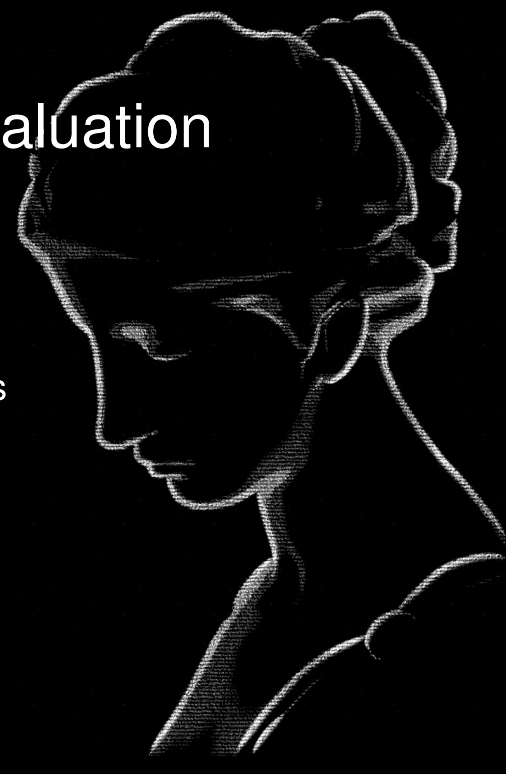
Part VI: Abstraction and Evaluation



Doug DeCarlo

Line Drawings from 3D Models
SIGGRAPH 2005 Course Notes

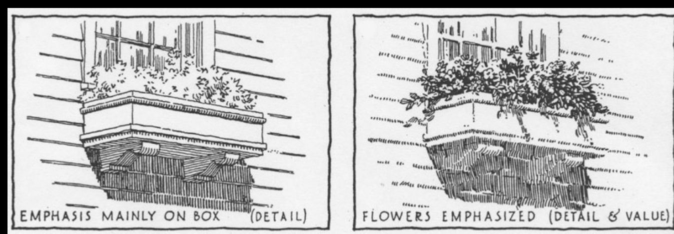
Abstraction

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

Artists can design effective imagery by altering 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 is a tool for effective visual communication.

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

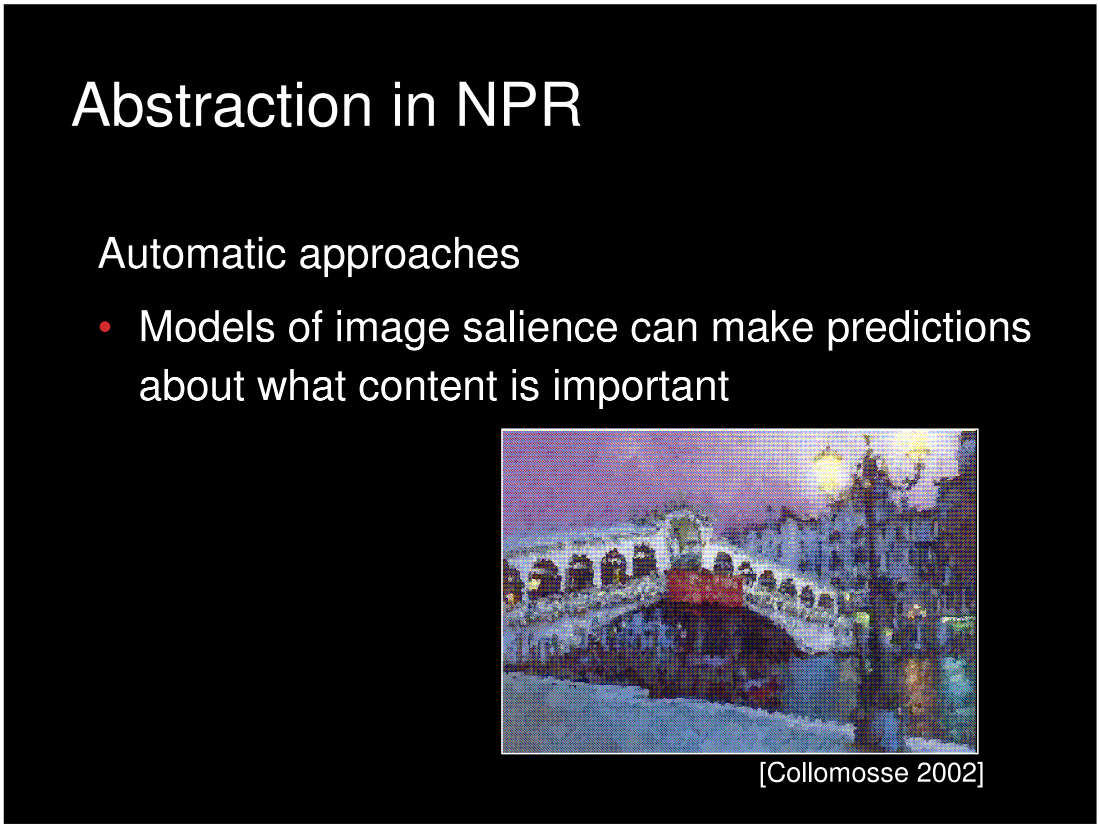
Abstraction



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

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.



How does this work in NPR? There are a range of techniques. The most important difference is in 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.

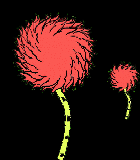
For painterly rendering, a fully automatic approach might attempt to preserve important content as determined by a computational model that predicts image salience.

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

Abstraction in NPR

Rendering specific content: trees

– automatically leave out lines in the center of the tree



[Kowalski 1999]

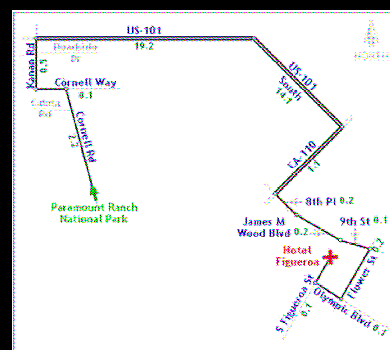
[Deussen 2000]

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

Rendering specific content: route maps

– constraints encode what can be omitted and how it can be distorted while still preserving the content necessary to follow the route



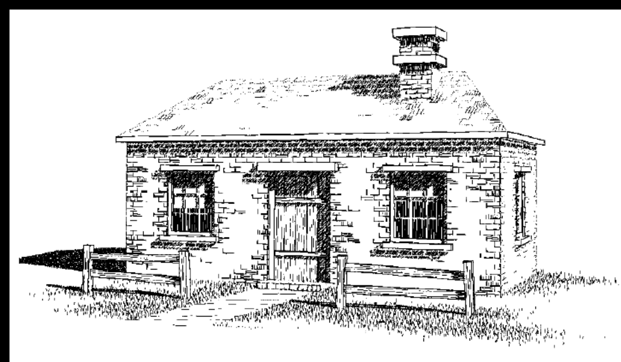
[Agrawala 2002]

Here's another example—constructing maps for driving from one location to another.

The LineDrive system, which is available on www.mapblast.com embodies the cognitive principles that describe how people effectively use such maps. The system can make sound decisions regarding what content can be removed, and how roads can be distorted or simplified, while ensuring the user's interpretation of the map is preserved.

Abstraction in NPR

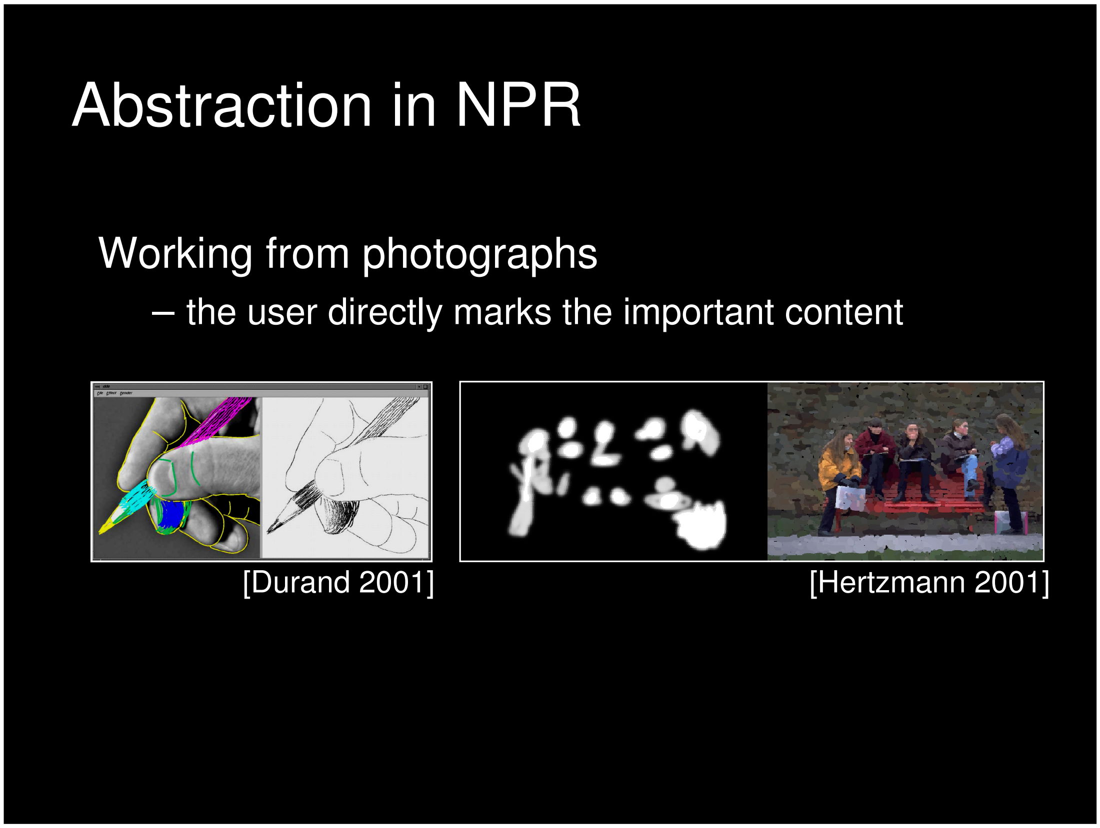
Indication in pen and ink illustration
- the user specified what content was important



[Winkenbach 1994]

In more general domains, a user is required to specify the important content.

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



The manual marking of photographs also produces effective artistic renderings.

Abstraction in NPR

Working from photographs

– infer important content from a user's eye movements

[DeCarlo 2002]

Finally, here is an example from my work with Anthony Santella. The 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 need to 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.

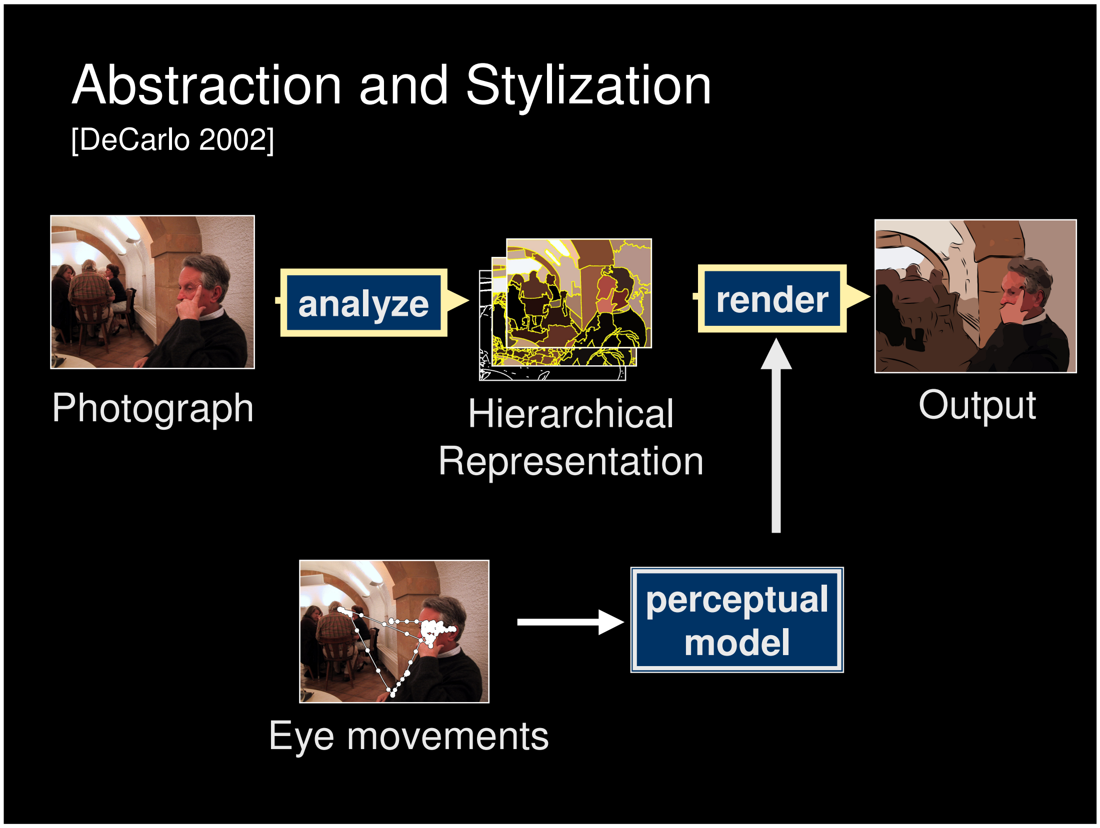
For the rest of this talk, I'll be explaining why this is a reasonable approach, and how we evaluated the effectiveness of our system.

Eye movements

Eyes dwell on particular locations during *fixations*

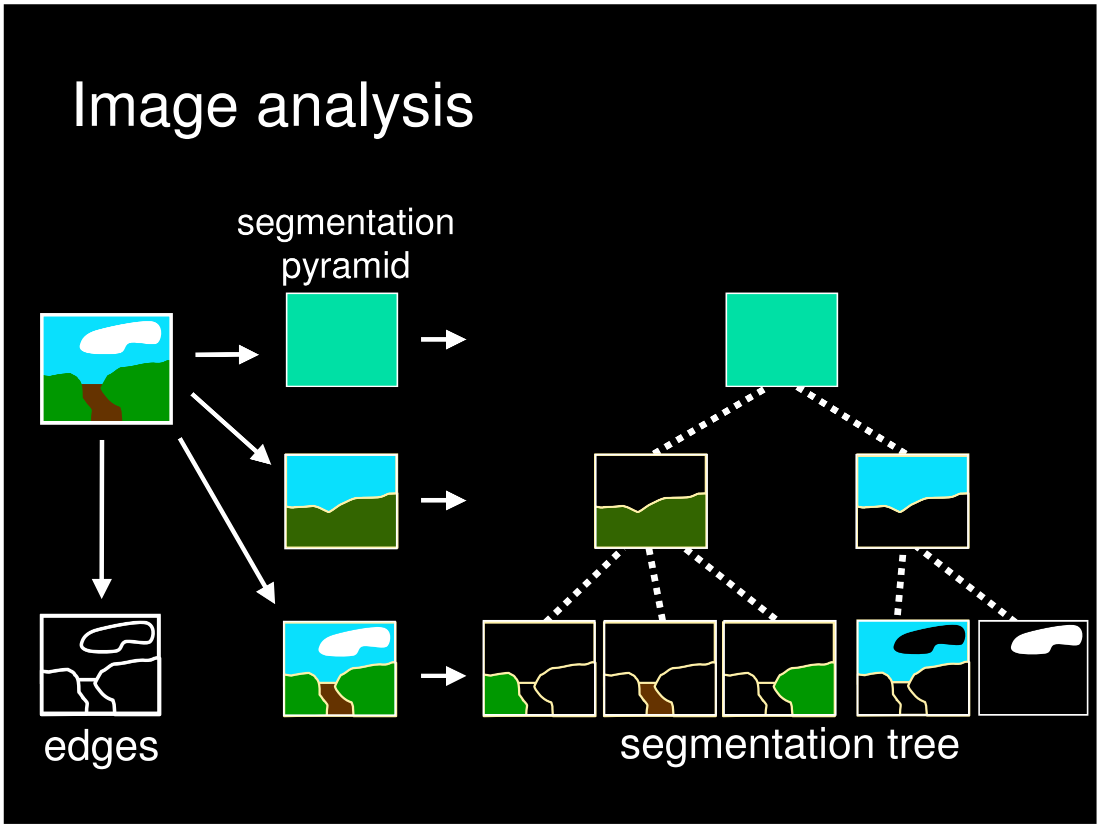
- Quick motions between these locations are made via *saccades*
- Longer fixations indicate viewer interest

Our eyes are constantly moving. A few 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 is has been demonstrated through a range of psychological studies that longer fixations indicate interest on the part of the viewer.



A system built with this in mind 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.

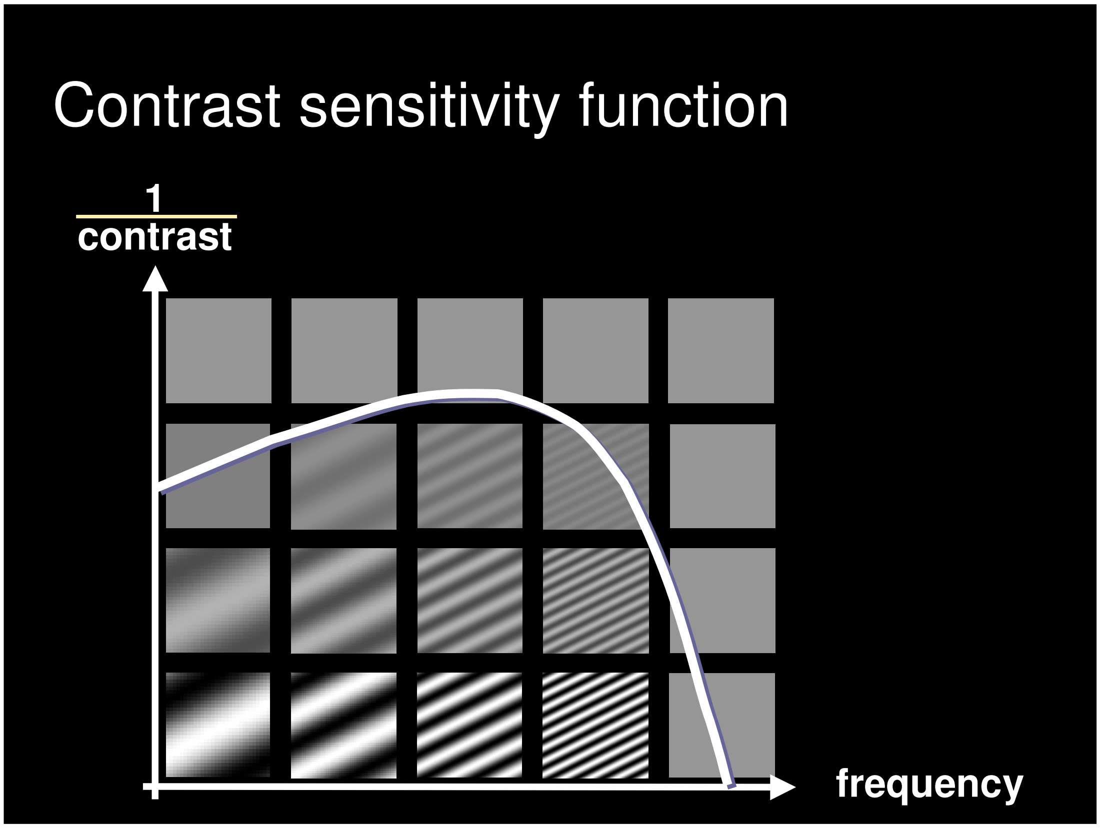
I'm going to give a quick overview of this system in the next few slides and then talk about how a more recent study where we've evaluated this system.



Our image analysis starts with an edge detection, and a set of image segmentations performed at a range of resolutions. Finer scale segmentations contain more detail.

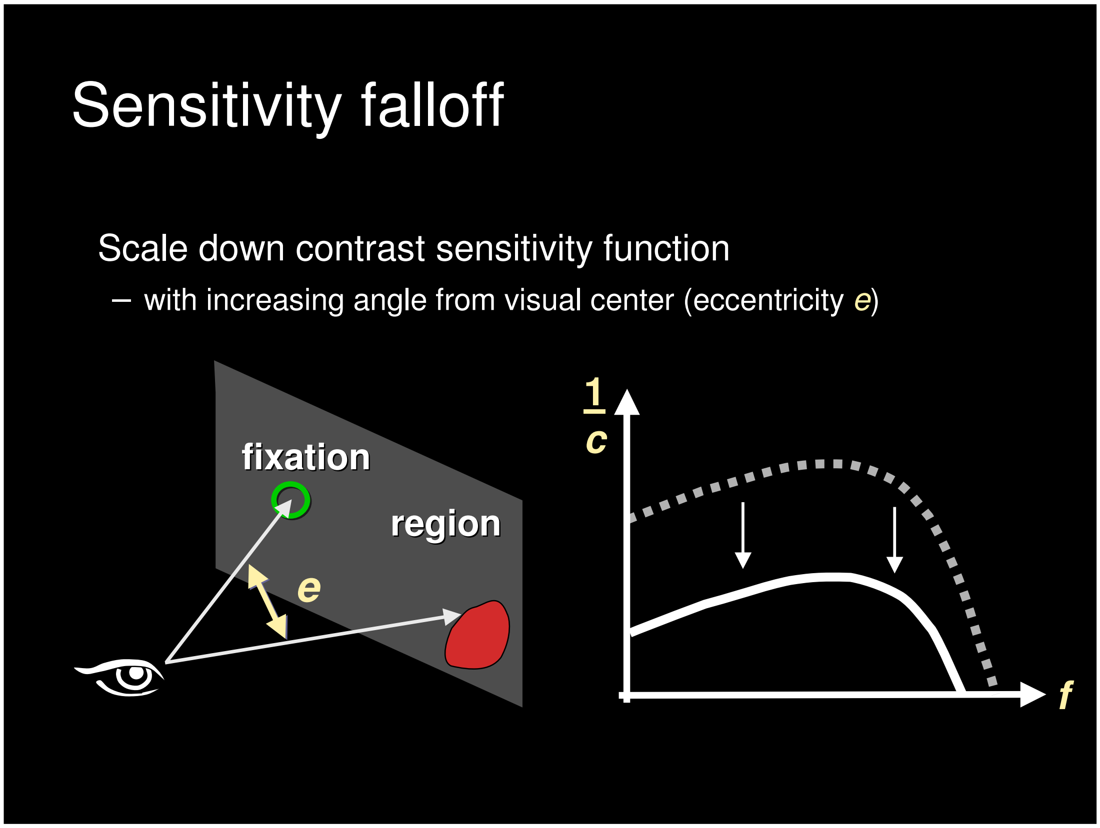
We build a hierarchical segmentation by inferring containment relationships between regions across resolutions.

This is our image representation. We now need to decide which of these visual elements will be included in the final output. This is where our perceptual model comes in.



Here’s a quick sketch of the perceptual model that’s used to do this. This model was constructed by showing people repeating patterns of lines of increasing frequency and decreasing contrast. If the frequency is too high or the contrast too low, the pattern appears as a solid color. The contrast sensitivity function is the curve that marks this change.

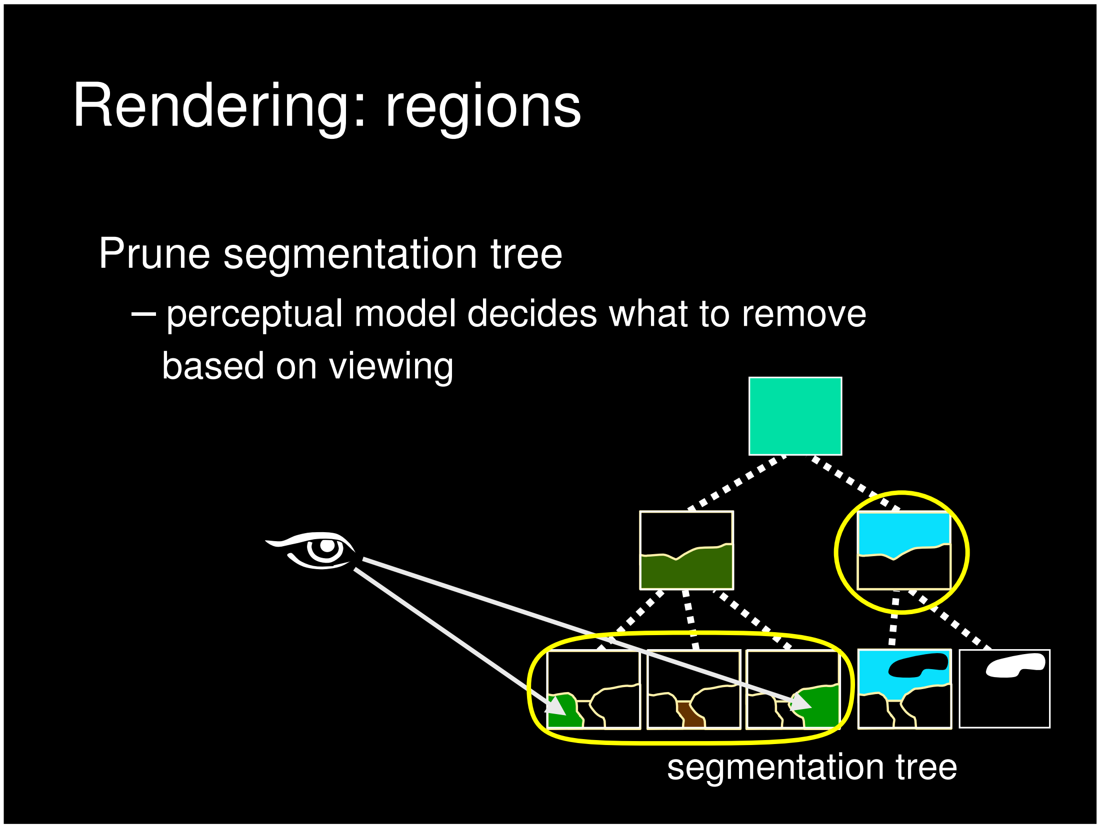
We can use this curve to judge the relative visibility of regions given their size and contrast with the background.



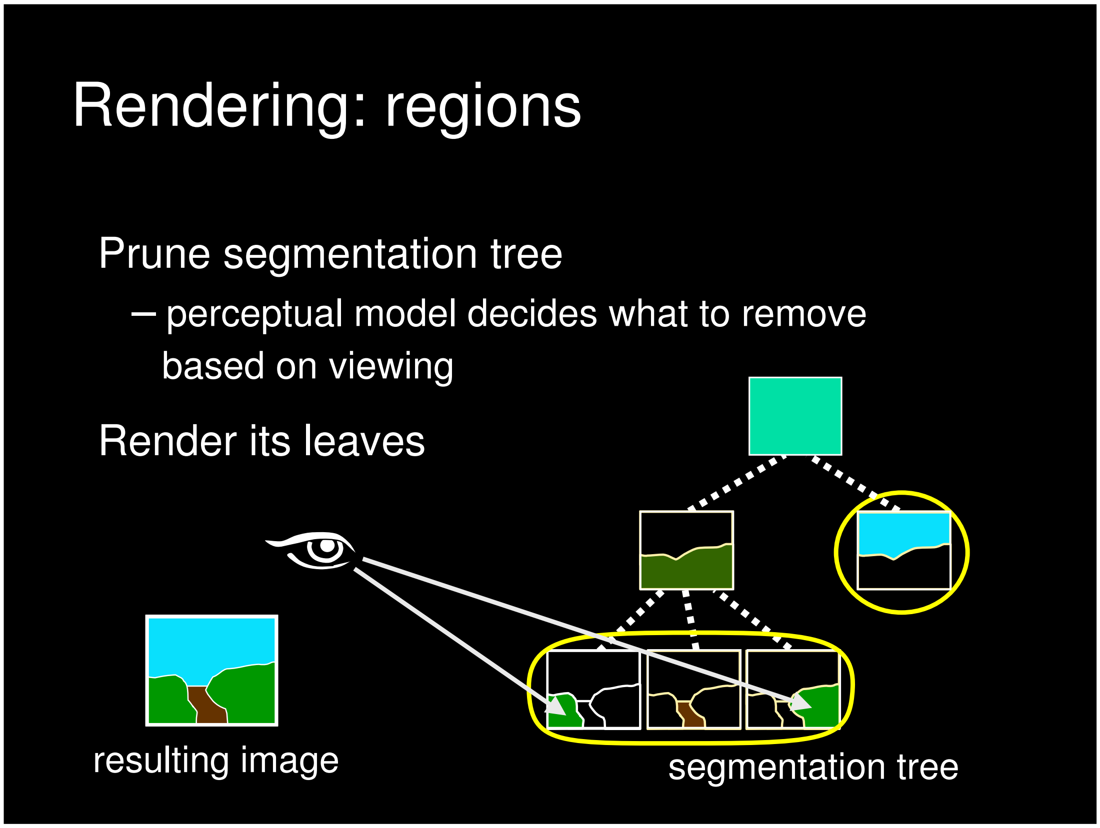
This model is used with the eye movement recordings as follows.

It uses another model that describes the fact that the further a feature is from where the viewer is looking, the harder it is to see. The angular distance between the fixation and the feature is called eccentricity. The contrast sensitivity curve is scaled down as a function of eccentricity, to model this drop-off.

So for each region, we compute its minimum eccentricity using the eye tracking data, and this gives us the appropriate contrast sensitivity function to apply.



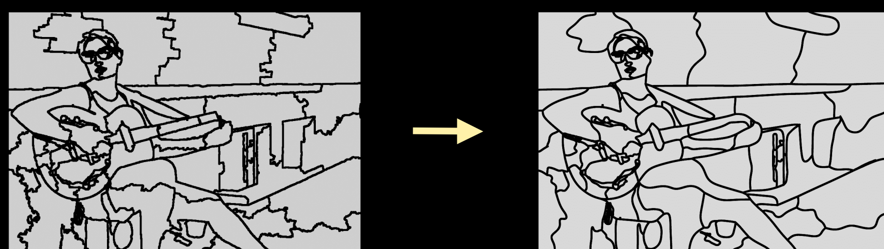
We then can prune the segmentation tree based on predictions made by the perceptual model. Anything that the viewer probably didn't notice will be removed.



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

Rendering: regions

Smooth boundaries based on their size



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

Rendering: lines

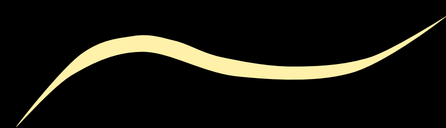
Select edges using perceptual model

Smooth lines by a fixed amount

- lines and regions don’t line up exactly
- seems to convey “sketchiness”

Line thickness

- increases with length
- tapers over first and last third

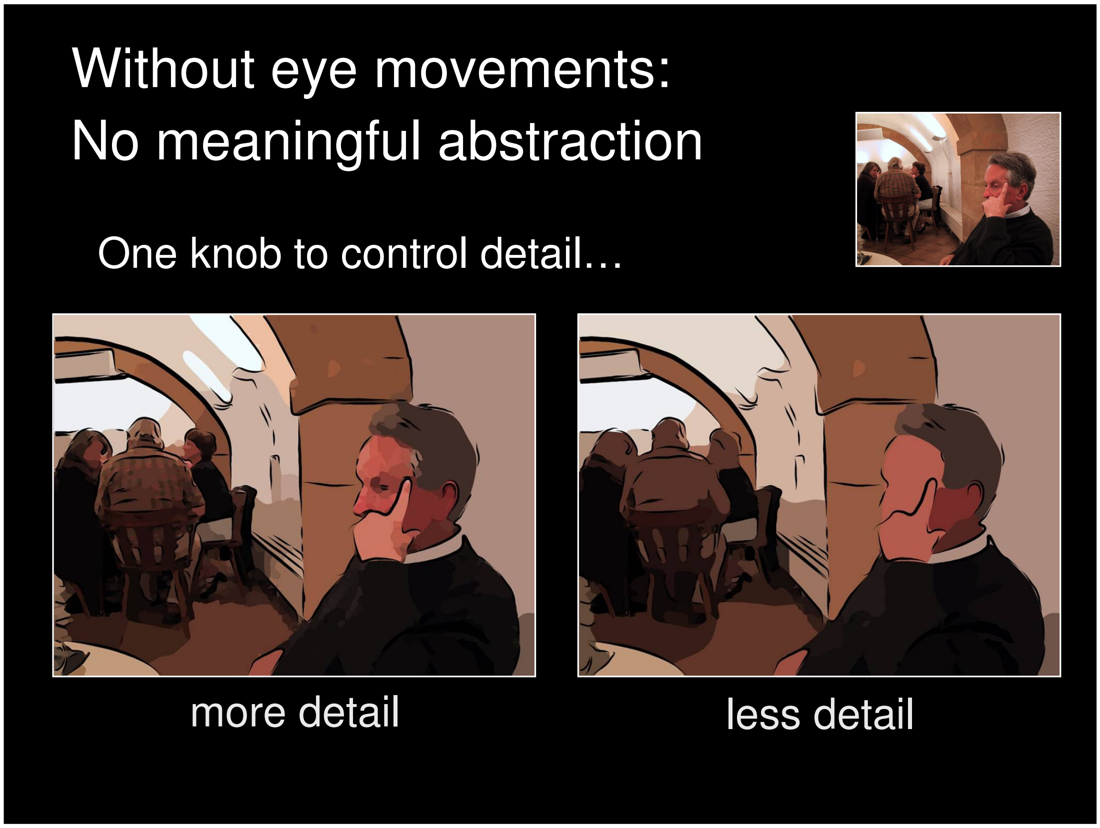


The lines are also selected using a perceptual model. They are smoothed by a fixed amount; this means that they won’t always line up with the regions. In areas where the regions are heavily simplified, this misalignment will be larger; the result of this is a sketchy look where detail was removed.

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, its size indicates fixation length. The scale at the lower left is one second.

Notice in the result, the foreground figure, where the viewer looked, is clear, while figures in the background have had most of their detail removed.



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

With high detail, the background looks distractingly fussy. While with less detail important features, like the face, are lost.

Evaluation

How do we measure success?

- Possibility 1:
  – measure performance using images in a specific task i.e. [Gooch 2004, …]
- Possibility 2:
  – measure of cognitive activity required to process the images [Santella 2004]

We can try and measure if 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. An example of this approach that is commonly used is evaluating an interface by how much user has to move the mouse while performing some task. It's not a performance measure but 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 have to provide efficient access to information. This includes everything from web pages to air traffic control displays.

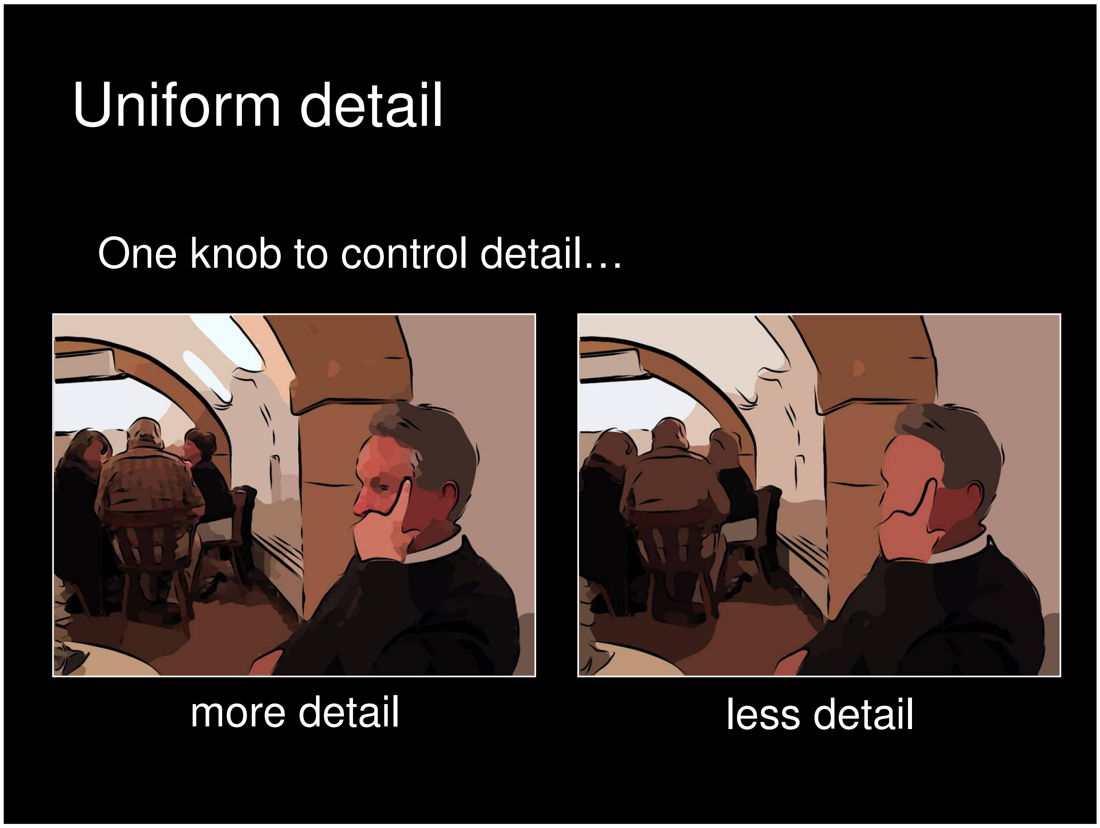
Evaluation via eye tracking

Expect viewers to concentrate interest on emphasized regions in our renderings

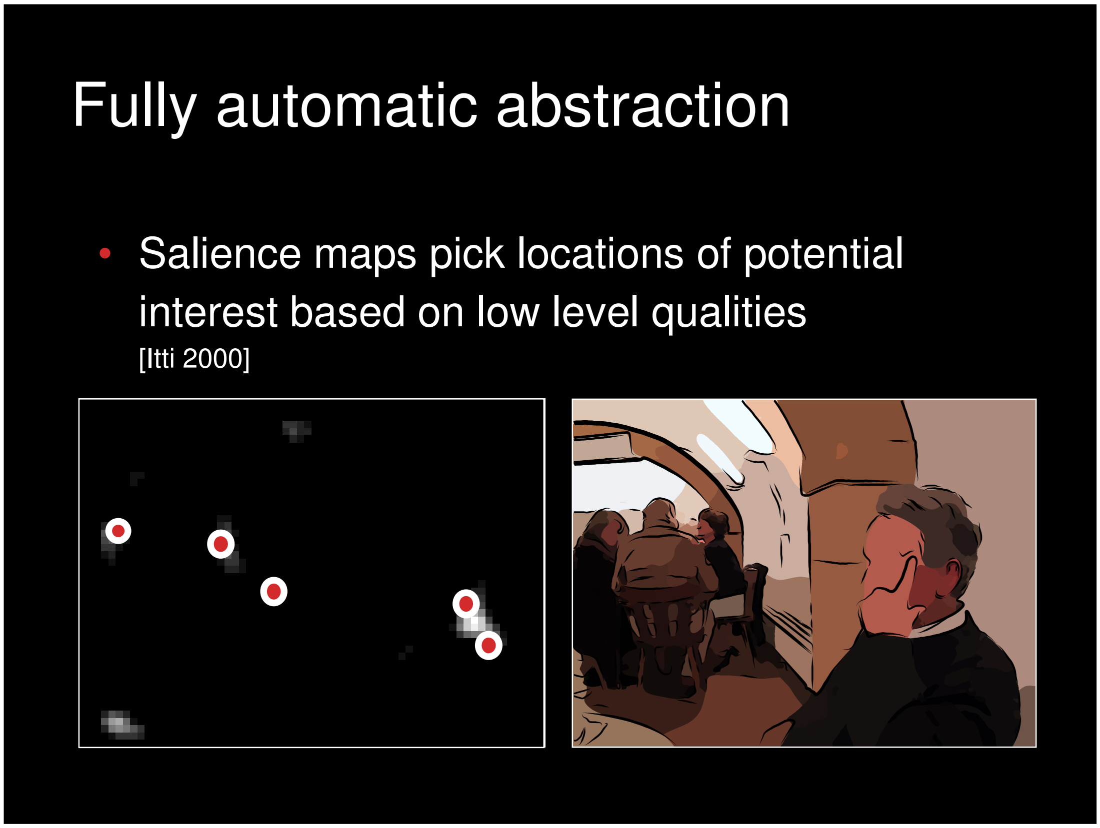
Measure and compare fixations on viewing:

- 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 photo. But there are some other interesting possibilities.



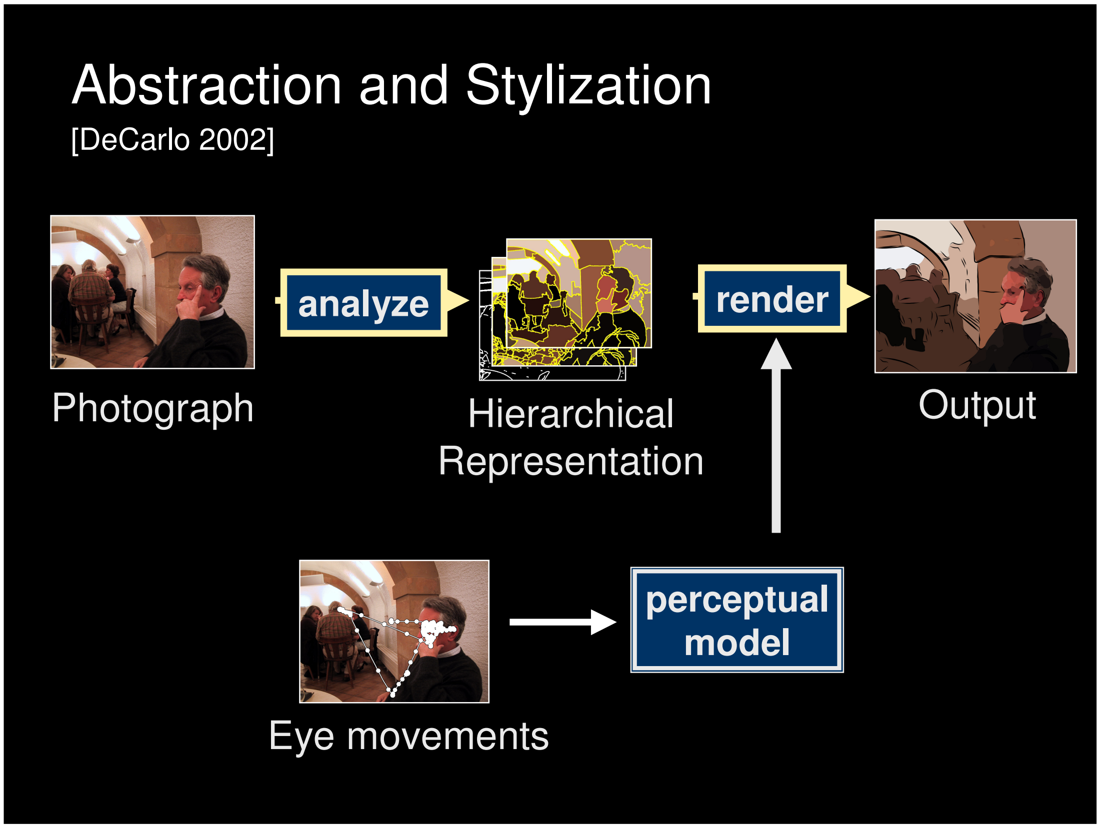
One possibility is uniform detail control, which use a global threshold in place of eye movement recordings.



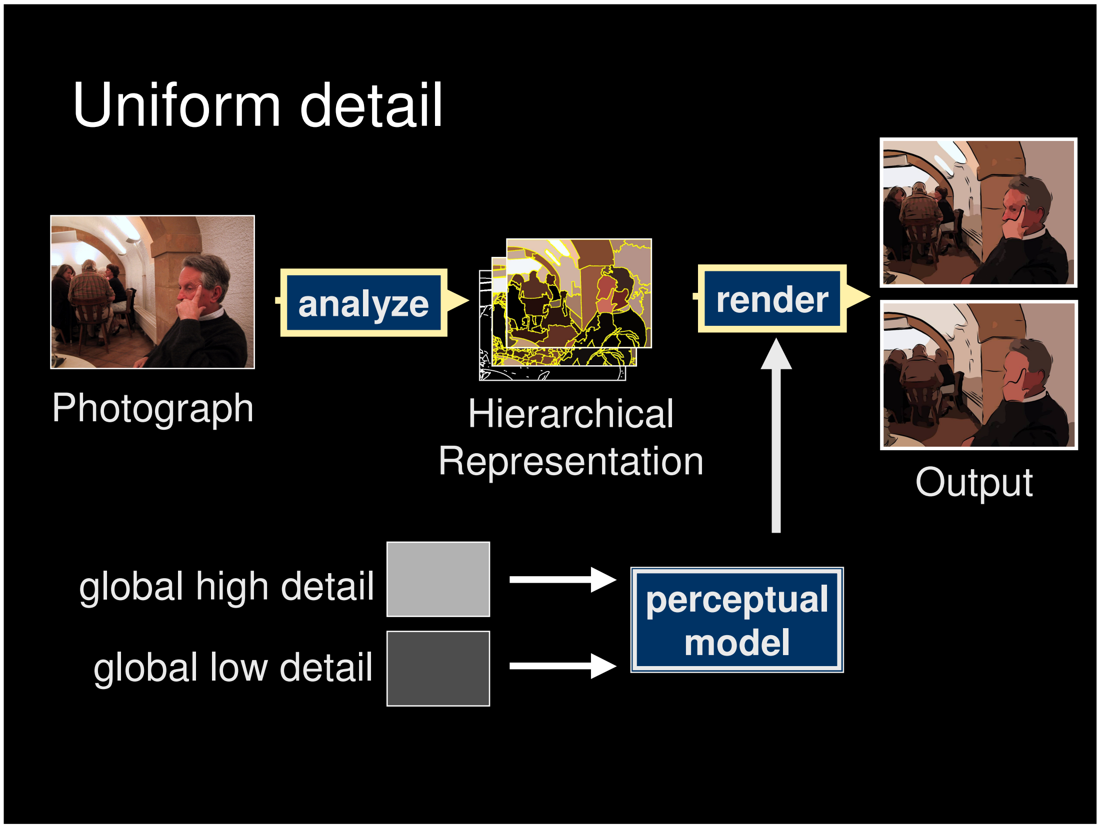
Another approach is to choose locations for increased detail automatically. Methods for predicting salience combine a number of filters to create a map of feature contrast for an image, like the one on the left. Bright areas are potentially interesting, and algorithms can use them to pick a set of locations completely automatically.

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

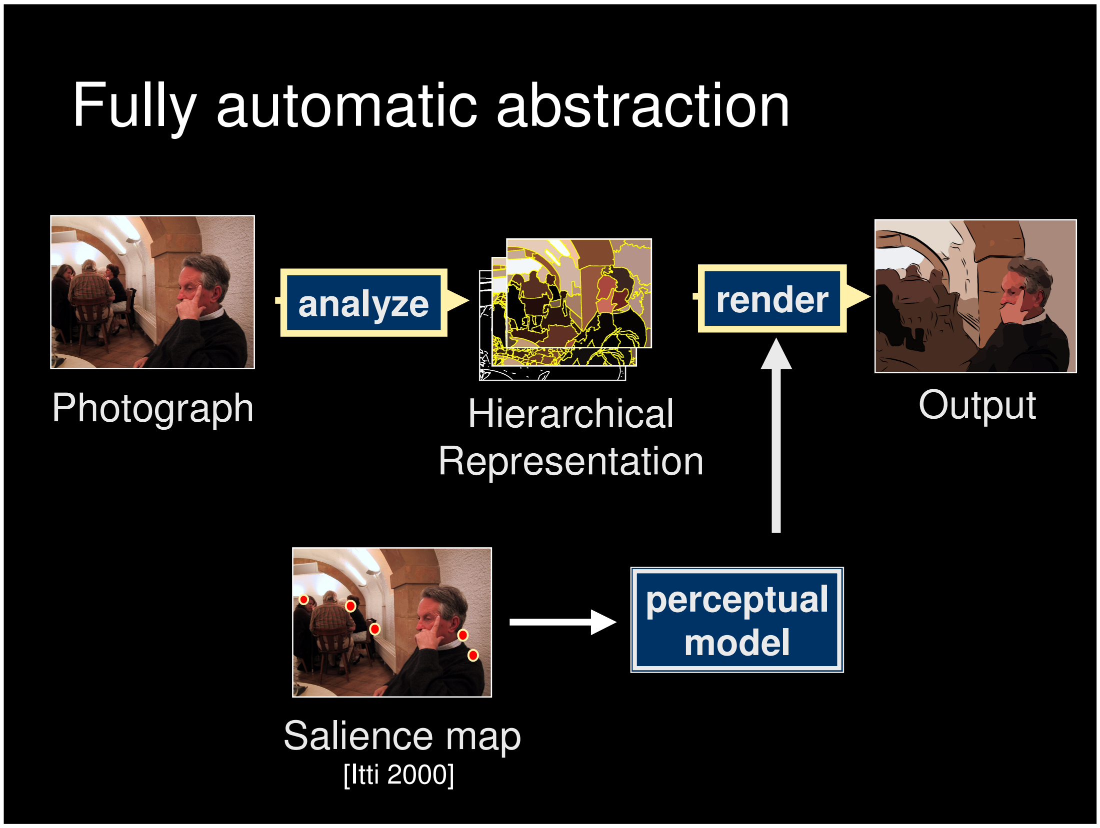
In the upcoming discussion, I'll refer to both fixations and salience points used to control detail as "detail points".



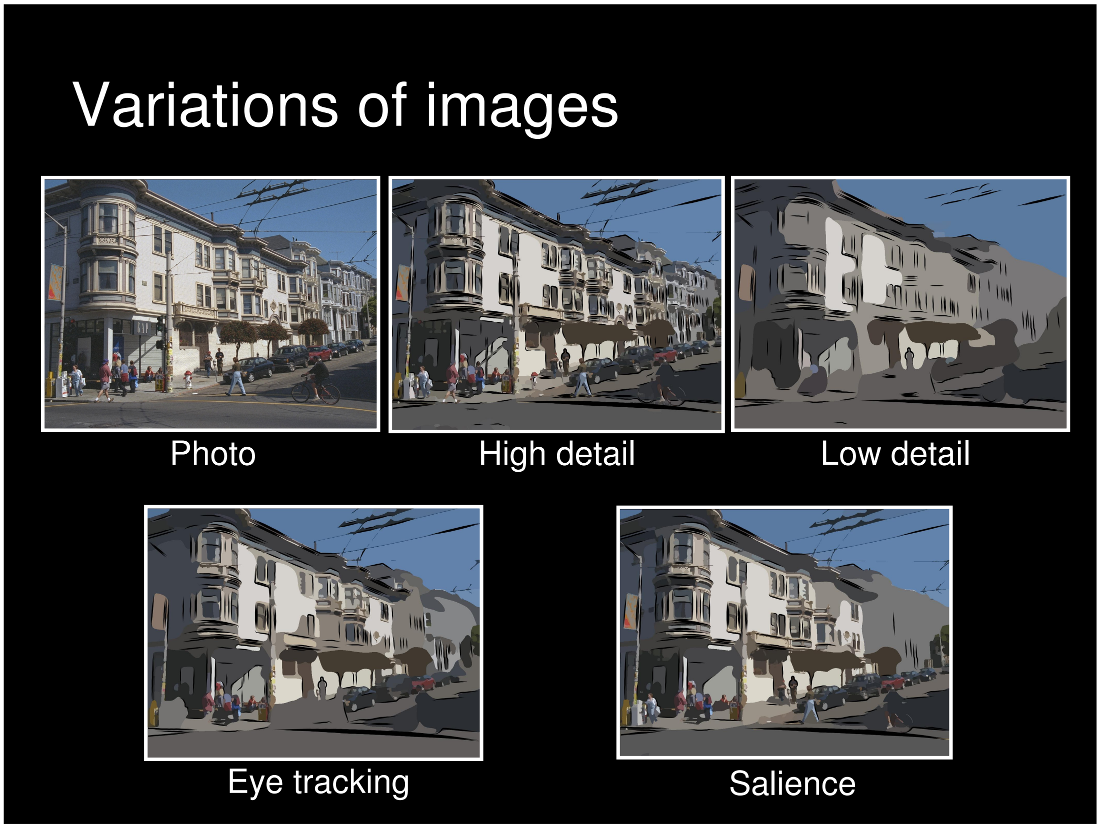
Here again is the design of our system. Eye movements are input to the perceptual model.



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.



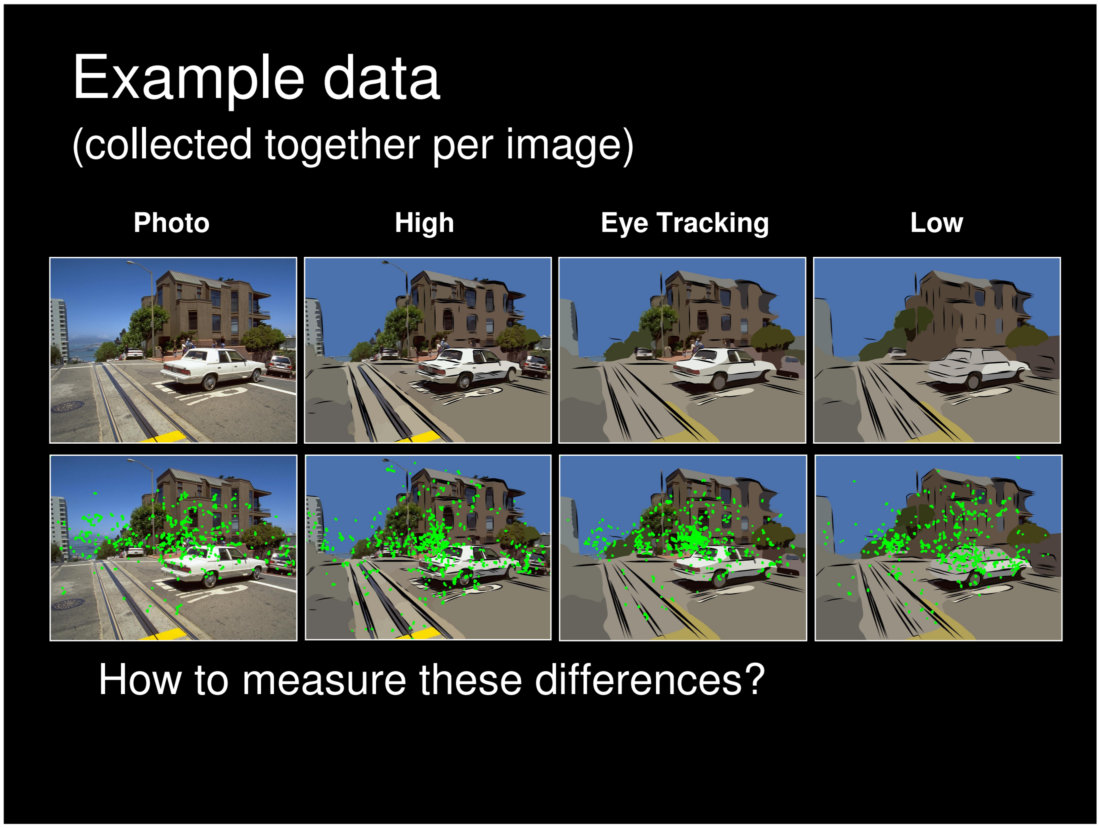
Here are the five conditions. We show each subject in our experiment one of these pictures.

Variations of images

This set of images separates:

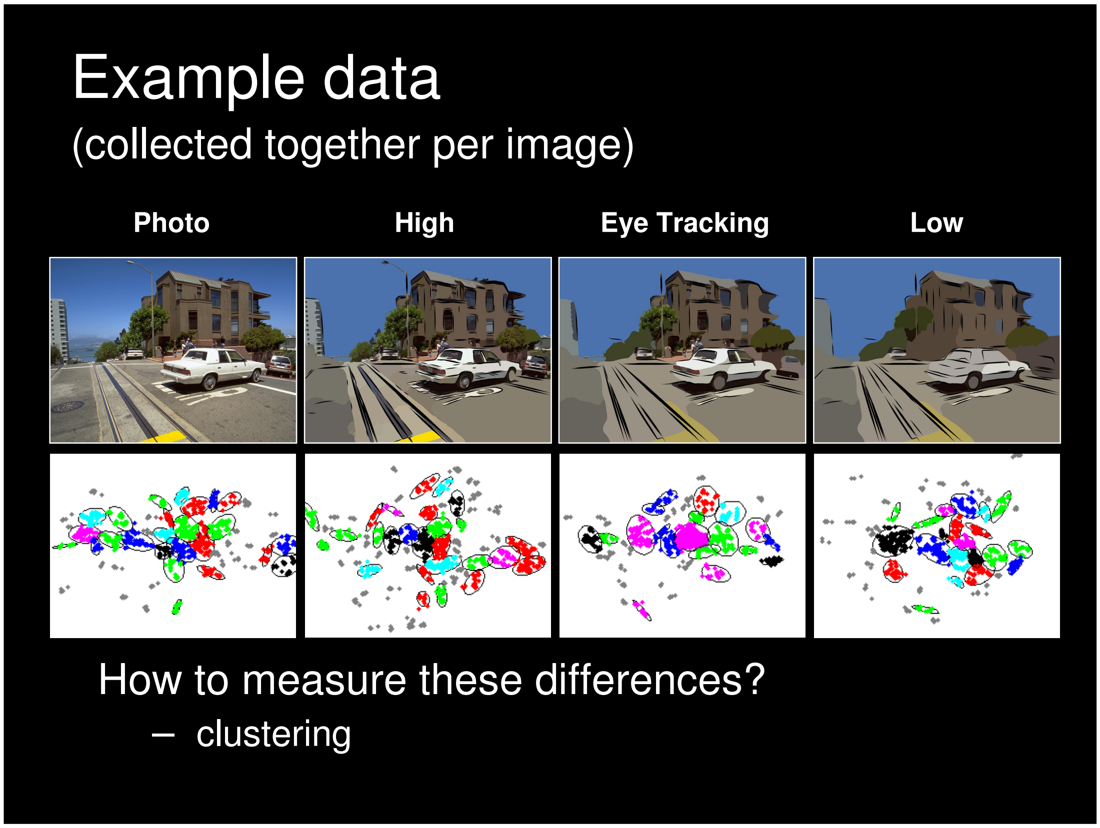
- the visual style
- global level of detail
- locally increased detail
- the locations of increased detail

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



The result is data like this. It 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?

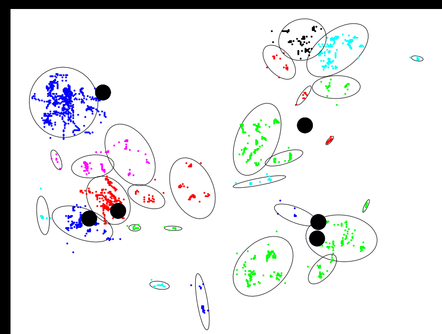


Our approach is to cluster the data—to divide it into limited regions of the image that were viewed coherently.

Quantitative analysis

Look across conditions at:

- number of clusters
- distance to locations where detail was preserved



Once we've cut the data up into clusters, we can compare the number of clusters, and the distance of cluster centers to the points we chose to emphasize in the eye-tracking and salience renderings. One set of detail points is marked here with large 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 more so than uniform detail. 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 as well.

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