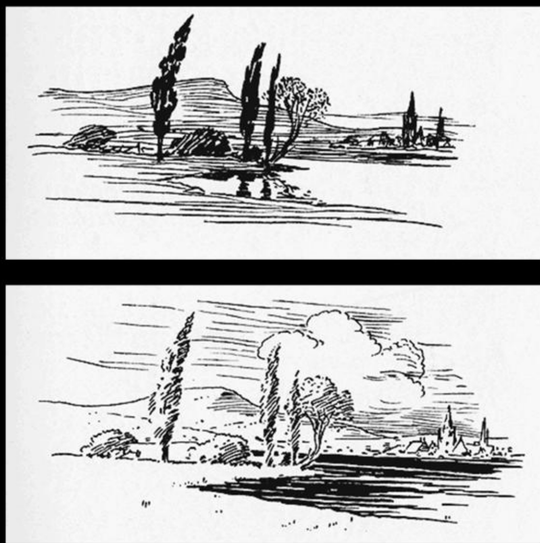
Contrast this with the detail Guptill included on the right
in order to stress the flowers.

Same for the cherries here,
where the focus can be on the leaves,
on the cherries,
or split between both.

There are a variety of means to do this.

The particular visual style and medium
determine the kinds of omissions and distortions that are possible.

Abstraction



from “Line and Form”, Walter Crane

In this example by Crane, for instance,

the form of the elements of a scene are very different

between these two different renderings.

Apart from selectively including various scene elements

such as the clouds, we see how Crane

adapts the details in the shape of the distant trees and buildings.

However, their rough shape remains

so that we can easily identify these objects.

Abstraction in NPR

Automatic approaches

- Models of image salience can make predictions about what content is important

[Collomosse 2002]

How does this work in NPR?
There are a range of possibilities.

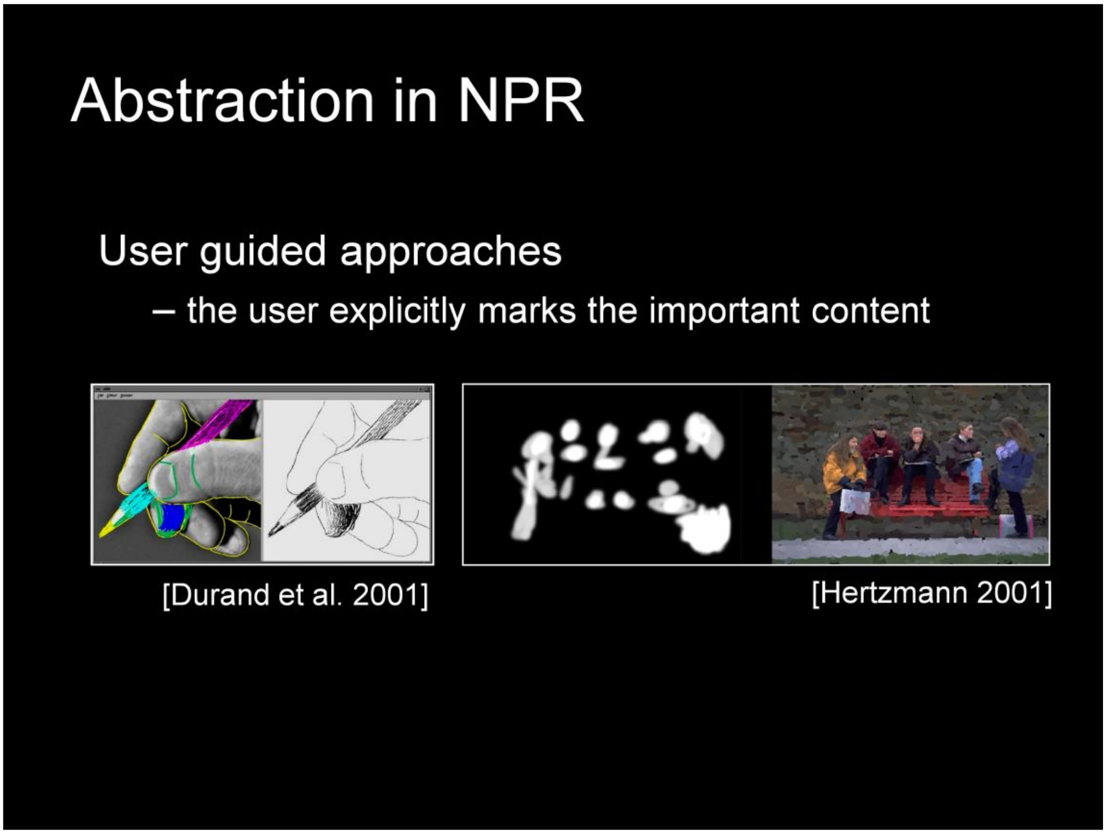
The most important difference comes from
how the important content is selected.

But techniques also differ
in how they go about retaining or removing content,
given a particular visual style and medium.

Let’s start with NPR approaches that work from photographs.

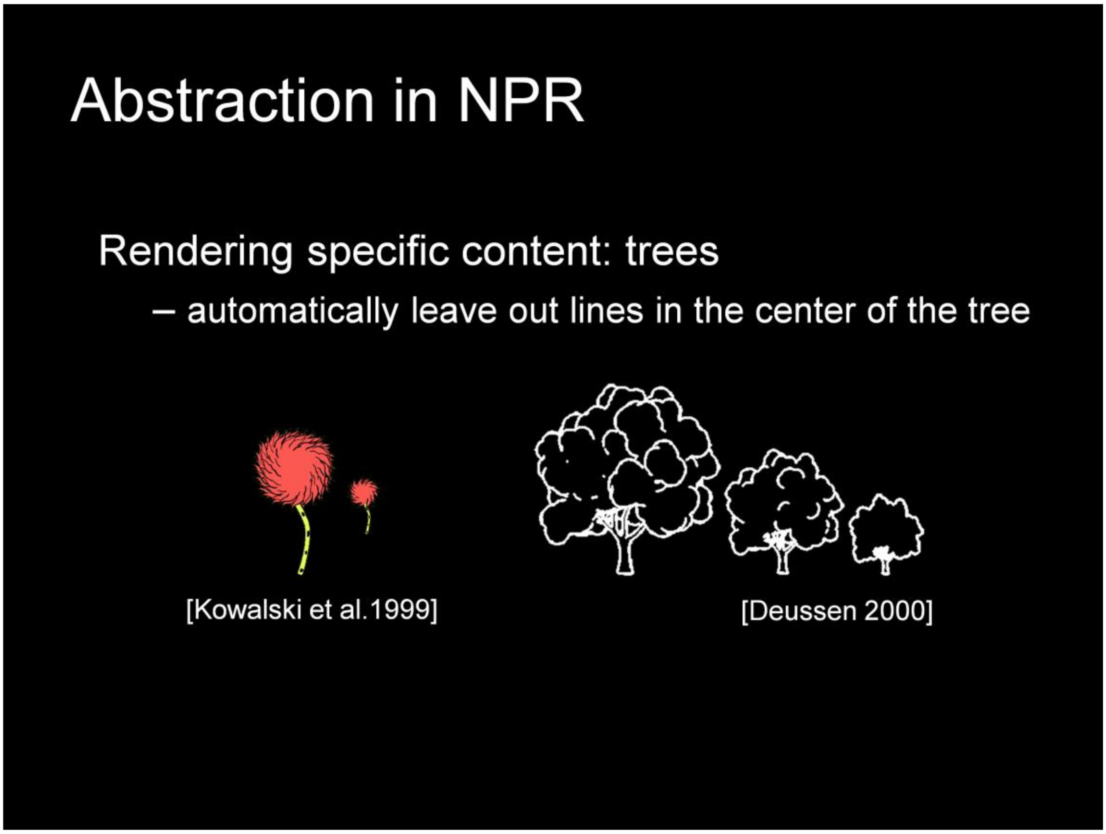
For painterly rendering, a fully automatic approach
might attempt to preserve important content
as determined by a computational model of image salience,
which predicts how noticeable particular parts of the image are.

When such predictions select the important content,
this is quite effective.



Another approach to producing effective artistic renderings is to have a user explicitly mark important regions of a picture.

The understanding is that these regions of the picture will be rendered with finer detail.



In 3D scenes, the important content is located based on where it is in the world.

However, in restricted domains where the content is known, such as the rendering of 3D models of trees, heuristics can be applied that create effective omissions.

Here, detail in the center is left out as the tree is drawn smaller.

Abstraction in NPR

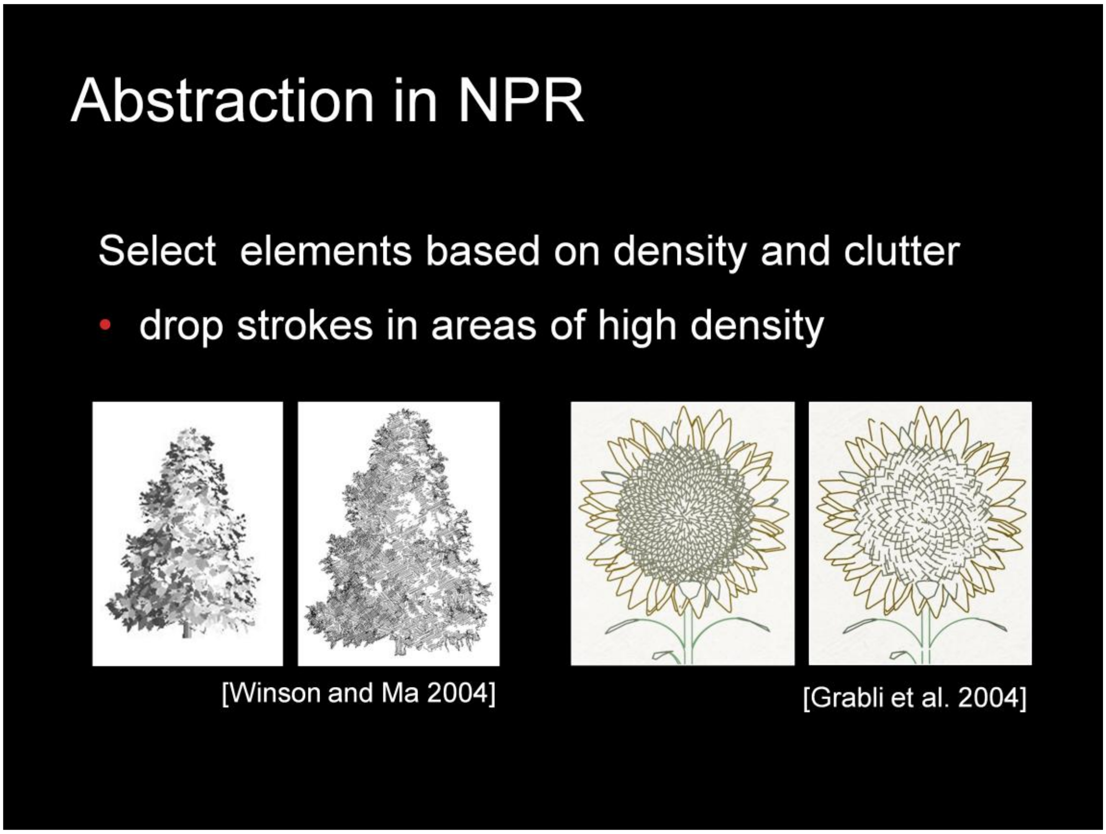
Indication in pen and ink illustration

– the user specified what content was important

[Winkenbach and Salesin 1994]

In more general domains,
a user must specify the important content, just as they do in images.

This was the case in Winkenbach's system
for pen and ink illustration,
where the process of texture indication
was guided by a set of marks drawn by the user.



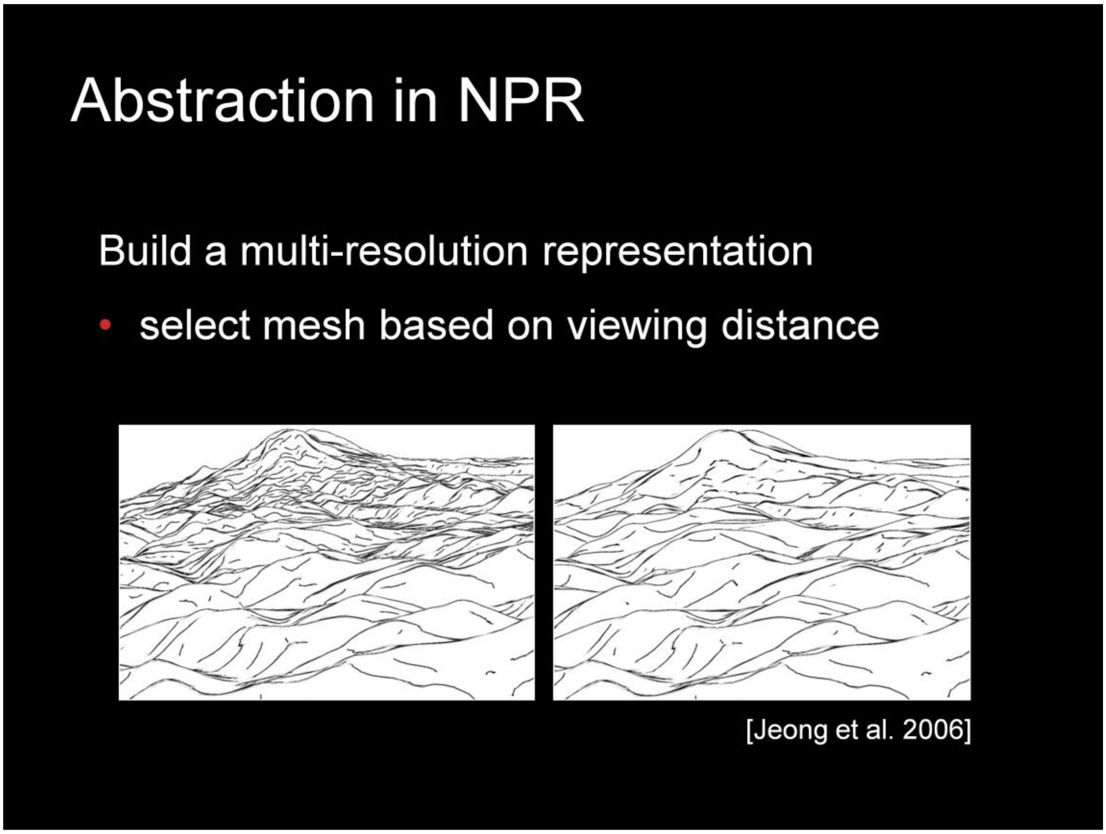
Working with 3D scenes presents a new problem
since the same objects can be viewed at various distances
by simply moving the camera.

Sometimes important information just can't be included
without causing clutter.

Thus, one approach is to explicitly measure
the density of lines in a potential rendering.

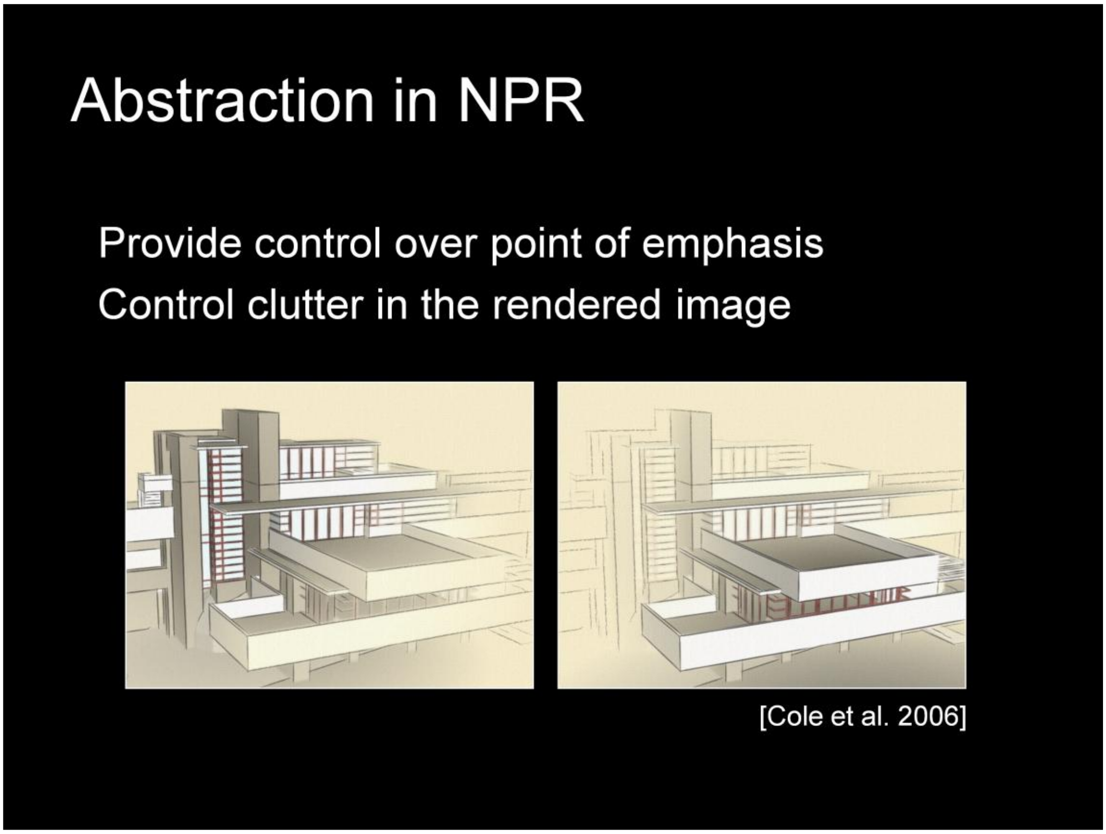
Some approaches work by using models of clutter effectively,
alongside methods for prioritizing content.

However, it's difficult to get this working in animation without
introducing temporal artifacts.



Another possible approach echoes work on level-of-detail.
Build a multi-resolution representation of a scene,
and select the scene elements based on viewing distance.

This simple approach **avoids** clutter,
although without any real guarantee.



It’s possible to combine both of these ideas,
which controls emphasis and manages clutter.

Forrester will be going into the details of such a system.

Here is an example from my work with Anthony Santella.

A photograph is transformed into a stylized version
which consists of black lines
and uniformly colored regions.

The interaction with the user is minimal:
they simply look at the photograph for a short period of time.

A recording of the user’s eye movements
provides the information required
to perform meaningful abstraction.

For the rest of this talk,
I’ll be explaining why this is a reasonable approach,
and how we evaluated the effectiveness of this system.

It’s all about how our attention shifts and how our eyes move.

Our eyes are constantly moving.

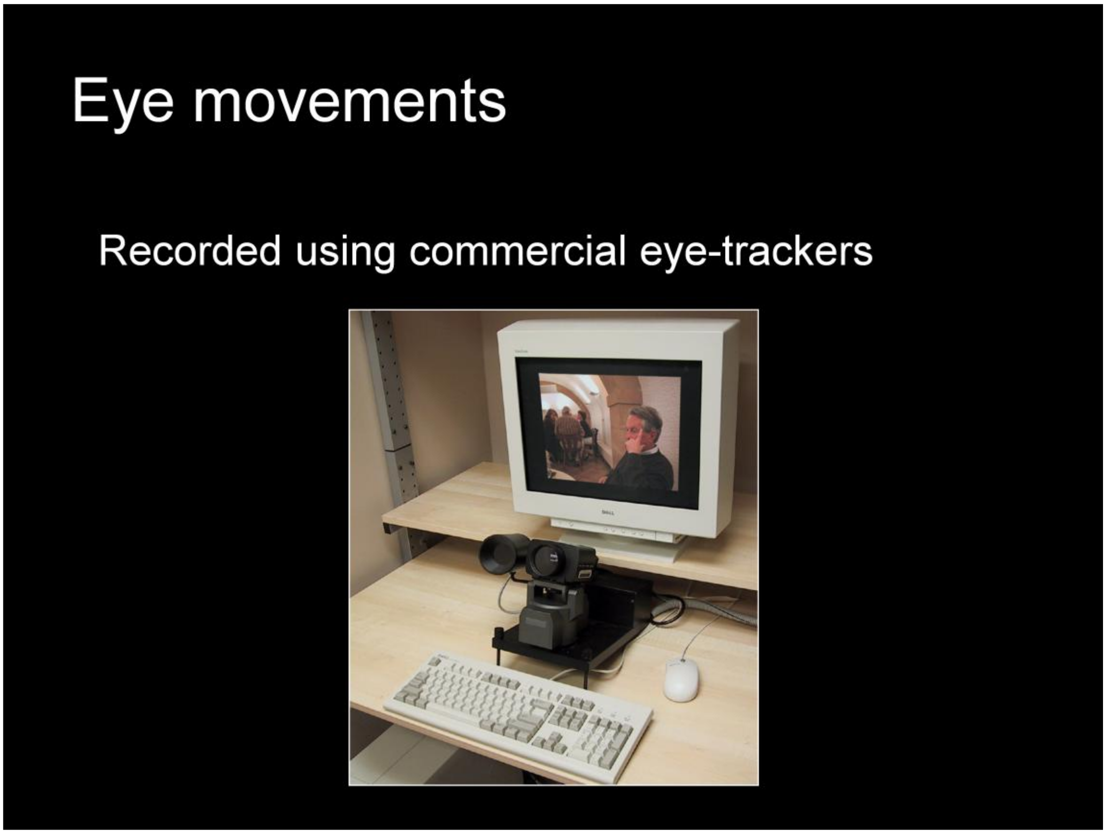
****

Here is an example recording.

Several times each second our eyes undergo
rapid motions known as saccades.

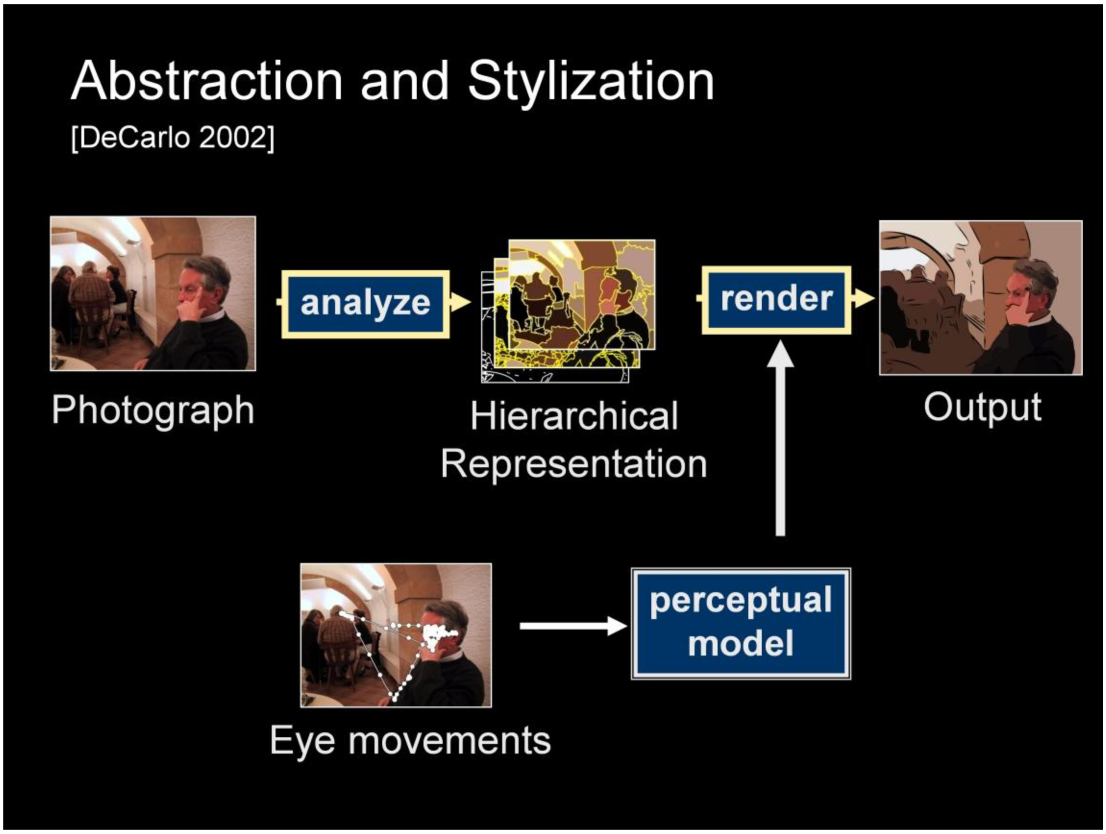
These are punctuated with stabilizing motions
known as fixations,
where our eyes are held fixed over a particular location.

It has been demonstrated
through a range of psychological studies
that longer fixations indicate INTEREST on the part of the viewer.



We can record eye movements using commercial eye trackers, such as this one.

So we can indirectly record
the content of interest for a particular viewer.

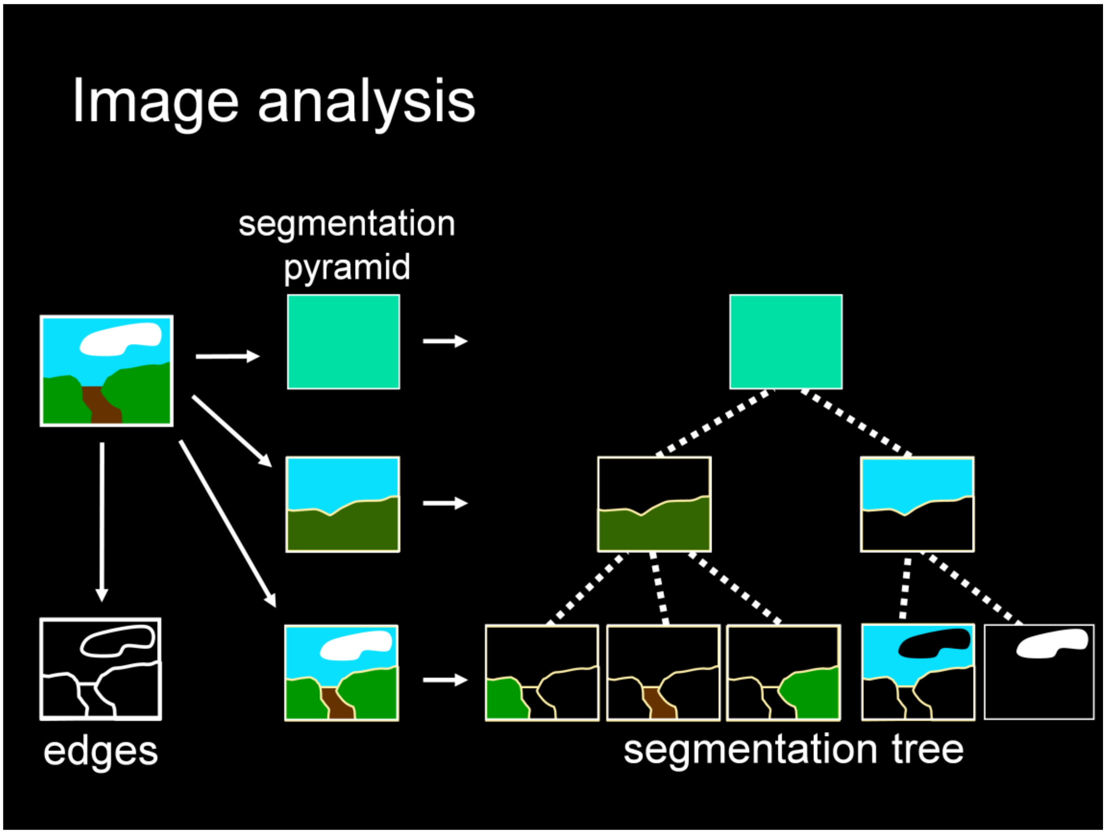


Our system starts with photograph,
****

decomposes it into a hierarchy of visual elements,
****

and renders a subset of these elements into an output image.
****

The features to render are selected by a perceptual model that draws upon a recording of a viewer's eye movements.



Our image analysis starts with edge detection,

****

and a set of image segmentations performed
at a range of resolutions.

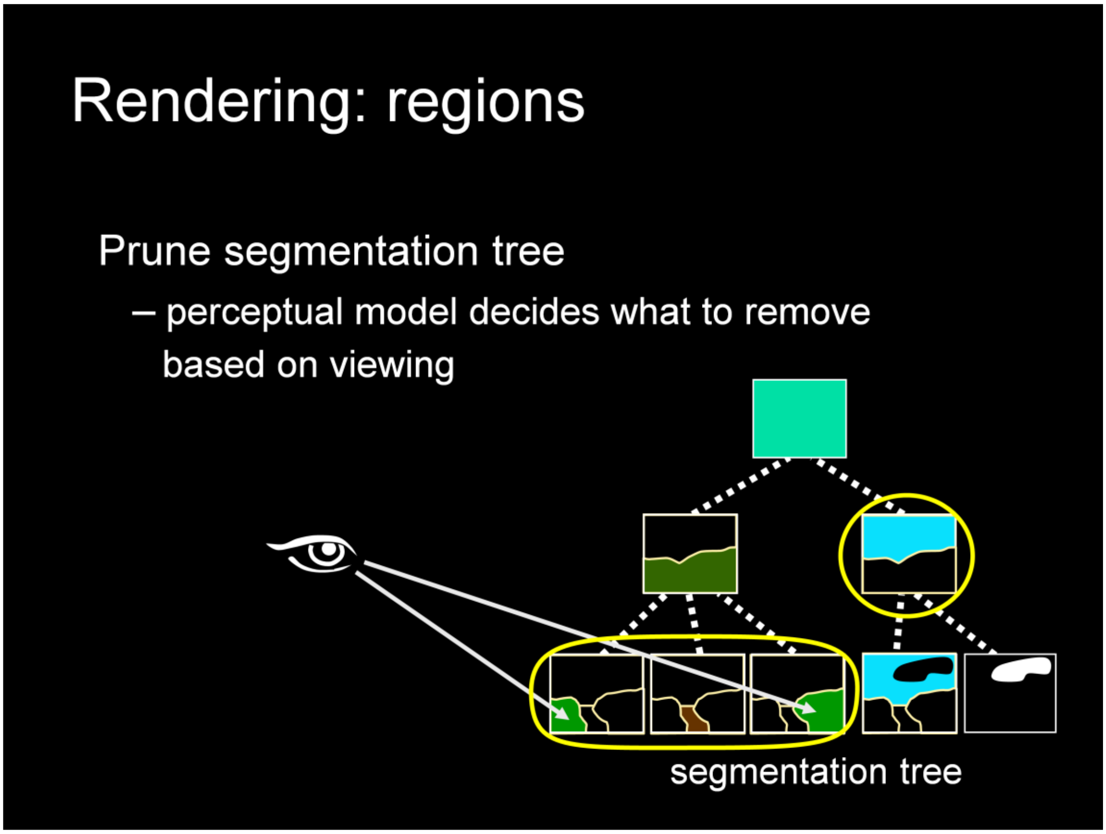
****

Finer scale segmentations contain more detail.

****

We build a hierarchical segmentation
by inferring containment relationships
between regions across resolutions.

This is our image representation.

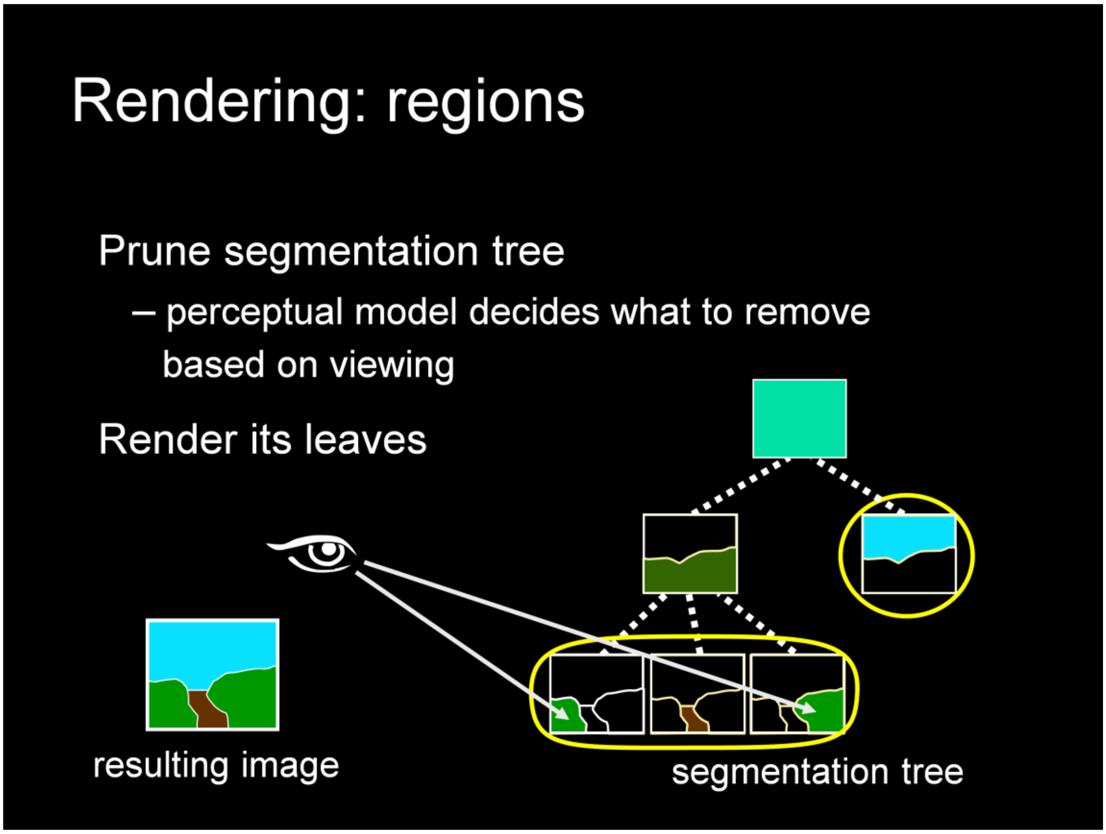


We then can prune the segmentation tree ****
based on predictions made by the perceptual model.

The perceptual model decides which of these visual elements will be included in the result.

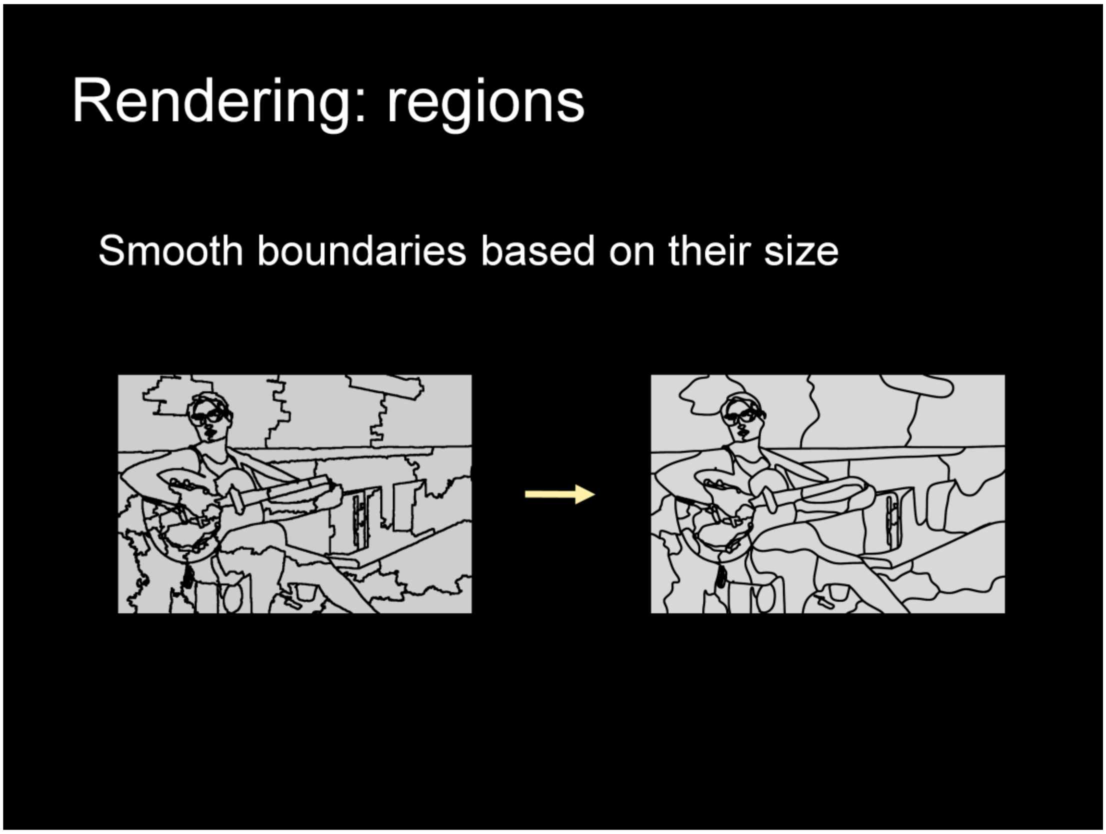
It does this based on the size and local contrast of each visual element, and for how long it was fixated.

In short, noticeable regions are more likely to be preserved, particularly when the viewer examined it closely.



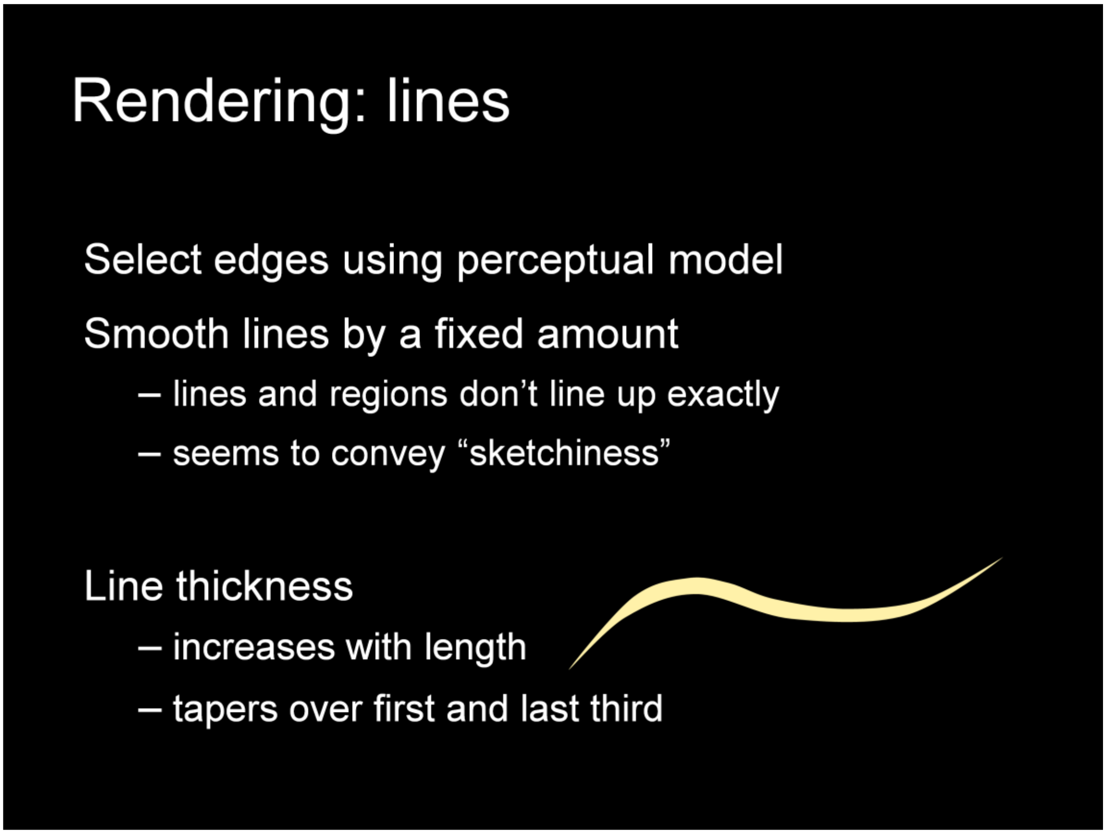
So anything that the viewer probably didn’t notice will be removed.

We render the leaves of the pruned tree into the output image.



We also smooth these regions
so that the detail in their boundaries reflect the appropriate scale.

Larger regions are smoothed more.



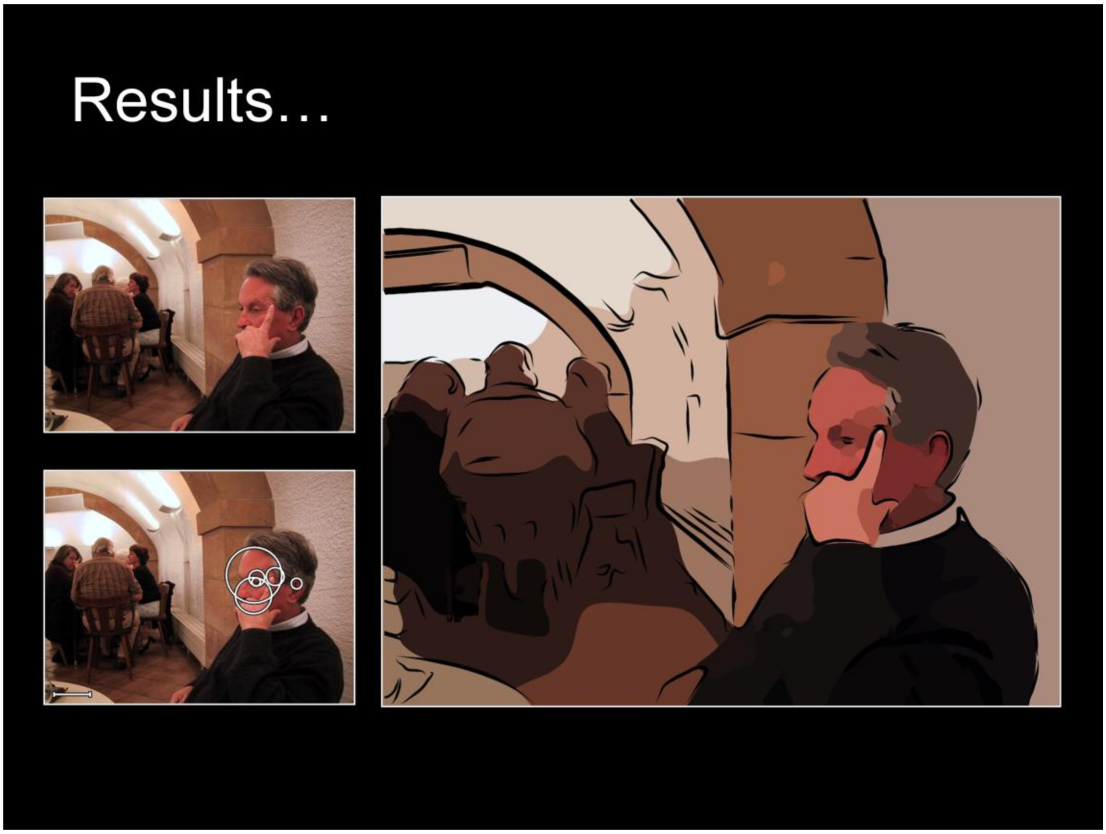
The lines are also selected using a perceptual model.

They are smoothed by a fixed amount.

This means that they won’t always line up with the regions.

In areas where the regions are heavily simplified,
this misalignment will be larger.

The result of this is a sketchy look where detail was removed.

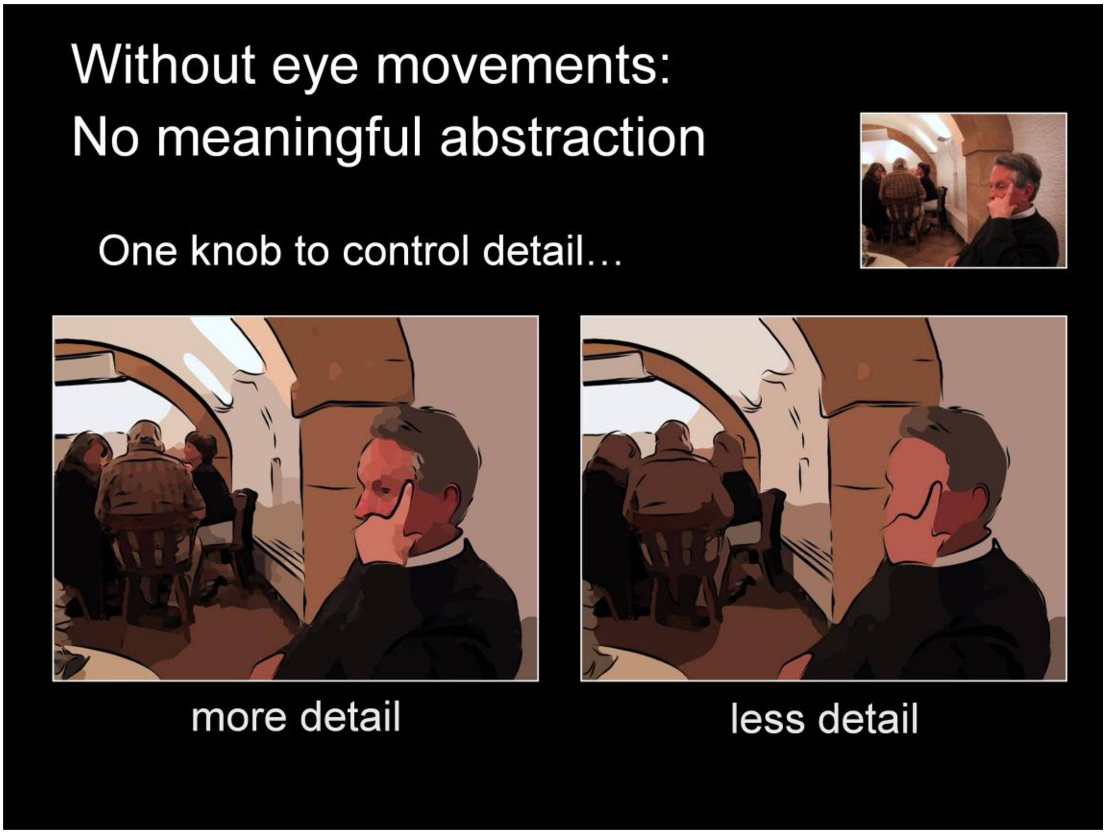


The final picture is made
by overlaying the lines on top of the regions.

Here, the small photograph on top is the original,
below are the users fixations, each white circle is a fixation,
where its size indicates fixation length.
The scale at the lower left is one second.

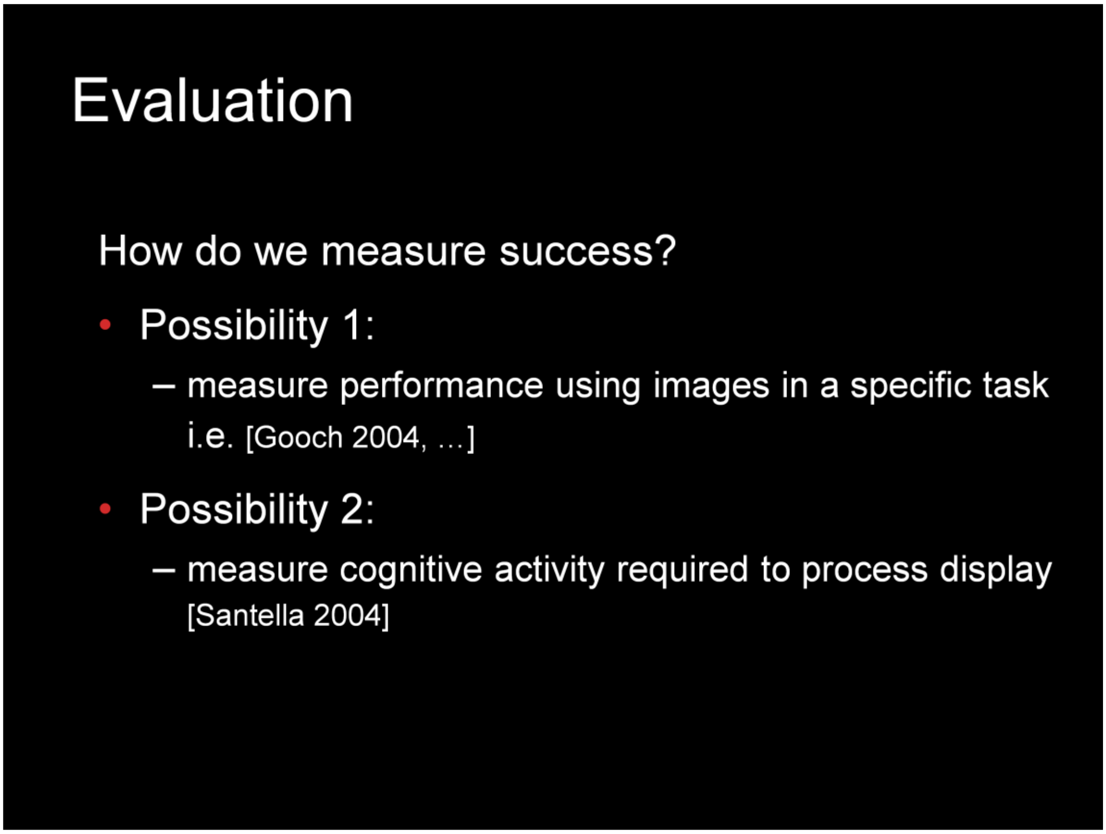
The foreground figure is clear,
as the viewer examined that location.

Figures in the background have had most of their detail removed.



You can compare this to results
where we don’t use eye movements,
but instead a global control for detail.

With high detail, the background looks distractingly fussy.
With less detail, important features such as the face are lost.



We can assess whether these images
achieve meaningful abstraction.

One approach to evaluating imagery
is to measure performance in a particular task.

Another is to measure the activity or effort required of the user.

A commonly used approach in evaluating an interface
measures how much the user must move the mouse
while performing some task.
It’s not a performance measure,
but rather an indirect measure of effort.

We take this approach,
and a natural activity to measure is eye movements.

As briefly mentioned before,
we know eye movements reflect viewers interests and goals.

Because of this link to cognition,
they’ve been used in the past to evaluate
complicated visual displays that must provide
efficient access to information.

Evaluation via eye tracking

Expect viewers to concentrate interest on emphasized regions in our renderings

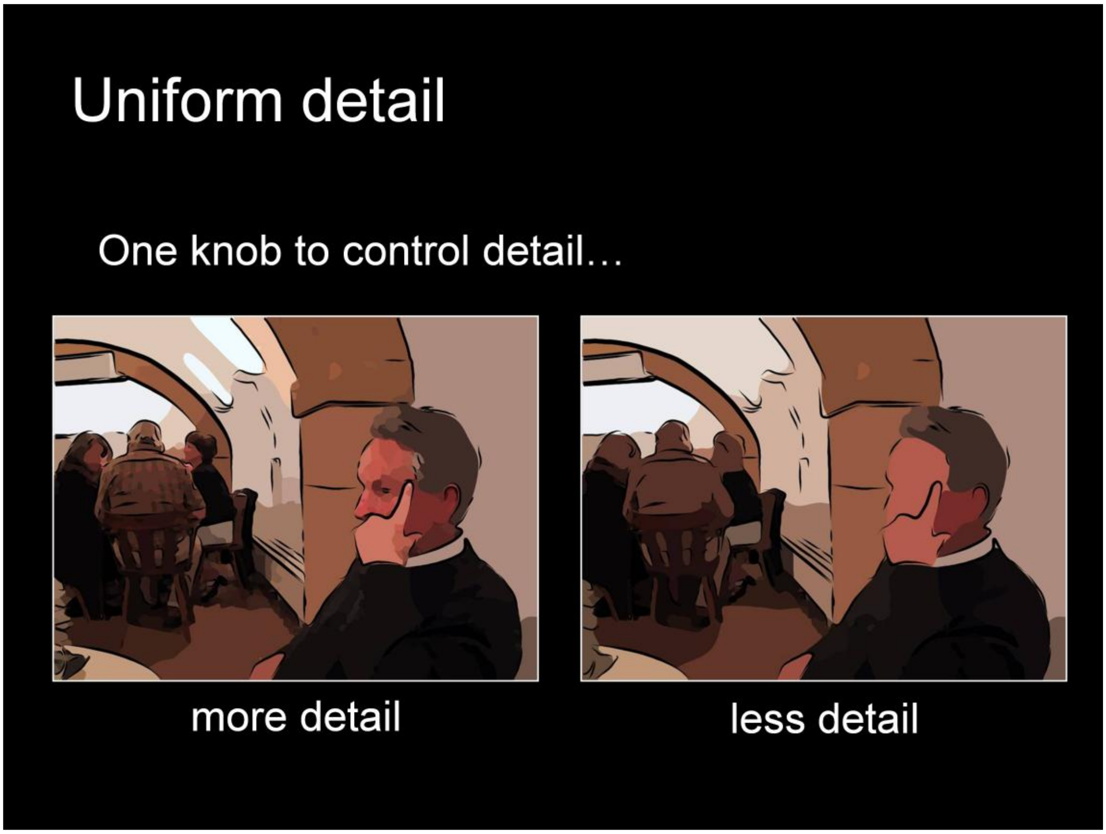
Measure and compare fixations on viewing:

- original image
- what else?

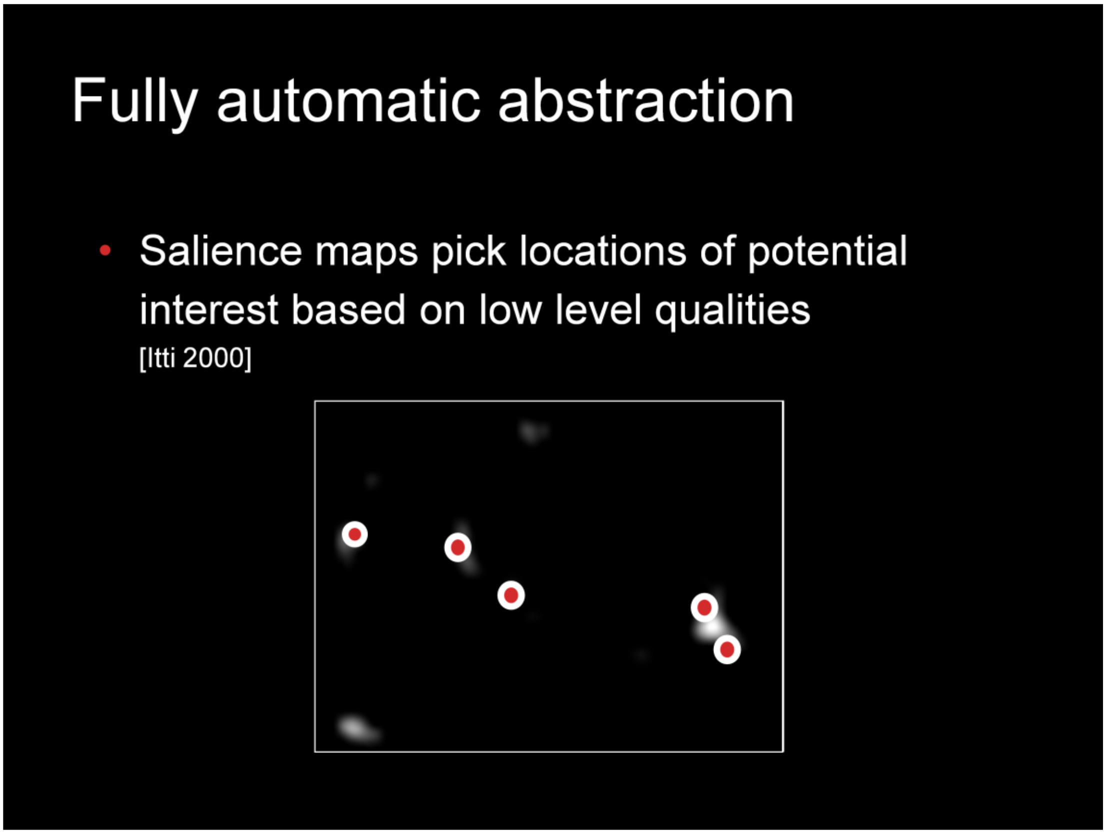
We hope to find that, for our images,
viewers concentrate more on the areas
that were highlighted with increased detail.

To test this we can compare fixations
over these images to those on the original photograph.

But there are some other interesting possibilities.



One possibility is uniform detail control,
which uses a global threshold
in place of eye movement recordings.



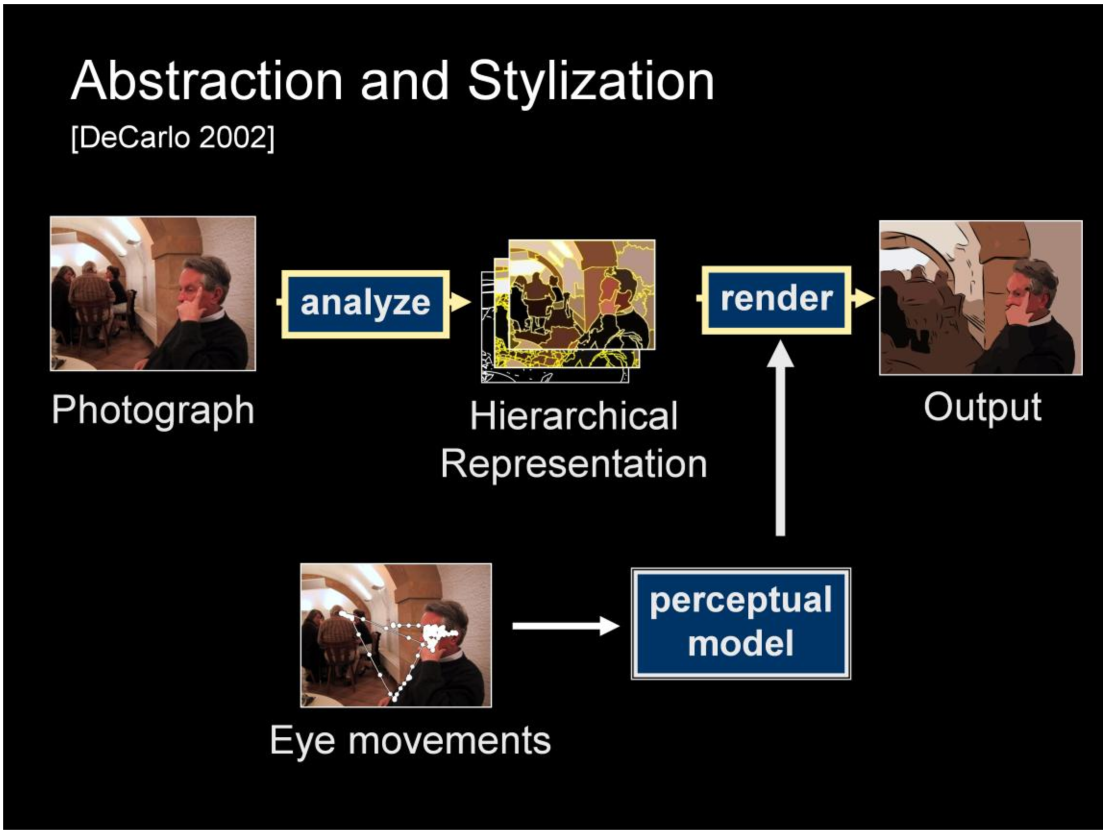
Another approach is to choose locations
for increased detail automatically.

Methods for predicting salience
combine a number of filters to create ****
a map of feature contrast for an image.

Bright areas are potentially interesting,
and algorithms can use them
to pick a set of locations completely automatically. ****

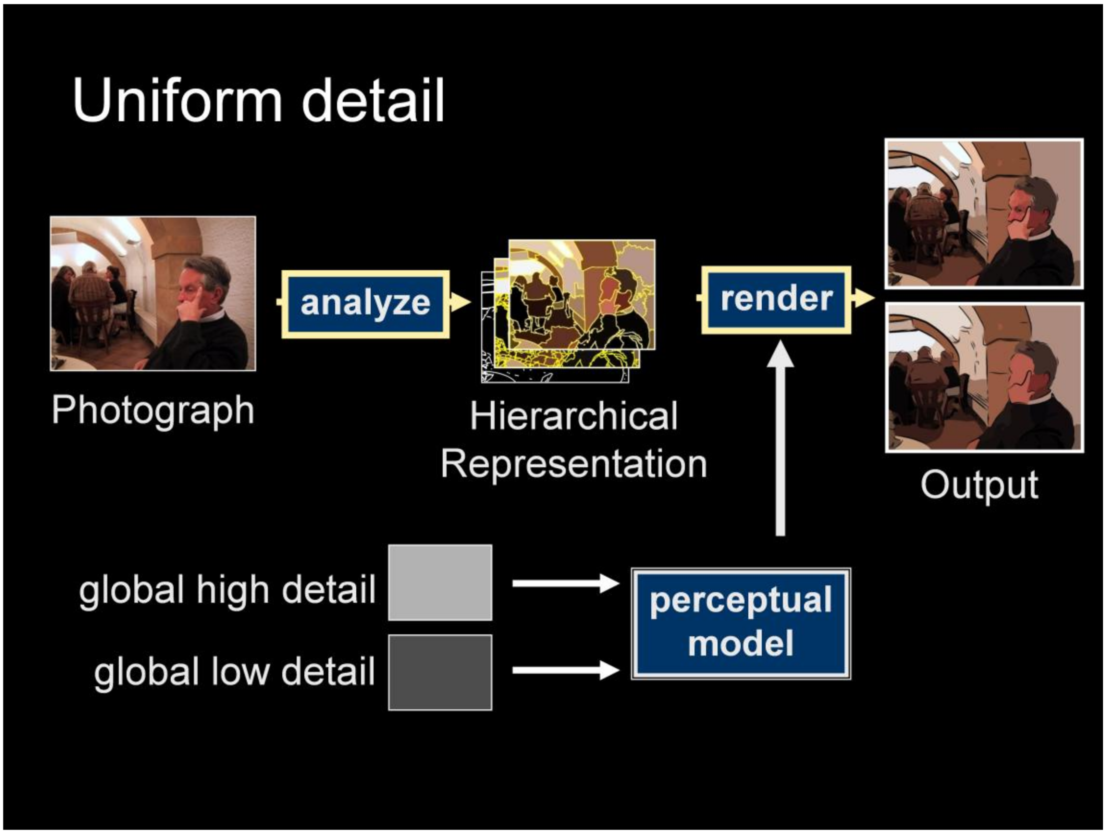
So like eye tracking, the output of the salience method
is a series of points
to be rendered with increased detail.

In the upcoming discussion, I'll refer to
both fixations and salience points used to control detail
as DETAIL POINTS.



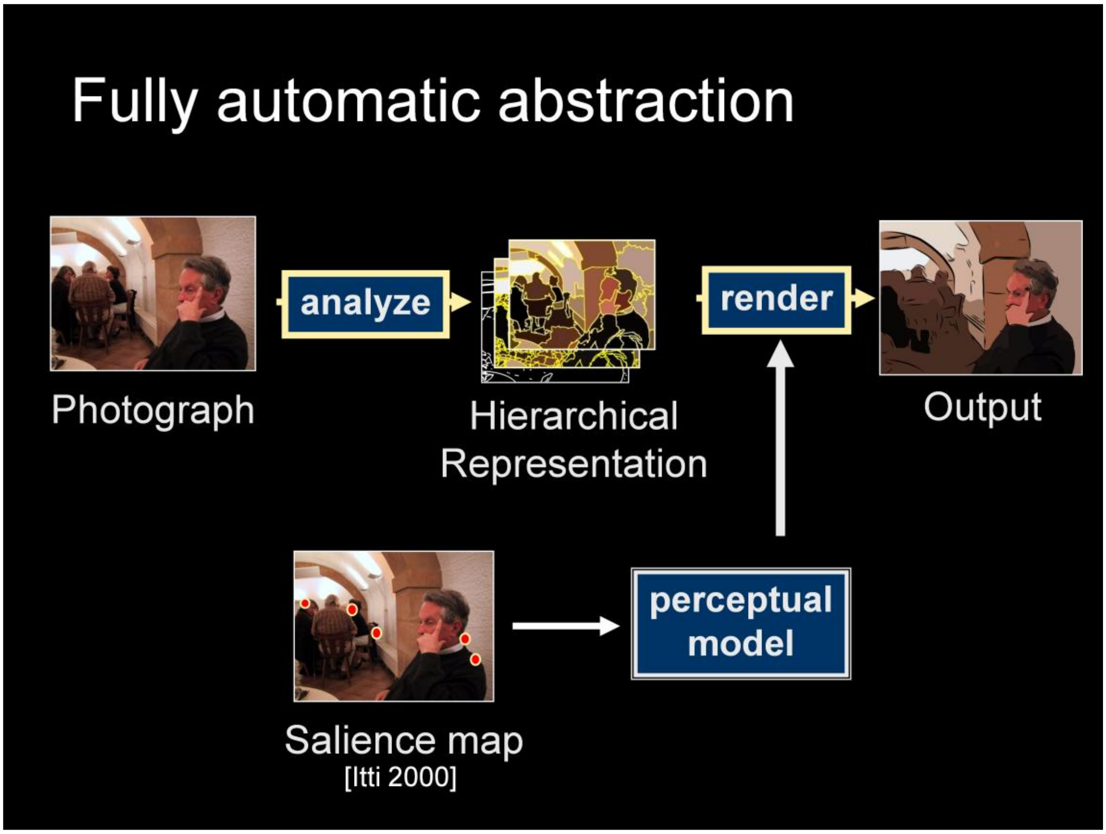
Here again is the design of our system.

Eye movements are input to the perceptual model.



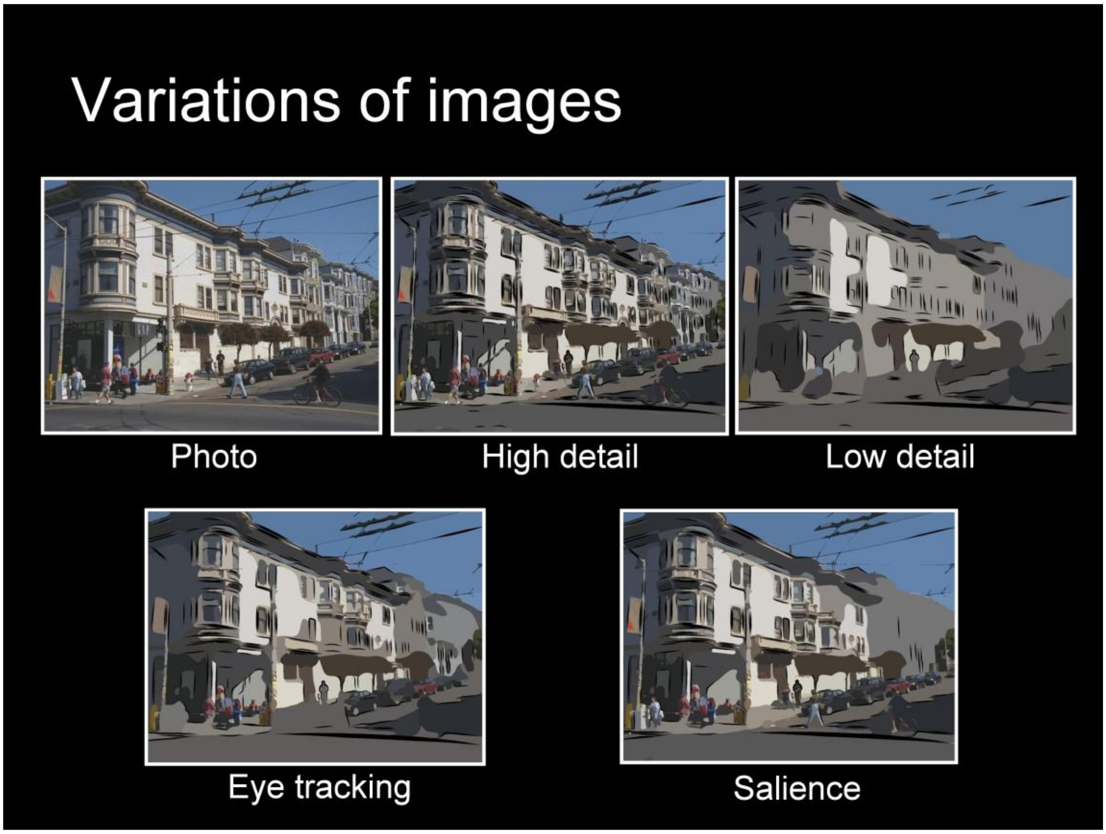
Uniform detail uses a global weight
to control detail across the image.

We do this for two levels of detail: low and high.



Finally, a model of salience
developed by Itti and Koch
determines the visually distinctive content.

This model selects a set of low-level image features
that might catch your attention.



Here are the five conditions.

We show each subject in our experiment one of these pictures.

We do this for 50 different photographs.

Variations of images

This set of images separates:

- the visual style
- global level of detail
- locally increased detail
- the locations of increased detail

These five images will let us distinguish between different hypothesis regarding how people examine images with modulated style and detail.

Specifically, we can analyze the effects of style, and of controlling detail both globally and locally.

The result is data like this.

This combines 10 viewers of each image together for 4 of the conditions.

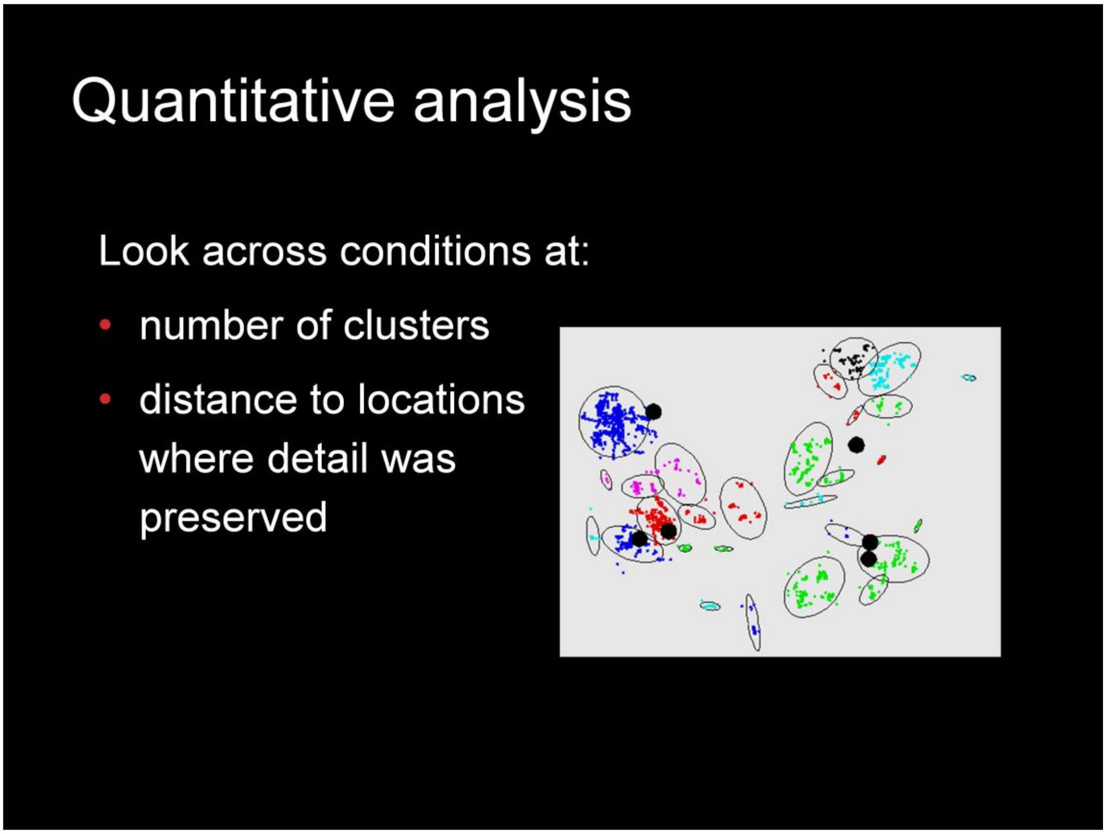
We also analyzed the data individually for each viewer, but here I'll only talk about results of analysis collapsed over viewers.

We can see differences in distribution of data across conditions.

But how can we quantify them?

Our approach is to cluster the data.

To divide it into limited regions of the image that were viewed coherently.



Once we’ve cut the data up into clusters,
we can compare the number of clusters,
and the distance from cluster centers to the detail points.

One set of detail points is marked here
with black circles.

Results:
Number of clusters

Both *eye tracking* and *salience* have significantly fewer clusters than *photo*, *high* and *low detail* (10-20% fewer)

- $p < 0.001 \ldots 0.05$

Eye tracking has significantly fewer clusters than *salience* (about 10% fewer)

- $p < 0.001$

When examining the number of clusters,
we found that modulating detail holds viewer interest moreso than uniform detail.

We found that viewers examine 10 to 20 percent fewer locations when the detail is modulated.

This effect is larger when the detail is meaningful.

Results:
Distance to detail points

- Clusters of interest are closer to the detail points when using *eye tracking* or *salience*
  - $p < 0.0001$

When measuring the distance from cluster centers to detail points, we can verify that people were in fact looking in the right places.

This might seem just like what you'd expect given what we know artists do.

But it wasn't clear that this is what we'd find.

It might have been that these techniques simply cannot capture how artists can guide our attention, as there are certainly many other tools that artists use to do this same thing.

Implications for NPR

Meaningful abstraction is important

- style alone is not enough
- global detail control is not enough
- low level salience is not enough

Use eye tracking to evaluate and understand NPR

In summary,

we find that in achieving meaningful abstraction,

altering the style is not enough,
globally changing the detail is not enough,
and using heuristic measures of importance is not enough.

We need to locally adapt the detail in a meaningful way.

We're also encouraged by the use of eye-tracking
to evaluate the effectiveness of NPR displays.

Summary

Abstraction

- omitting and adapting content
- get importance information from a user/artist

Evaluation

- eye movements provide one way of assessing abstract displays