Video Interaction at Scale

2018.07.12
Juho Kim
Computer Graphics

internal model of the scene

layout, shape, material, etc.

mathematical representation
suitable for graphical computations

From “A Short Introduction to Computer Graphics” by Frédéric Durand
Computer Graphics

internal model of the scene

layout, shape, material, etc.

mathematical computation
algorithm functionality
performance robustness
HCI

interaction

user model

perception
cognition
usability
learning
satisfaction
experience

system model

mathematical
computation
algorithm
functionality
performance
robustness
Ubiquitous computing, IoT
Interaction at Scale
Large-scale online platforms

Large-scale interaction data from users
How can we understand & improve interaction with large-scale interaction data?

My HCI Research Mission:
Build & study interactive systems powered by crowds, data, and AI.
AI: Computers imitate humans

photo credit: KAIST Hubo Lab
Human Computation:
Humans & computers work together

Image courtesy of Christine Daniloff at MIT
Crowd + Engineered Process Wins

https://www.ted.com/talks/shyam_sankar_the-rise_of_human_computer_cooperation
Advanced Chess

Human computation

human intelligence + computational process
Interactive systems for supporting video interaction at scale

LectureScape  
UIST 2014

RecipeScape  
CHI 2018

ToolScape  
CHI 2014

Crowdy  
CSCW 2015

ConceptScape  
CHI 2018

RIMES  
CHI 2015

Mudslide  
CHI 2015

DynamicSlide
Four types of video interfaces powered by...

**Collective knowledge construction**
- **ToolScape**: Step-by-step information [CHI 2014]
- **Crowdy**: Summary of steps [CSCW 2015]
- **ConceptScape**: Concept maps [CHI 2018]

**Learner responses**
- **RIMES**: Multimedia responses [CHI 2015]
- **MudSlide**: Confusion about the lecture [CHI 2015]

**Data mining**
- **LectureScape**: Data-driven content navigation [UIST 2014]
- **RecipeScape**: Mining & visualizing instructions at scale [CHI 2018]

**Reference extraction**
- **DynamicSlide**: Slide-narration links [ongoing]
Four types of video interfaces powered by…

**Collective knowledge construction**
- **ToolScape**: Step-by-step information [CHI 2014]
- **Crowdy**: Summary of steps [CSCW 2015]
- **ConceptScape**: Concept maps [CHI 2018]

**Learner responses**
- **RIMES**: Multimedia responses [CHI 2015]
- **MudSlide**: Confusion about the lecture [CHI 2015]

**Data mining**
- **LectureScape**: Data-driven content navigation [UIST 2014]
- **RecipeScape**: Mining & visualizing instructions at scale [CHI 2018]

**Reference extraction**
- **DynamicSlide**: Slide-narration links [ongoing]
how-to videos online

This part is also option, im going to add a wind trail for a

The Pythagorean Theorem
Introduction to the Pythagorean Theorem

Pythagorean Theorem
- hypotenuse
- 90° angles
- right angle
- Right triangle
HTML5 and CSS3 Beginner Tutorial 19 - Div and Span

By: Quentin Watt

Wiki Video Outline

**Introduce the idea behind `<div>` and `<span>`**

1. Create a div element using the `<div>` tag
2. Below the div, create a span element using `<span>`
3. Span and div elements were designed to be used with CSS
4. Div and span elements don’t really make visual changes unless you use CSS to style them

**Demonstrate how to use a `<div>` tag**

5. Open a div tag
6. Within the div tag, create an `<h2>` element
7. Beneath the `<h2>`, create a paragraph with some text

**Demonstrate how to use a `<span>` tag**

8. Put a span tag around two words in the text
9. Give one of the span tags class="red_text"
10. Set the other span tag to have class="brown_text"
11. View the page in firefox to see the changes

**Styling a div tag**

12. You can’t see the div or span tags yet because we haven’t styled them
13. Go to your css file and create a div {} statement
14. Set background-color:yellow
15. Set border: 1px solid red

**Styling span tags using classes**
Interactive, Collaborative, Crowd-powered Video Learning

Which subject are you interested in learning?

Statistics

Web Programming

Choose from the how-to videos below

As you watch the videos, we'll periodically check in and ask what you're learning. It's an experiment to help you learn better from videos.

Featured videos

CSS
Introduction to styling with CSS
By: BeginnerTutsdotcom

CSS
Making Divs Side by Side using CSS
By: I Teach Stuff

jQuery
Introduction to Selectors
By: thenewboston
Choose from the how-to videos below

As you watch the videos, we'll periodically check in and ask what you're learning. It's an experiment to help you learn better from videos.

Featured videos

CSS Introduction to styling with CSS
By: BeginnerTutsdotcom

CSS Making Divs Side by Side using CSS
By: I Teach Stuff

jQuery Introduction to Selectors
By: thenewboston

Basic statistics tutorials from Khan Academy
Crowdy coordinates learner tasks to generate a video outline.

What’s the overall goal of the section you just watched?

Learners are prompted to summarize video sections.

UI presents a video outline.

System coordinates learner tasks for a final summary.

Introduce the idea behind `<div>` and `<span>`

1. Create a div element using the `<div>` tag
2. Below the div, create a span element using `<span>`
3. Span and div elements were designed to be used with CSS
4. Div and span elements don’t really make visual changes unless you use CSS to style them
How to extract & present solution structure at scale?

Crowd-powered workflow for complex tasks
- Soylent [UIST 2010], CrowdForge [UIST 2011], PlateMate [UIST 2011], Turkomatic [CSCW 2012]
Multi-stage learnersourcing workflow

Learnersourcing Subgoal Labels for How-to Videos.
Sarah Weir, Juho Kim, Krzysztof Z. Gajos, & Robert C. Miller.
CSCW 2015.
Stage 1. Generate summary labels

- Learner: summarize

What was the overall goal of the video section you just watched?

e.g., Create event handlers

Note: Other users will see your outline to help them better understand the steps in the video.
Stage 2. Evaluate candidate labels

• Learner: compare

Which of the following best describes the video section you just watched?

Choose the best answer (submitted by other users) or add your own.

- importance of css, comparison to html
- defining css
- learn css

I have a better answer: 

Submit Cancel
Stage 3. Proofread summary labels

• Learner: inspect

Does the below statement (submitted by other users) accurately summarize the steps?

*Tip: Good statements usually begin with a verb.*

**Statement:**

basic steps to creating an HTML template

**Steps:**

1. Begin HTML page with `<html>` tag and end with `</html>`
2. Start the `<body>` tag above content, end `<body>` below content
3. Create head section using `<head>` tag
4. Use the `<title>` tag to create page title

- Yes, this statement applies
- No, these steps don't require summarization
- **No, and I want to revise the statement:**
  
  basic steps to creating an HTML
ConceptScape: Collaborative Concept Mapping for Video Learning

Ching Liu, Juho Kim, Hao-Chuan Wang.

CHI 2018. Best of CHI Honorable Mention Award.
ConceptScape Workflow: Hierarchical multiple stages
Four types of video interfaces powered by...

**Collective knowledge construction**
- **ToolScape**: Step-by-step information [CHI 2014]
- **Crowdy**: Summary of steps [CSCW 2015]
- **ConceptScape**: Concept maps [CHI 2018]

**Learner responses**
- **RIMES**: Multimedia responses [CHI 2015]
- **MudSlide**: Confusion about the lecture [CHI 2015]

**Data mining**
- **LectureScape**: Data-driven content navigation [UIST 2014]
- **RecipeScape**: Mining & visualizing instructions at scale [CHI 2018]

**Reference extraction**
- **DynamicSlide**: Slide-narration links [ongoing]
“Draw a picture of an organism showing one or more of the characteristics of life.”

“Sketch a picture that represents how a Renaissance person would see the relationship between a star, a flower and a mole.”
Authoring and submitting multimedia responses

RIMES: Embedding Interactive Multimedia Exercises in Lecture Videos. 
Juho Kim, Elena L. Glassman, Andrés Monroy-Hernández, Meredith Ringel Morris. 
CHI 2015.
RIMES: Embedding Interactive Multimedia Exercises in Lecture Videos.
Juho Kim, Elena L. Glassman, Andrés Monroy-Hernández, Meredith Ringel Morris.
CHI 2015.
Reviewing exercise: Look over the following sentences. Then, record yourself as you identify the subject and the verb of each sentence. Circle the subject in green and circle the verb in red.

1. My dog and his best friend knocked over the garbage to look for food.

2. Rummaging through the trash, who's been friends with my dog for years, found the greatest of all treats: an uneaten bagel.
Mudslide: Spatially anchored confusion via learnersourcing

Four types of video interfaces powered by...

**Collective knowledge construction**
- **ToolScape**: Step-by-step information [CHI 2014]
- **Crowdy**: Summary of steps [CSCW 2015]
- **ConceptScape**: Concept maps [CHI 2018]

**Learner responses**
- **RIMES**: Multimedia responses [CHI 2015]
- **MudSlide**: Confusion about the lecture [CHI 2015]

**Data mining**
- **LectureScape**: Data-driven content navigation [UIST 2014]
- **RecipeScape**: Mining & visualizing instructions at scale [CHI 2018]

**Reference extraction**
- **DynamicSlide**: Slide-narration links [ongoing]
Video lectures in MOOCs
Classrooms: rich, natural interaction data
it's like i'm talking to a wall
First MOOC-scale video interaction analysis

<table>
<thead>
<tr>
<th>Learners</th>
<th>Videos</th>
<th>Mean Video Length</th>
<th>Processed Video Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>127,839</td>
<td>862</td>
<td>7:46</td>
<td>39.3M</td>
</tr>
</tbody>
</table>

Data Source: 4 edX courses (fall 2012)

Domains: computer science, statistics, chemistry

Video Events: start, end, play, pause, jump
Factors affecting video engagement

Metric: session length

Shorter videos
- significant drop after 6 mins

Informal shots over studio production
- more personal feel helps

Tablet drawing tutorials over slides
- continuous visual flow helps
How do learners navigate videos?

• Watch sequentially
• Pause
• Re-watch
• Skip / Skim
• Search
Collective interaction traces

Learner #7888
Learner #7887
......
Learner #4
Learner #3
Learner #2
Learner #1

video time
Collective interaction traces into interaction patterns
Interaction peaks

Temporal peaks in the number of interaction events, where a significant number of learners show similar interaction patterns.
Observation: **Visual / Topical transitions** in the video often coincide with a peak.
Returning to content

before transition

after transition

# play button clicks

interaction peak

visual transition

video time
Video player adapts to collective learner engagement

Learners watch videos.

UI provides social navigation & recommendation.

System analyzes interaction traces for hot spots.

- Learner3879, Video327, “play”, 35.6
- Learner3879, Video327, “pause”, 47.2

Other learners rewatched this part.

How to solve this recursively?
- First, convert the string to just characters, by stripping out punctuation, and converting upper case to lower case
- Then
  - Base case: a string of length 0 or 1 is a palindrome
  - Recursive case:
    - If first character matches last character, then is a palindrome if middle section is a palindrome
LectureScape: Lecture video player powered by collective watching data

Data-driven interaction techniques for improving navigation of educational videos.

Rollercoaster Timeline
Embedded visualization of collective interactions

How to solve this recursively?

- First, convert the string to just characters, by stripping out punctuation, and converting upper case to lower case
- Then
  - Base case: a string of length 0 or 1 is a palindrome
Visual clip highlights

Interaction data + frame processing

- Recursion on non-numerics
  - How could we check whether a string of characters is a palindrome, i.e., reads the same forwards and backwards
  - “Able was I ere I saw Elba” – attributed to Napoleon
  - “Are we not drawn onward, we few, drawn onward to new era?”

- How to solve this recursively?
  - First, convert the string to just characters, by stripping out punctuation, and converting upper case to lower case
  - Then
    - Base case: a string of length 0 or 1 is a palindrome
    - Recursive case:
      - If first character matches last character, then is a palindrome if middle section is a palindrome
Learning about iterative algorithms...

Iterative algorithms

- Looping constructs (e.g. while or for loops) lead naturally to iterative algorithms.
- Can conceptualize as capturing computation in a set of “state variables” which update on each iteration through the loop.
Pinning:
Automatic side-by-side view

Pinned slide

Video stream

Recursion on non-numerics

• How could we check whether a string of characters is a palindrome, i.e., reads the same forwards and backwards
  - “Able was I ere I saw Elba” – attributed to Napoleon
  - “Are we not drawn onward, we few, drawn onward to new era?”
RecipeScape: Mining & Visualizing Instructions at Scale

“What makes the cookies so tasty? The ingredients aren’t really all that unique. It’s the process that does it.” - J. Kenji López-Alt
What can we learn from analyzing the **embedded processes** in hundreds of recipes for a single dish?

- What makes a chocolate chip cookie a chocolate chip cookie?
- Average recipe? Creative variants or outliers?
- Simplest vs most complex recipe?
- Can we discover ordering constraints?
- Can we cluster fundamentally different ways of making a cookie?
RECIPE VISUALIZATION

- Each cluster can be toggled.
- Prototypical recipes are marked with stars.
- Each recipe can be selected.
COMPARISON

Mini Chocolate Sandwich Cookies vs Rich Chocolate Rum Cookies

Instructions
- whisk
- over
- stir
- whisk
- chill
- form
- preheat
- cool
- cut
- bake
- cool

Ingredients
- baking powder
- hazelnuts
- simmering water
- melted chocolate
- dough
- logs
- cookies

By Cooking Process

By Ingredients

Chocolate-Dipped Ice Cream Sandwiches

Recipe View

Ingredients
1. Nonstick vegetable oil spray
2. 1 cup all-purpose flour
3. 1/4 teaspoon baking soda
4. Pinch of salt
5. 1/2 cup (1 stick) unsalted butter, cut into 4 pieces
6. 1/2 cup (packed) golden brown sugar
7. 2 tablespoons golden syrup (such as Lyle’s) or light corn syrup
8. 2 large eggs
9. 1 tablespoon vanilla extract
10. 2-3 cups premium ice cream (such as dulce de leche or banana), softened
11. 9 ounces bittersweet chocolate (do not exceed 65% cocoa), chopped
12. 3 tablespoons vegetable oil
13. Assorted decorations (such as chopped nuts, colored sprinkles, toffee bits, and cocoa nibs)
14. Even easier: Make sandwiches with your favorite ice cream and ice cream. Dip in the melted chocolate, then decorate.
15. Test-kitchen tip: If you use a soft ice cream (such as dulce de leche),

Tree View
Temporal usages pattern of top 10 cooking ingredients / actions
Bake early, use cookies in other dishes  vs  Cookies as the final result
Recipe Text: Stir in remaining 6 tablespoons corn syrup and vanilla.

Part-Of-Speech Tag: Stir in remaining 6 tablespoons corn syrup and vanilla.

Human Annotation: Stir in remaining 6 tablespoons \textcolor{red}{corn syrup} and \textcolor{red}{vanilla}.

Tree Representation:

Calculate distance between Trees:

Plotting Distances:

Distance Matrix:

Gram Matrix:

Hierarchical Clustering:
HUMAN & MACHINE HYBRID LABELING

14,988 cooking actions (54%) confirmed by human annotators

Human annotator corrected 9,987 cooking actions the NLP tagger mislabeled.

~40% of the 24,975 cooking verbs.
INGREDIENTS  Sugar, Peanuts, Corn Syrup, Cocoa Butter, Fractionated Palm Kernel Oil, Chocolate Liquor, Whole Milk Powder, Non Fat Dry Milk, Whey Powder, Soy Lecithin, Cocoa Powder, Butter, Egg Shell, Enzyme Modified Soy Beans

Stronger analytical capabilities

Interpretations  Explanations  “Sensemaking”
Ongoing Work: RecipeScape for Video

Authentic Italian Pizza Recipe (Pizza Margherita & Pizza Bianca)

Main Video

Related Snippets

How To

Set as Main Video
recipescape.kixlab.org
Four types of video interfaces powered by...

**Collective knowledge construction**
- **ToolScape**: Step-by-step information [CHI 2014]
- **Crowdy**: Summary of steps [CSCW 2015]
- **ConceptScape**: Concept maps [CHI 2018]

**Learner responses**
- **RIMES**: Multimedia responses [CHI 2015]
- **MudSlide**: Confusion about the lecture [CHI 2015]

**Data mining**
- **LectureScape**: Data-driven content navigation [UIST 2014]
- **RecipeScape**: Mining & visualizing instructions at scale [CHI 2018]

**Reference extraction**
- **DynamicSlide**: Slide-narration links [ongoing]
DynamicSlide: Reference-based Interaction Techniques for Slide-based Lecture Videos

Lecture slide and narration are complementary
- Lecture slide: Visual, brief, hierarchy
- Narration: Audio, detailed, less structured

How can we find and leverage the link between slide and narration (reference)?

Ongoing work in collaboration with Hyeungshik Jung (KAIST) & Hijung Valentina Shin (Adobe Research)
Post-hoc dynamic video presentation

Attribute Classification

Discrete Attribute
Has a finite or countably infinite set of values
- Examples: zip codes, click counts, set of words in a collection of documents
- Often represented as integer variables
- Binary attributes are a special case of discrete attributes

Continuous Attribute
Has real numbers as attribute values
- Examples: temperature, height, or weight
- Continuous attributes are typically represented as floating-point variables
Post-hoc video editing

What to Expect from the Session

- Understand key differences between NoSQL and RDBMS
- Learn the fundamentals of Amazon DynamoDB
- Identify suitable use cases for DynamoDB
- Plan and execute a migration from RDBMS to DynamoDB
- Explore a customer use case: Sony’s migration from MySQL to DynamoDB
Computational Pipeline

Slide Image
- Slide #1
- Slide #2
- Slide #3

Text Segments
- OCR + Heuristics
- Shot Boundary Detection / Classification

Script (sentence)
- Cosine similarity btw.
- TF-IDF weighted bag of words

Video
- Text difference + Pixel difference
- Content & Layout Detection

Transcript
- Text to Script Alignment
Interesting research directions

• High-level video synthesis
  – `createVideo(task="cookie", style="slide")`

• Post-hoc, lightweight editing
  – Breaking the encoding wall

• Support for different types of video
  – VR, 360, mobile, distributed, egocentric, etc.

• Rethinking video as a medium
  – Personalized presentation, Hybrid media formats