IDC HERZLIYA

Efi Arazi School of Computer Science

The Interdisciplinary Center, Herzlia

Efi Arazi School of Computer Science
M.Sc. program - Research Track

Video Summarization using Causality Graphs

by

Shay Sheinfeld

M.Sc. dissertation, submitted in partial fulfillment of the requirements for the M.Sc. degree, research track, School of Computer Science The Interdisciplinary Center, Herzliya

June 2016

This work was carried out under the supervision of Prof. Ariel Shamir from the Efi Arazi School of Computer Science, The Interdisciplinary Center, Herzliya.

Abstract

Video summarization is useful for many applications such as content skimming and searching. However, automatic summarization is extremely challenging due to problems such as detecting scene changes, determining who or what is in each scene, recognizing the words and actions, and above all assigning meaning, causal relationships, and importance to the displayed events. Computer vision algorithms can provide some solutions to the former tasks, but the latter task is semantic in nature and challenging for machine computation.

We present a reliable, crowdsourced solution to video summarization based on human computation that addresses one of the main semantic challenges in story understanding: recognizing cause and effect. Our approach first automatically divides the video into simple shots as atomic elements. Since it is more natural for humans to reason about semantics using text, our approach converts each atom to text using the crowd. We then utilize the recent context tree approach of Verroios and Bernstein (2014), to gain global understanding. Finally, our algorithm addresses causality relations by explicitly building a causality graph between story units in the context tree. Our evaluation shows that information from the causality graph creates better summarizations of the original video.

Table of Contents

Introduction

Video summarization, in which a long video is shortened, is useful for many applications such as content skimming and search. Although fully automatic summarization algorithms are highly desired, they require solutions to many difficult tasks: detecting scene changes, determining who or what is in each scene, recognizing the words and actions, and above all assigning meaning, causal relationships, and importance to the displayed events. Computer vision algorithms can provide some solutions to the former tasks, but the latter task is semantic in nature and therefore the most challenging for a fully automatic algorithm (machine computation). Rather than automating this process entirely, we describe a reliable, crowdsourced solution to video summarization that addresses one of the main semantic challenges in story understanding: recognizing cause and effect.

One naive solution for video summarization would be to ask a human to watch the movie and then summarize it in text. This demands high involvement and effort of one person, and requires a follow-up conversion of the text summary back to video, which is a difficult problem in itself. Another possibility would be to segment the video into short pieces, distributing the task (assigning importance) among many individuals. However, this jeopardizes the global understanding of context and cause and effect of each individual piece.

We present a hybrid automatic (machine) and human computation approach to the problem. Our approach first automatically divides the video into simple shots as building blocks. Video, unlike text, may or may not be verbal and does not have built-in demarcated divisions (e.g. sentences or paragraphs). Our algorithm automatically considers shot boundaries and separates the video into smaller units. Using single shots, rather than scenes, also provides flexibility in the final stage of composition of the video summary.

Since it is more natural for humans to reason about semantics using text, our approach converts each shot to text using the crowd while maintaining inter-crowd consistency in naming characters. To gain global understanding, we then utilize the recent context trees approach of Verroios and Bernstein (2014), which was designed to create global understanding in tasks where each contributor only has access to local views. For instance, given a long text, the context tree approach divides the text among the crowd for summarization and then recursively builds a tree by merging and re-summarizing until one root node is reached. In a second pass, the partial summary in each node is ranked in terms of its overall importance.

One problem with relying on the context tree alone for summarization is that long-term dependencies between events are hard to detect. For instance, an important event near the end of the movie could be caused by an event near the beginning and these links may be missed even in the context tree. Furthermore, the importance of the context tree does not provide knowledge about cause and effect. The root cause of an important event could be dropped from the summary because its own importance is low, causing a gap in the story and misunderstanding. Our method addresses such causality relations by explicitly building a causality graph between story units in the context tree (Figure 1). This graph is then used for better summarizations of the original video.

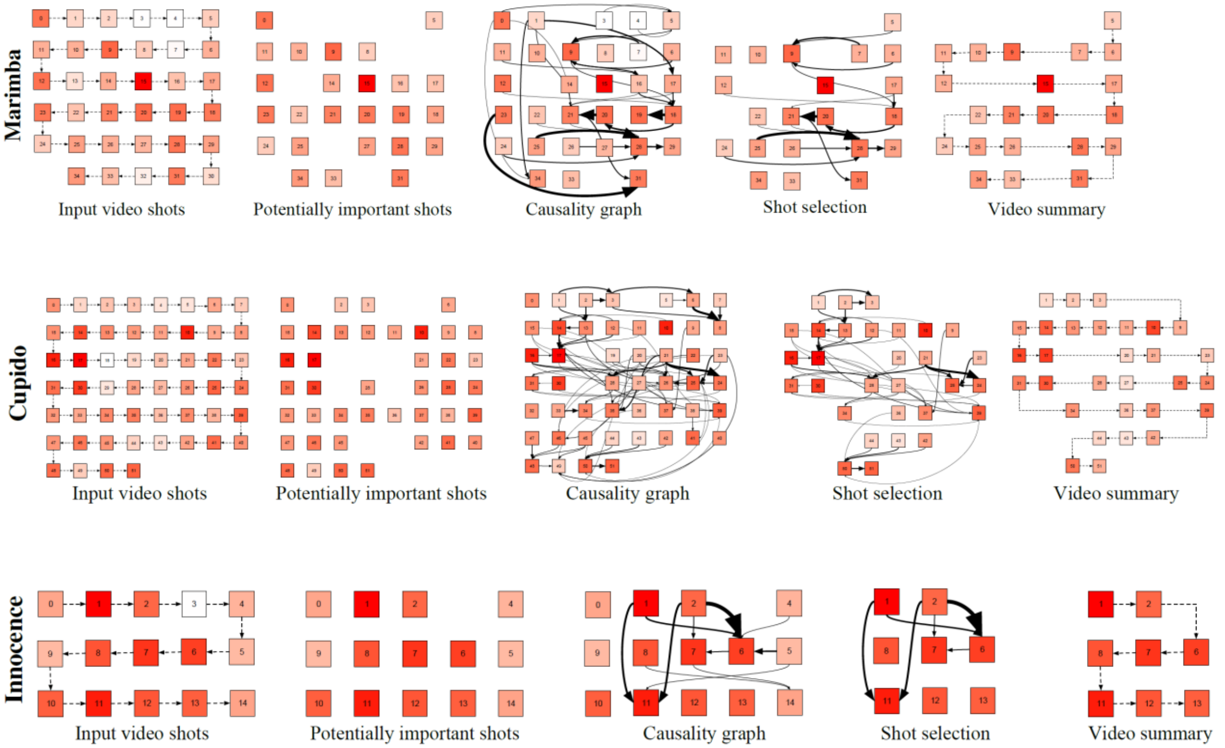


Figure 1: The entire video is automatically decomposed into shots. The importance of each shot is computed by the crowd via the context tree algorithm (Verroios and Bernstein 2014). (More important shots are darker.) Potentially important shots are input into our crowd-based causality graph algorithm which computes causal relationships between shots. (Thicker edges in the causality graph are higher weight.) Finally, important shots and their causal dependencies are selected to create a video summary.

We tested our approach on several videos and compared it to baseline (naive) summaries as well as summaries created using the context tree alone. Our evaluation shows that information from the causality graph creates better summarizations of the original video.

Background

Extensive work has been done in the field of video abstraction and summarization. Truong and Venkatesh (2007) conducted an extensive survey of video abstraction techniques which produce either a sequence of representative still images or a shorter video containing only the most important parts of the input video. (The latter is our goal.) In general, the output of video abstraction varies and includes comic books (Uchi- hashi et al. 1999), skimmable static images (Assa, Caspi, and Cohen-Or 2005), and shorter videos. These techniques rely on various heuristics for "understanding" video contents. For example, Zhuang et al. (1998) use unsupervised learning to detect salient content as key frames. Their approach clusters all frames and selects the centroids of big enough clusters to be key frames. Uchihashi et al. (1999) outputs a summary of a video as a Japanese comic (manga) in which more important key frames will be bigger in the comic strip. This approach determines the semantic importance of video segments using unsupervised learning and heuristics. The learnt video segments are said to be important if they are long and different from other segments. These criteria do not always correlate with importance. Smith and Kanade (1998) introduced an algorithm for shot characterization that makes use of object detection, camera movement, transcript analysis, and histogram changes. Heuristics are used for scene detection. The understanding of which shots are semantically important is performed via a set of rules created based on research into video production practices. Pritch et al. (2008) shorten an input video with a static camera by shifting events in time to show actions in non-chronological order and simultaneously. Without humans in the loop, these automatic approaches are limited to heuristics for understanding the video contents. We do, however, employ an automatic approach to shot detection in Step 1 of our algorithm.

If the input videos are from a restricted domain, additional heuristics can be used. For example, He et al. (1999) shortened recordings of presentations with accompanying slides. They used heuristics like the pitch of the speaker's voice and when information is most meaningful after slide transitions. Shin et al. (2015) converted blackboard-style lecture videos into a static representation which interleaved blackboard drawings with transcript text. Heck et al. (Heck, Wallick, and Gleicher 2007) created blackboard lecture videos with high production values via an algorithm that makes videography choices given multiple simultaneous recordings. Pavel et al. (2014) presented approaches for creating video digests, browsable and skimmable static representations of lecture videos with links to instantly jump to the relevant portion of the video. Their approach supports two pipelines: (1) manual authoring of video digests and (2) automatically segmenting the video and crowdsourcing summaries of the segments. Relatedly, Kim et al. (2014) and Weir et al. (2015) presented approaches for crowdsourcing static representations of how-to videos. In contrast, our aim is to shorten fiction videos while maintaining the same output medium (videos). Similarly to the latter three approaches, we employ crowdsourcing to access human-level understanding.

Crowdsourcing has been used for video and text summarization in the past (as well as visual tasks more generally (Goldman and Brandt 2011; Gingold, Shamir, and Cohen-Or 2012)). Soylent (Bernstein et al. 2010) shortens text documents via human computation. Soylent operates by shortening paragraphs, but will not remove them entirely. Our approach to summarization will eliminate shots entirely if they are unimportant. Wu et al (2011) described a technique which summarizes a movie by asking each worker to watch the entire movie, hitting a dedicated button when they think the video is reaching a highlight. This approach does not make use of the parallelism of the crowd. Bernstein et al. (Bernstein et al. 2011) introduced an approach to find high quality still images in short videos with parallel crowdsourcing. Relatedly, Di Salvo et al. (2013) presented a game with a purpose for a crowd to play to generate high quality annotations of objects in videos. Our approach relies on parallel crowdsourcing. We introduce the causality graph in order to “rescue” seemingly less important moments in a movie if they are causes of important moments.

Our causality graphs are closely related to plot graphs used in the story generation literature. Plot graphs were first described by Kelso et al. (1993) as a structure for modeling events cause (or, equivalently, must precede) other events in interactive storytelling (Weyhrauch 1997; Young 1999). Li et al. (2013) learned a plot graph for a user-requested type of story (e.g. bank robbery) by analyzing many example stories obtained via crowdsourcing. Gupta et al. (2009) modeled simple cause and effect actions in baseball videos, such as when a baseball is pitched, the batter can hit, miss, or not swing at all; a hit can lead to a run; etc. Their aim is to learn these cause and effects possibilities by analyzing many annotated videos of the same genre. We make use of causality graphs from the opposite point of view. We create a causality graph for a single story which already exists, the input movie, in order to preserve important events when shortening.

Algorithm Overview

In our algorithm each worker only sees a local part of the input video. The algorithm distributes and combines all these local views into a global understanding of the video's plot, including the cause and effect graph of events in the video, crucial for creating high quality summarization.

Our algorithm is composed of the following steps. Some of these steps (namely, 1, 5, 7 and 8) are automatic in nature, and others (steps 2, 3, 4, 6) use *human computation*:

1. Shot detection
2. *Character naming*
3. *Creation of textual descriptions for shots*
4. *Importance scoring using the context tree*
5. Filtering of shots
6. *Construction of the causality graph*
7. Shot selection
8. Video summary creation

Examples of all human computation tasks are provided at the end of this chapter.

Our algorithm utilizes the context trees algorithm of Verroios and Bernstein (2014). The context tree is a human computation algorithm that tackles the problem of global understanding when each contributor only has access to local views. The algorithm uses a two phase approach: in the first phase, a summarization tree is built bottom up, and in the second stage traversal of this tree is performed top down to rate the importance of parts.

The output of the context tree algorithm is an importance score for each leaf. This importance score can then be used to construct a video summary. However, such a summary will not adhere to causalities in the story. For instance, some events may be dropped due to low scoring, even though they cause events that do appear in the summary.

To rectify this, we propose a causality graph that encodes causality relations between nodes and allows for better summarizations. We first use the importance score to prune the context tree and then use human computation to extract causality relationships between the important parts of the movie.

To provide simpler means for humans during the creation of the context tree and the causality graph, we first translate the input video into a textual representation. The video is segmented into simple shots and each one is translated. By using simple shots as building blocks, we can more easily combine them to create the output video summarization. Our final summarization is based on both the context tree scores and the causality graph.

Preliminaries

Step 1: Shot detection

The first step of our method creates the basic building blocks that cannot be separated during the various stages of the algorithm. The building blocks must be short enough to contain a single event in the video, but not so short that an event will be distributed over several blocks. We assume that the input video is built from multiple shots that tell a story. Therefore, we chose to use single shots as the building blocks of our algorithm. We further constrain the shots to have a minimum length of 4 seconds.

The requirement to have one block contain a single event is important, since the algorithm is based on rating the importance of blocks. If one block contains more than one event, it would be difficult to assess its importance. Moreover, the final summary video output is constructed using such building blocks. If a building block contains several events or only part of an event, then the output summary is more likely to contain discontinuities, as well as events that are not important to the understanding of the video.

Extensive research has been done on the shot detection problem (Boreczky and Rowe 1996). We use a region-base histogram comparison to detect shot boundaries. Every two successive frames are split into four equal regions, and the HSV histogram of every two corresponding regions are compared using correlation. If the average correlation is more than a threshold and the length of the current segment is more than 4 seconds, a shot boundary is marked and the video is split.

Steps 2–3: Textual descriptions for Shots & Naming

As mentioned earlier, it is more natural for humans to reason about semantics using text. Therefore, each extracted shot is translated into a textual representation. This text will be used in the context tree algorithm as well as the causality graph creation. We use human computation to translate the shots into text. As this translation will be performed by different workers, the names of the characters must be unified such that each character will be addressed in the same manner, and two different characters will have two different names. This naming process is crucial as each worker only sees a small part of the input video. Without consistent naming, workers would not be able to infer which characters other workers are referring to or understand the plot of the movie.

Computer vision algorithms can detect human characters in a video, but they must also identify characters consistently across different shots in the movie. In addition, some characters, such as background actors, are not important to the understanding of the story and need not be named. Lastly, for non-human characters—as in cartoons or computer animated movies—even the detection problem becomes a challenge. In our case, either we use well known characters (e.g. Popeye) or we resort to manual labeling. In our algorithm, important characters' pictures and names are provided to the worker in the description of the task.

To convert shots to text, a worker is shown a single building-block shot cut from the movie and is asked to write a summary of the shot. The worker is instructed to use the correct character names when referring to a character in the summary.

After a worker completes the task, another task is generated to iteratively refine the description (Little et al. 2010). The worker is shown all information shown to the first worker (the shot and characters), as well as the previous description, and is asked to refine the summary or mark it as “good enough.”

This step is important for producing a comprehensive summary. A flawed or incorrect description of an event may be difficult to correct in later stages of the algorithm. Moreover, an event that seems unimportant by itself may prove to be very important in the context of the whole story and may later be the cause of an important event. Because of this, in the up-phase of constructing the context tree (next section) we not only show the textual summaries but also representative (important) shots from the subtree (following Verroios and Bernstein (2014)). In this case, if the description of the shot misses some semantically important detail, then the worker will be able to correct it in context.

Step 4: Context Determination

The basic building blocks of the context tree algorithm are small parts of the original input. These parts are mutually exclusive, their union is the original input, and they constitute the leaves of the tree. For example, Verroios and Bernstein (2014) used their context tree algorithm to summarize a whole book by splitting the book into paragraphs. By definition, paragraphs are sections of text that deal with a single theme, and therefore commonly describe a single event. Unlike text, videos do not contain markers splitting them into well-defined sections. Verroios and Bernstein (2014) also described an application of their algorithm to a video divided into scenes whose average length was 3.5 minutes. Our approach, for reasons described above, divides the video into smaller building blocks, each containing a single shot.

We augment each shot with the textual description created in step 3. In essence this means that the leaves of the context tree are the text description of the shots, and not the shots themselves. There are several reasons for this. First, text is easier to skim, examine, and manipulate than a video. Second, it is easier for humans to capture semantics in text than in a video. Lastly, the original context-tree algorithm has proven to be very effective in summarizing text.

Constructing the Context Tree

We apply the context tree algorithm to obtain the importance of each shot. Our context tree construction algorithm follows the original two phase approach:

1. In the up-phase (from the leaves to the root), workers are shown b successive parts (e.g. shots) of the input video and asked to write a text summary of these shots. At the end of this phase, the root contains a context-less text summary of the whole input video.
2. In the down-phase (from the root back to the leaves), workers provide context to the text summary. They are asked to rate the importance of each inner node for understanding the parent node summary. At the end of this phase, all leaves have a normalized importance score.

There are two parameters that govern the tasks of workers and the shape of the context tree. First, we use a branching factor of $b = 3$. This means that in the up-phase, every three consecutive nodes are joined under a single parent (i.e. shots 1, 2, 3 and shots 4, 5, 6, etc.). Second, we use a replication factor of $r = 4$, which means that every corresponding task is replicated four times, to be answered by four different workers.

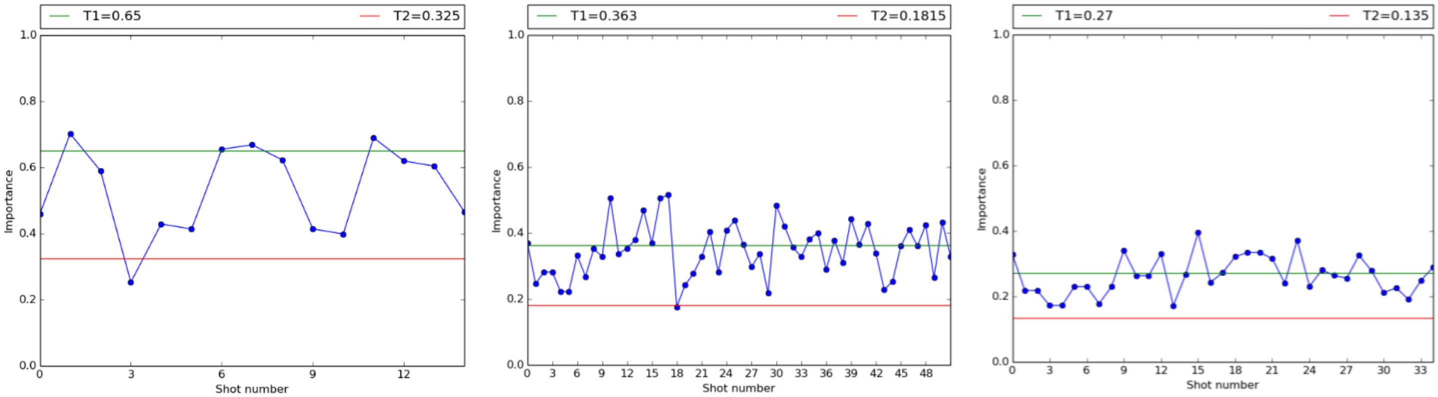
Up-phase In this stage workers are shown b sections, each with r summaries. A worker is first asked to choose, for each section, the best summary. The worker is then asked to write one summary for all b sections. Lastly, the worker is shown some representative video shots and asked to choose one shot that best illustrates the summary she just wrote. The representative video shots are all the shots that were chosen by the $b \cdot r$ workers in the children of the current node.

For the summarization portion of this task, workers were instructed that summaries appear in chronological order. This instruction is important since some movies contain repetitions, and summaries can look like alternative summaries for the same part.

The first level in the up-phase, the parents of leaf nodes, are treated somewhat differently, since leaf nodes simply correspond to atoms (shots). As a result, the shot descriptions are not replicated and the worker is simply shown and asked to summarize the *b* shot descriptions. The representative shots are all *b* shots.

Down-phase In this stage workers are shown all ***r*** textual summaries of the whole video and all ***r*** summaries of the current node. The worker is asked to describe the most important event in the node for understanding the whole-video summaries. The worker is then asked how important that event is for understanding the overall story on a 7-point Likert scale. That number is the node's importance.

Due to the fact that ratings are subjective, the range varies between workers (Herlocker, Konstan, and Riedl 2002). Moreover, each worker rates a node locally. To achieve a global rating, we follow the nor scheme of Verrios and



(a) (b) (c)

Figure 2: The importance score of every shot in Innocence (a) Cupido (b) and Marimba (c), as computed by the context tree algorithm. We use two thresholds: every shot above threshold T_1 is marked as important, as well as local maxima shots above threshold T_2 (for example, shot number 4 in (a)). All other shots are pruned before the construction of the causality graph.

Step 5: Filtering

The output of the previous context tree step is a numeric score for every node, including leaf nodes, which represent individual shots.

Hence, the importance score of each leaf node provides a way to rank the shots. However, the absolute value of importance may still be affected by the local perspective of workers. We have found that the relative value of the important score, i.e. the value of a shot relative to its neighboring shots, can be indicative of the shot's importance. Therefore, we define two types of shots:

1. Top rated shots: A highly rated shot means that the shot is one of the most important for understanding the whole video.
2. Local maxima shots: A shot which is more important than the shot before and after implies that the events of the movie reach an importance peak at that shot.

Our filtering scheme is based on two importance threshold parameters. T_1 is the threshold for the top rated shots. Any shot whose importance score s_i satisfies $s_i \geq T_1$ will be considered as potentially important. $T_2 =$ *12*T_1 is the threshold for local maxima shots. Any local maxima shot whose importance score s_i satisfies $s_i \geq T_2$ will also be marked as potentially important. Before the next stage of creating the causality graph we filter out all shots that were not marked as important (see Figure 2).

Given a target length L for the video summary, it is straight forward to find T_1 such that the length of important shots is approximately L. However, we wish to consider additional nodes as important based on their causality relationship to important nodes. Therefore, we choose T_1 such that the length of important shots is 1.5L.

We note, however, that other values for T_1 and T_2 may produce better results for a given input video. One can use these parameters to characterize the summary of the movie. For example, if the movie has one major event, then using only top rated shots will be sufficient, i.e. setting $T_1 = T_2 = t$, where t is the importance of the least important shot that is added to the summary. Another example is a movie with a lot of repetition. If the repetition is somewhat important to the understanding of the plot, then all repeated scenes in the movie will have a local maxima shot, and T_2 should be adjusted to include all of them in the summary.

Step 6: Causality Determination

Although the previous stage provides a way to determine important events in the movie, there are no connections between these events and the plot structure is missing. Our causality graph provides a way to connect these events in a meaningful manner.

A causality graph is a weighted directed graph. Its nodes are events (shots) from the video. The directed edges represent a causal relationship between two events in the video: the source event is a direct cause of the target event. The weight of each edge represents the strength of the relationship, measured by the evidence found for this relationship.

A naive crowdsourcing algorithm for constructing the causality graph is to ask, for every pair of shots, if one causes the other. This will yield $O(n^2)$ queries. Furthermore, the resulting causality graph could contain incorrect edges, since a worker cannot correctly identify a causal relationship between two events without context. For example, if the first event is about someone making a phone call, and the second event is about someone answering a phone call, a worker may indicate a causal relationship between the events, despite the fact that the first person called someone else. Instead, our algorithm for constructing the causality graph uses the context tree created in Step 4 and the important shots identified in Step 5 both for efficiency and for solving the context understanding problem.

Constructing the Causality Graph

First, we prune not potentially important nodes from the context tree. The importance of a node is computed in the same way as the importance of a shot using the same two thresholds; local maxima nodes consider adjacent nodes at the same level of the context tree. Using the pruned context tree, we build a sequence of causality graphs whose source nodes are nodes from level *i* of the context tree and target nodes are the leaves (i.e. individual shots). The construction of the causality graph for level *i* depends on the causality graph of the previous (higher) level. The final causality graph will include only leave

We define a query node as a text leaf node describing one shot in the movie. Note that the set of query nodes is fixed, and does not change when passing from one level of the context tree to the next. We define a causal node as any node in the context tree that could potentially be a cause of a query node.

The causality graph for level *i*, denoted CG_i, is a causality graph whose nodes consists of all the query nodes and the causal nodes from level *i* of the pruned context tree. Our final causality graph will consist of only query nodes (i.e. shots) and edges between them, but along the way we recursively construct several causality graphs.

Base case The base case of the algorithm constructs CG_1, the causality graph for the level below the root of the tree. Using the crowd, we find, for each query node, all the possible causal nodes in the first level of the tree. For a given query node, we consider only nodes in the first level of the pruned context tree whose subtree contains shots which chronologically happened before the query node.

Recursion The causality graph $\mathrm{CG_i}$ of level i is a graph that consists of edges from causal nodes at level i in the pruned context tree to query nodes. To construct graph $\mathrm{CG_{i+1}}$, the potential causes (causal nodes) for query nodes are calculated as follows: Let v be a node from level i+1 of the pruned context tree[1], $p(v)$ its parent, and q a query node. We say that v is a potential cause for q if $(p(v), q)$ is an edge in $\mathrm{CG_i}$.

While constructing $\mathrm{CG_{i+1}}$, every query node q has a set of potential causes which are causal nodes from level i+1. We use the crowd to determine whether each potential causal node v is actually a direct cause of q or not. We ask each query $d = 5$ times. The count $c(v, q)$ of an edge (v, q) is the number of workers who indicate that v is a direct cause of q. If the count is 0, we do not create an edge at all.

Edge Weights The weights of the edges in the final causality graph CG (all of whose nodes are shots) are defined recursively by adding the counts of their parents (see Figure 1). The weight of the edge between nodes v and q in CG is defined (recursively) as:
$w(v, q) = d(v) \cdot c(v, q) + w(p(v), q)$,
where $d(v)$ is the depth of v (distance from the root), $p(v)$ is the parent of v in the context tree, and $c(s, q)$ is the count of edge (s, q) in a causality graph (of any level) connecting s and q, i.e. the number of workers indicating a causal relationship between s and q.

The Crowdsourcing Task

A cause and effect relation may be interpreted differently among different workers. Therefore, the workers are asked to mark an event as a cause only if they think it is a direct cause of the event. In addition, the workers see an example for a summary, query, and expected causal node.

Every task consists of sections of query nodes. Each section starts with a textual summary of the query node's event followed by checkboxes next to the textual summaries of potential causal nodes. The last checkbox in every section is "None of the above." Workers were instructed "Which of the following are possible direct causes of this event? Check all that apply. (If none apply, you may check none)".

Workers were asked about up to 4 query nodes per task. Each query appeared twice and in random order, so workers saw up to 8 sections of query nodes. A potential cause is considered to be a cause only if it was chosen both times.

[1] In the last level, which constructs the final causality graph, the causal nodes are individual shots. We consider all shots as possible causes, i.e. v are nodes from the original rather than pruned context tree.

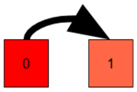
Causality relations between inner nodes

Due to the fact that the level of abstraction increases for higher levels of the context tree, sometimes it can be useful to look at causality relations between nodes in the same inner level of the tree. One example, which will be discussed in depth in the next section, in the creation of textual summaries for the input.

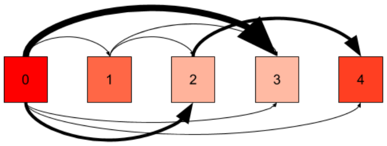
After constructing the causality graph for leaves level, a causality graph for inner levels can be induced by it. The definition for the causality graph for inner levels is: let *u*,*v* nodes in inner level *i* for the context tree. Let *leaves(v)* the group of leaves in the subtree that starts with *v*. An edge (*u*,*v*) exists in the causality of level *i* iff exists an edge (*u*,*w*) for some node *w* in *leaves(v)*. The weight of that edge is defined to be the maximal weight on an edge (*u*,*w*) for *w* in *leaves(v)*, and the score for each node *v* in level *i* is defined to be the maximal score among all leaves in its subtree.

For example, below are the causality graphs for all inner levels for the movie Innocence. Bigger arrows represent heavier edges, and the colors of the nodes depict their score.

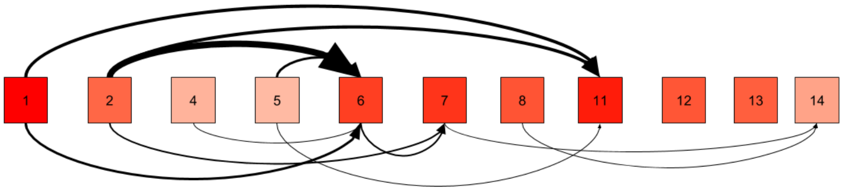
<u>Level 1</u>:



<u>Level 2</u>:



<u>Level 3</u>:



Steps 7–8: Shot Selection & Video Summary Creation

Using the final causality graph CG and the importance rating of nodes from the context tree, the final steps in our algorithm choose the leaf nodes (i.e. shots) and then stitch them together to create the final video summary.

Given a target video summary duration (or percentage of the input movie duration), our algorithm first sorts all the un-pruned context tree nodes according to their importance. Until this target duration is exceeded, we choose the most important node q not yet added for inclusion in the summary. We also include shots v with causality edges that point to q if the causality edge weight is above a threshold $w(v, q) > T_3$.

In our experiments, we set T_3 to be half of the largest edge weight in the graph. We continue this process until we exceed the desired movie duration.

After all shots contained in the summary are selected, the actual output video summary needs to be created. To do so, all the selected shots are simply concatenated in temporal order of the original movie. This simple method can produce a hard cut between shots. However, our evaluations show that such summaries are still adequate for a video summary, largely because we chose fine grained shots of the input video as building blocks.

Creation of textual summaries

Another application of our algorithm is creation of textual summaries for the input. The causality graph induces causality graphs for all levels of the context tree, and therefore allows the creation of textual summaries with different level of abstraction. Using higher levels yield textual summaries with higher level of abstraction.

The textual summaries are created using a similar mechanism based on the causality graphs for inner levels of the context tree: given a target length for the output summary (for example, number of words), and an inner level *i,* inner nodes at level *i* can be add to the textual summary in the same manner as the leaves, until their best summaries (which were crowd sourced in the up phase of the context tree creation) exceed the threshold length.

Results

We have implemented our algorithm and created video summaries for several movies. In the Evaluation Section (below), we compare our video summaries to the entire movie (ground truth) and to video summaries created using only the context tree (baseline) on three examples.

Implementation

We implemented our algorithm with the Amazon Mechanical Turk paid crowdsourcing platform. Examples for all tasks (HITs) can be seen in the supplemental materials, including payment per task.

Video Summaries

We selected three short films to evaluate our algorithm. The films are each challenging to summarize for different reasons. For all videos, we set the target length at 50%. Our video summaries can be seen in the supplemental materials, including more drastic summarization of 25%.

Cupido - Love is blind[2] is a 7.4 minute long animated movie about Cupid in his quest to make two people fall in love. Near the beginning of the movie, Cupid gets distracted by a butterfly and accidentally shoots two arrows into the same person, who then falls in love with himself. Cupid spends the rest of the movie trying to fix the mistake. This movie offers several challenges for our algorithm: (1) It has a long-term causality relationship, since an event at the beginning of the movie is solved at the end. (2) It contains shots which are difficult to translate into text with only local information. (3) Various artistic rendering styles are used which are not significant for the plot of the movie. Based on the 50% target duration, T_1 was set to 0.365. The resulting video summary is 3.8 minute long and can be seen in the supplemental materials. A graph of the importance of events as determined by the context tree can be seen in Figure 2

The top three levels of Cupid's context tree can be seen in Figure 3. The building block shots, their importance, the causality graph, and the shots used in the final video summary can be seen in Figure 3. The list of shots can be found in the appendix.

The causality graph (Figure 1) contains causality arrows which reach far across time. By pruning the possible nodes in the causality graph according to their context tree importance, we reduce the nodes for which we must obtain causality relationships. Without the causality graph, important shots would be left out of the video summary (see Evaluation).

[2] https://www.youtube.com/watch?v=Pe0jFDPHkzo

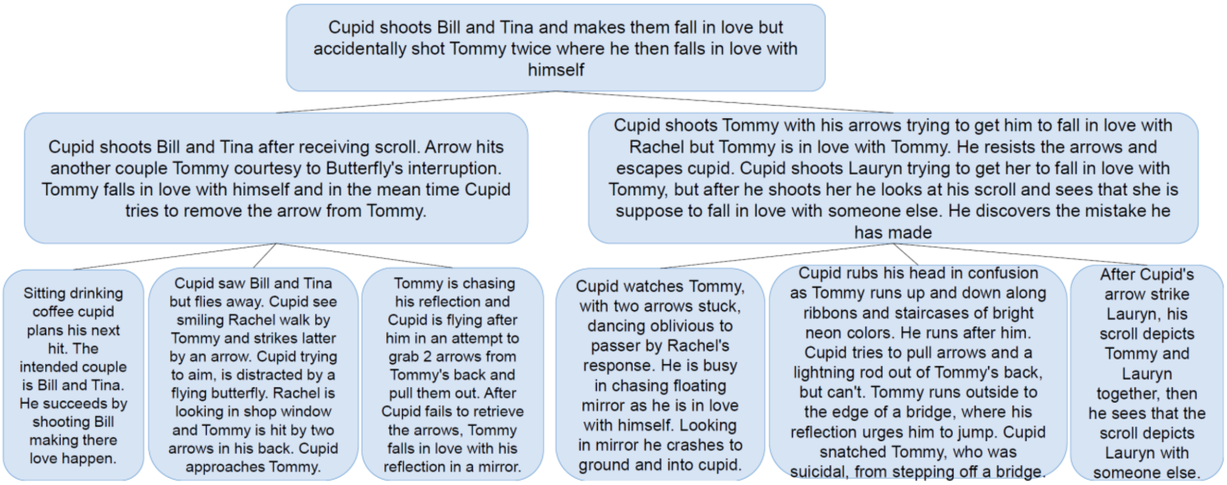

Cupid shoots Bill and Tina and makes them fall in love but accidentally shot Tommy twice where he then falls in love with himself
Cupid shoots Bill and Tina after receiving scroll. Arrow hits another couple Tommy courtesy to Butterfly's interruption. Tommy falls in love with himself and in the mean time Cupid tries to remove the arrow from Tommy.
Cupid shoots Tommy with his arrows trying to get him to fall in love with Rachel but Tommy is in love with Tommy. He resists the arrows and escapes cupid. Cupid shoots Lauryn trying to get her to fall in love with Tommy, but after he shoots her he looks at his scroll and sees that she is suppose to fall in love with someone else. He discovers the mistake he has made
Sitting drinking coffee cupid plans his next hit. The intended couple is Bill and Tina. He succeeds by shooting Bill making there love happen.
Cupid saw Bill and Tina but flies away. Cupid see smiling Rachel walk by Tommy and strikes latter by an arrow. Cupid trying to aim, is distracted by a flying butterfly. Rachel is looking in shop window and Tommy is hit by two arrows in his back. Cupid approaches Tommy.
Tommy is chasing his reflection and Cupid is flying after him in an attempt to grab 2 arrows from Tommy's back and pull them out. After Cupid fails to retrieve the arrows, Tommy falls in love with his reflection in a mirror.
Cupid watches Tommy, with two arrows stuck, dancing oblivious to passer by Rachel's response. He is busy in chasing floating mirror as he is in love with himself. Looking in mirror he crashes to ground and into cupid.
Cupid rubs his head in confusion as Tommy runs up and down along ribbons and staircases of bright neon colors. He runs after him. Cupid tries to pull arrows and a lightning rod out of Tommy's back, but can't. Tommy runs outside to the edge of a bridge, where his reflection urges him to jump. Cupid snatched Tommy, who was suicidal, from stepping off a bridge.
After Cupid's arrow strike Lauryn, his scroll depicts Tommy and Lauryn together, then he sees that the scroll depicts Lauryn with someone else.

Figure 3: The top three levels of Cupido's context tree. The summaries were written in the up phase, and so are context-less. The root node contains the entire movie summary.

Marimba[3] is a 7.1 minute long live-action movie with a moral message. The protagonist lives a comfortable and materialistic life. He is not sympathetic to a homeless man who lives across the street. After witnesses an act of kindness, he changes but is too late to help the homeless man. This movie is challenging because: (1) Seemingly unimportant events near the beginning of the movie become important because of the decisions the character makes near the end. (2) The moral message requires global context. Based on the 50% target duration, T_1 was set to 0.267. The resulting video summary is 3.4 minutes long and can be seen in the supplemental materials. A graph of the importance of events as determined by the context tree can be seen in Figure 2. The list of shots can be found in the appendix.

The causality graph and the shots used in the final video summary can be seen in Figure 1.

Innocence[4] is a 5 minute long live-action movie about a theft. A boy finds a phone on the beach and decides to steal it. The thief then loses his wallet, but someone returns it to him. The thief has a change of heart. The movie implies but does not explicitly show that the thief returns the phone to its original owner. This movie is challenging because (1) The cause-and-effect of the phone theft involves characters in different locations who never meet. (2) The change of heart requires global context to understand. (3) The return of the phone is implied but not shown. Based on the 50% target duration, T_1 was set to 0.65. The resulting video summary is 2.7 minutes long and can be seen in the supplemental materials. A graph of the importance of events as determined by the context tree can be seen in Figure 2. The list of shots can be found in the appendix.

The causality graph and the shots used in the final video summary are depicted in Figure 1.

[3] https://www.youtube.com/watch?v=PrXBiZD7a28

[4] https://www.youtube.com/watch?v=B2V8lKgjc9I

Textual Summaries

As well as video summaries, textual summaries were created for each movie. The summaries were built using the causality graph of the middle level of the context tree, and the target length of the summary was fixed to 10% of number of words in the whole description of the movie.

Our evaluations show that in most cases, simple concatenation of the summaries of inner nodes yield summaries with duplications. This causes the output textual summary to be longer than it can be.
To solve this problem we suggest either sending a HIT to a worker to eliminate redundancy, or using NLP algorithms.

In addition, the video summaries are easier to understand than the textual summaries. One picture's worth a thousand words, and a video's worth even more. Therefore, one who sees a shot will understand it fully, as oppose to one who reads the shot's description. In the same manner viewing the video summary of the video yields better understanding of the plot than reading the textual summary at any level.

Evaluation

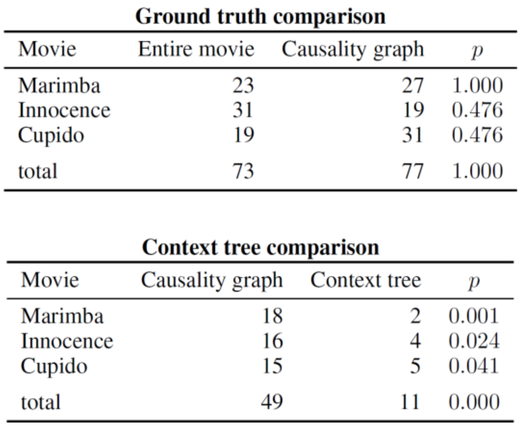
We evaluate our video summaries by comparing viewer comprehension against (a) the entire unabridged movie and (b) video summaries created using only the context tree. The goal of our algorithm is to create a summary video that shortens an input video while still capturing the plot of the input. We consider the following criteria to be important for a good summary:

1. Complete: The summary captures the entire plot. Someone who sees the summary but not the original video still understands the overall plot.
2. Concise: The summary is as short as possible.

Ground truth comparison To assess whether our video summaries successfully summarize the entire films, we recruited, for each film, two groups of $N = 5$ participants from Amazon Mechanical Turk to act as text summarizers. The entire movie group watched the unabridged movie and then wrote a text summary. The causality graph group watched our video summary and then wrote a text summary. A screenshot of the task (HIT) can be seen above. Participants were compensated \$0.25 to summarize the entire movie and \$0.20 to summarize our (shorter) video summaries.

We then recruited a second group of participants to act as raters. Raters were shown the original movie and two text summaries, one written by the entire movie group and one written by the causality graph group. Raters were asked to induce comprehension and, therefore, a thoughtful choice. Finally, raters were asked to choose the better summary. Every pairing of the text summaries was rated twice ($5 \cdot 5 \cdot 2 = 50$ total ratings). A screenshot of the task (HIT) can be seen above. Raters were compensated \$0.25 for each rating. Participants preferred the summaries written by the entire movie group 49% of the time (23, 31, and 19 votes for Marimba, Innocence, and Cupido, respectively) and summaries written by the causality graph group 51% of the time (27, 19, and 31 votes for Marimba, Innocence, and Cupido, respectively). These differences are not statistically significant (see Table 1), and we conclude that text summaries of the unabridged video and our video summaries are of similar quality. In other words, our video summaries successfully convey the plot of the movie to the degree where someone who watches our video summary and the original movie would produce an equivalent summary of events. The average length of text summaries for the entire movie group was 136 words (standard deviation of 66 words) and for the causality graph group was 84 words (standard deviation of 41 words). This suggests that our shorter videos lead to shorter text summaries which are nevertheless perceived to be better summaries by viewers of the entire movie.

Table 1: Evaluator preference. Top: The number of evaluators who preferred text summaries written from the entire movie versus text summaries written from our causality graph-based video summary. Bottom: The number of evaluators who preferred our causality graph-based video summary versus a video summary using only the context tree information. Statistical significance (p) values are computed with a two-sided binomial test and adjusted with the Holm-Bonferroni correction for multiple tests (Holm 1979)



Ground truth comparison

Movie	Entire movie	Causality graph	p
Marimba	23	27	1.000
Innocence	31	19	0.476
Cupido	19	31	0.476
total	73	77	1.000

Context tree comparison

Movie	Causality graph	Context tree	p
Marimba	18	2	0.001
Innocence	16	4	0.024
Cupido	15	5	0.041
total	49	11	0.000

Context tree comparison In our second experiment, we compared our video summaries to video summaries created with only information from the context tree. The context tree video summary was created using steps 1–5 and 8 of our algorithm. We set the target length of the video summary to 50%. The importance threshold T_1 was then automatically determined based on the total duration of shots above the T_1 and T_2 thresholds. The context tree only video summaries can be seen in the supplemental materials. We recruited a group of $N = 20$ participants to act as raters. Raters first watched the entire movie. Raters were asked to write a text summary of the movie. This was intended to induce comprehension and, therefore, a thoughtful choice. Finally, raters watched the two video summaries and indicated which they thought was better. A screenshot of the task (HIT) can be seen above. Raters were compensated $0.30. Raters preferred the summaries written by the causality graph group 82% of the time (18, 16, and 15 votes for Marimba, Innocence, and Cupido, respectively) and summaries written by the context tree group 18% of the time (2, 4, and 5 votes for Marimba, Innocence, and Cupido, respectively). These differences are all statistically significant ($p < 0.05$), taking into account the Holm-Bonferroni multiple comparison adjustment (Holm 1979). See Table 1. We conclude that our video summaries which use the causality graph better convey the plot of the movie than video summaries which use only information from the context tree. For example, shot 42 in which Cupid saves Tommy from falling off a ledge and, incidentally, detaches one arrow from Tommy's back was not important enough to be used in the context tree video summary. However, the causality graph correctly determined that this shot was a cause of the very important later shot 50 in which Tommy falls in love with Lauryn. Therefore it is included in our causality graph video summary, and the result is a summary without a large, unexplained gap in the plot.

Textual summaries

Cupido has a total of 873 words. Using the causality graph based on the middle level of the context tree, we created the following summary for threshold of 87 words:

Cupid while sitting on a roof gets a love contract between Bill and Tina. He finds them at a coffee shop and shoots them both with his arrows.
Cupid accidently shoots two arrows into Tommy, which causes Tommy to fall in love with himself. Cupid tries to get one of the arrows back, but Tommy runs away.
Cupid trys to make Tommy fall for Rachel but only succeeds in making him fall in love with himself. Cupid would like his arrows back!
Cupid tries to hook Lauren and Tommy up when it was suppose to be Lauren and Bob.

Marimba has a total of 850 words. Using the causality graph based on the middle level of the context tree, we created the following summary for threshold of 88 words:

Ali attends a party with Hannan and Bashar where they take a selfie. Bashar drops Ali off at home, they notice Mustafa digging through the trash and Ali says he doesn't know anything about Mustafa. The two go over plans for tomorrow.
Bashar is driving Ali home and he drops him off. Ali is putting his things on the table and taking his shoes off when he notices Mustafa digging in a dumpster.
Ali comes home and sees Mustafa, a poor beggar, digging through the trash in front of his house. Ali watched him for a moment and unlocked the door and went inside. While driving in the car Ali notices he is out of cigarettes.

Bashar goes to a food store and a boy takes something from the bakery before Bashar pays the bill. When Bashar comes out, he sees the boy giving the food to a woman who blesses him and a girl who is hungry.
While driving his car near a tunnel, Ali calls up Bashar and tells him he'll be 10 minutes late. He then drives by a park and sees people.
Bashar stopped in his car looking out at the 4 men who were standing together looking confusedly at an object on the ground, then Ali get a text.

Innocence has a total of 274 words. Using the causality graph based on the middle level of the context tree, we created the following summary for threshold of 27 words:

Patrick walks onto the beach and sees a cell phone on the sand. He looks to see if anyone is watching then takes it and turns the power off.
While Larry and Benny play ball on the beach, Patrick sits on the sand nearby. Patrick stands up, but his wallet falls on the sand, Benny sees this, picks up the wallet and runs to Patrick to give it back.
Jose lost his cell phone on the beach and Patrick found it. Patrick called Jose to let him know he found the cell phone and sits on the beach waiting for Jose.

Conclusion and Future Work

We have presented a human computation algorithm for video summarization that takes meaning into account and preserves the plot of the movie. Our algorithm takes advantage of the crowd by splitting the input into mutually exclusive shots and running tasks in parallel, so that each task is completed by a different worker with only local information. Our approach constructs a context tree to provide global context for isolated and distributed crowd workers and a causality graph to capture causal relationships important for understanding. Video summaries created with our algorithm led viewers to come away understanding the same plot as viewers of the entire, unabridged video. Moreover, our video summaries were judged to be better than video summaries created with information from the context tree. Our approach is able to overcome:

1. The difficulty of detecting causality locally. If event A causes an event B to happen, and different workers see them, then none of them will be able to understand the cause and effect relation.
2. The difficulty of summarizing movies with a non-linear plot. For example, movies where a viewer is only able to understand what she saw at the end.
3. Meta-understanding, like the moral of a movie, is sometimes hard to understand in general, and always harder to understand locally.
4. The difficulty of locally translating a shot into text. For example, an event which may seem unimportant out of context may be very important to the understanding of the movie.

There are limitations to our method. For one, we work at the granularity of shots. If one shot is very long and is chosen for the summary, we do not cut it and it may take up too large a portion of the summary. In the future, we plan to develop methods for rating and possibly shortening individual shots. There are also possibilities of confusion in the causality graph for very complicated plots or ones that are repetitive.

Other directions for the future are automate character identification, with either human or machine computation. We would also like to scale our algorithm to feature length films and adapt it to lecture videos. We wish to explore additional applications of the semantic information generated by our human computation video summarization algorithm. This information could be used to create different kinds of summaries, such as static arrangements (tapestries or storyboards or comics). This information could also be used to create video remixes.

Appendix: Shot Descriptions

Cupido

Shot number	Shot description
0	Rachel is walking passed a sign that changes to show the credits.
1	Cupid sit on a roof, sipping a drink.
2	Cupid sits on a roof drinking coffee as a scroll floats down and bonks him on his head.
3	Cupid sits on a rooftop drinking coffee while his scroll floats towards them. He then sees an image depicting Bill and Tina, with a heart stamped in between them
4	Cupid lets the Scroll float away.
5	From the top of a roof, Cupid flies away.
6	Cupid flies down above Tina who is sitting outside reading a newspaper
7	Cupid floats above Bill and Tina, while aiming his arrow towards them
8	Cupid shoots an arrow at both Bill and Tina in front of a cafe. Tina and Bill turn around and exchange smiles
9	Cupid glances down at Bill and Tina, while they turn around to look at each other at the outdoor cafe.
10	Cupid flying.
11	Cupid is hovering in the air above a building when a glowing scroll floats down from above him. He grabs the scroll, unravels it, and begins reading it.
12	Tommy is checking out his phone. Cupid is looking around and notices Rachel walking by.
13	Cupid draws his bow and is about to hit Tommy with an arrow.
14	Tommy is sitting on a bench looking at his phone when Cupid strikes him with an arrow. He looks in lust as Rachel walks by.
15	Cupid prepares his bow and arrow, when he is distracted by a beautiful blue butterfly.
16	Cupid accidentally hit Tommy with two arrows!
17	Tommy sits on a bench in a haze, Cupid adjusts his focus to Rachel.

18	Cupid tries to make Tommy look at Rachel, but Tommy instead sees his own reflection.
19	Cupid shoots Tommy with love arrows. Tommy falls in love with himself
20	Tommy watches his reflection adoringly as cupid unsuccessfully attempts to remove the arrow from his back
21	Cupid accidentally shoots Tommy with two arrows instead of one and Rachel with the other.
22	Tommy sprays himself with perfume as Rachel walks by. Cupid reacts in frustration.
23	Cupid chases after Tommy
24	Cupid flies down the street after Tommy to retrieve his arrows out of Tommy's back.
25	Tommy is running around a street with Cupid hanging on to the arrows in Tommy's back. Tommy stops to check his reflection out in a window.
26	Tommy see's himself in the mirror.
27	Tommy is dancing on stage as some tango music plays.
28	Cupid is watching Tommy dance. He has two arrows in his back.
29	Rachel begins to walk by in the background as the unknown man Salsa dances. Cupid sees her and tries to push the unknown man to dance with her.
30	Rachel walks by as Tommy spins in love.
31	Rachael walks down the street past rows of houses as music plays. Tommy runs by her with Cupid on his shoulders.
32	Tommy chases after a piece of paper as Cupid rides his back. Cupid sticks an arrow with him and then Tommy jumps off a ramp.
33	Tommy is falling through the air, with arrows is his back and looking at himself in the mirror. Once he reaches the ground, he sees Cupid
34	An upside down Tommy sees Cupid as Rachel walks on some crazy ramp.
35	Tommy is walking on a path, and Cupid is chasing after him, pulling on the arrows in Tommy's back.
36	Tommy is running along a path, and Cupid is being carried along behind him.
37	Cupid seems frustrated as there is some crazy lights around him.

38	Cupid tries to reach Tommy but he cant as Tommy runs up steps.
39	Cupid tries to reach Rachel but cant.
40	Cupid is looking up the staircases and follows a person into a magic mirror.
41	Tommy is about to jump off of a building.
42	Cupid saves Tommy from jumping off the bridge. By doing this he pulls one of the arrows out of Tommy's back.
43	Cupid is about to hit someone with an arrow, and he's looking around to see who walks by
44	Lauryn walks sadly past Cupid. Cupid is about to hit Lauryn with the arrow!
45	Cupid shoots Lauryn with an arrow. Lauryn immediately stops walking and stares straight ahead. Cupid seems pleased himself as he stands beside scroll, but then realizes that nothing will happen as the camera pans up to show both Lauryn and a previously shot Bob(?) are not looking at each other.
46	Scroll appears by Cupid, resulting him to pull out his quill and secretly begin writing in it.
47	Cupid writes down in his scroll with a feathered pen
48	Cupid glances at a picture of Tommy and Lauryn.
49	Cupid has shot Lauryn and Tommy who are walking away from each other on a quiet Parisian street. The arrows begin to do their work and the pair are about to turn to face one another. Cupid checks his scroll and celebrates another match made.
50	Cupid is dancing because Tommy and Lauryn seem to like each other. There's a clap of thunder and a Scroll appears.
51	Cupid looks at his scroll with depicts Lauryn with someone who is not Tommy

Marimba

Shot number	Shot description
0	Ali tells Bashar he will pick it up tomorrow. Bashar asks about Mustafa sitting on the side of the road.
1	Bashar drops Ali off somewhere. Before Ali gets out of the car Bashar asks him what's up with the old man (Mustafa) in the distance. Ali confesses he doesn't know but the man has been around for several days before saying goodbye. He then gets out of Bashar's car.
2	Ali (?) says goodbye to someone he's talking to on his cellular phone and

	exits his car. An exterior shot of a home. Ali begins to put on a black-and-white checkered shirt over an undershirt.
3	Ali puts on a black and white checkered shirt and, with a serious expression, looks at himself in the mirror.
4	Ali looks at himself in the mirror while buttoning up his shirt.
5	Ali is taking off his shirt in the mirror and receives a text.
6	Ali is looking at himself in the mirror and gets a text from Mirimba. He says he will come down. He asks Leni how long it takes to wipe a pair of shoes.
7	Ali gets into a car with Bashar and Hannan and asks if they brought any alcohol. Hannan shows him that she has a bottle in a brown bag.
8	Bashar slowly drives away and as his car leaves, it reveals in the background a man sitting on the corner of the road next to a black trash can.
9	Bashar, Ali, Hannan and many other people are socializing and enjoying themselves at a party at someone's home.
10	People are chatting and having a good time at a party.
11	There are a bunch of people hanging out at a party. Ali is there. Mustafa is sitting on the side of the road by himself.
12	There is a homeless man on the side of a road coughing.
13	There are a bunch of people in a living room, Ali is smoking.
14	Bashar tells a girl a secret at a party. Mustafa is alone on the side of the road. Ali is enjoying the same party that Bashar is at.
15	Someone is holding a bag of trash.
16	Ali takes a selfie of himself with Hannan and Bashar at a party. Bashar then drives Ali home in his car. The two pull up on a street while Mustafa digs through a trashcan across the street. Bashar jokingly tells Ali to go home so he can go home and urinate. As Ali is leaving the car Bashar asks him what their plans for tomorrow are.
17	Ali says he will pick up the car from the mechanic and get a new phone.

	Bashar asks about Mustafa sitting on the side of the road and Ali says he doesn't know whats up with him.
18	Bashar is driving a car and the car is stopped. Ali is in a passenger seat and gets out of the car. After Ali gets out of the car, Bashar drives away.
19	Ali appeared to have just got home. He set his things on the table and started to take off his shoes.
20	Ali picks up something from the ground. Standing up, he's surprised to see Mustafa digging in a dumpster.
21	Mustafa digs through a trash can while Ali looks on, slightly distressed looking.
22	Mustafa, who appears to be a beggar of some sort, is digging in a pile of refuse on the street in front of what I think is Ali's home. Ali looks at Mustafa pensive for several seconds before finally turning the key and walking into his home.
23	Ali is driving in a car and sees he is out of cigarettes.
24	A man (who I do not believe to be any of the characters listed) sells something to a young boy in what appears to be a convenience store. The cost of the item is 75 piasters.
25	A kid buys something at a convenience store.
26	Ali is waiting in line at a store while texting. A kid is buying something in front of him.
27	Bashar went to some food place and a boy took something from a bakery or restaurant.The boy went out before him and Bashar paid the bill, when he came out he saw the boy giving the eatables to a lady along with a girl who is hungry and the lady blessed the boy.
28	A kid gives something to an old lady and young girl and Ali watches.
29	Ali is watching a woman help a little girl.
30	Ali shows emotion.
31	Ali is driving in his car near a tunnel. He tells Bashar on his cellular phone that he'll be 10 minutes late because there's something he needs to do.
32	Someone tells Bashar they will be 10 minutes late. He then drives by a park and sees people.
33	Bashar is in his car, stopped, looking out the window at a group of 4 young men. The young men are standing around an object on the ground, looking at the object, seeming confused.

34	Ali sees some men looking at a bag of trash and then he receives a text.

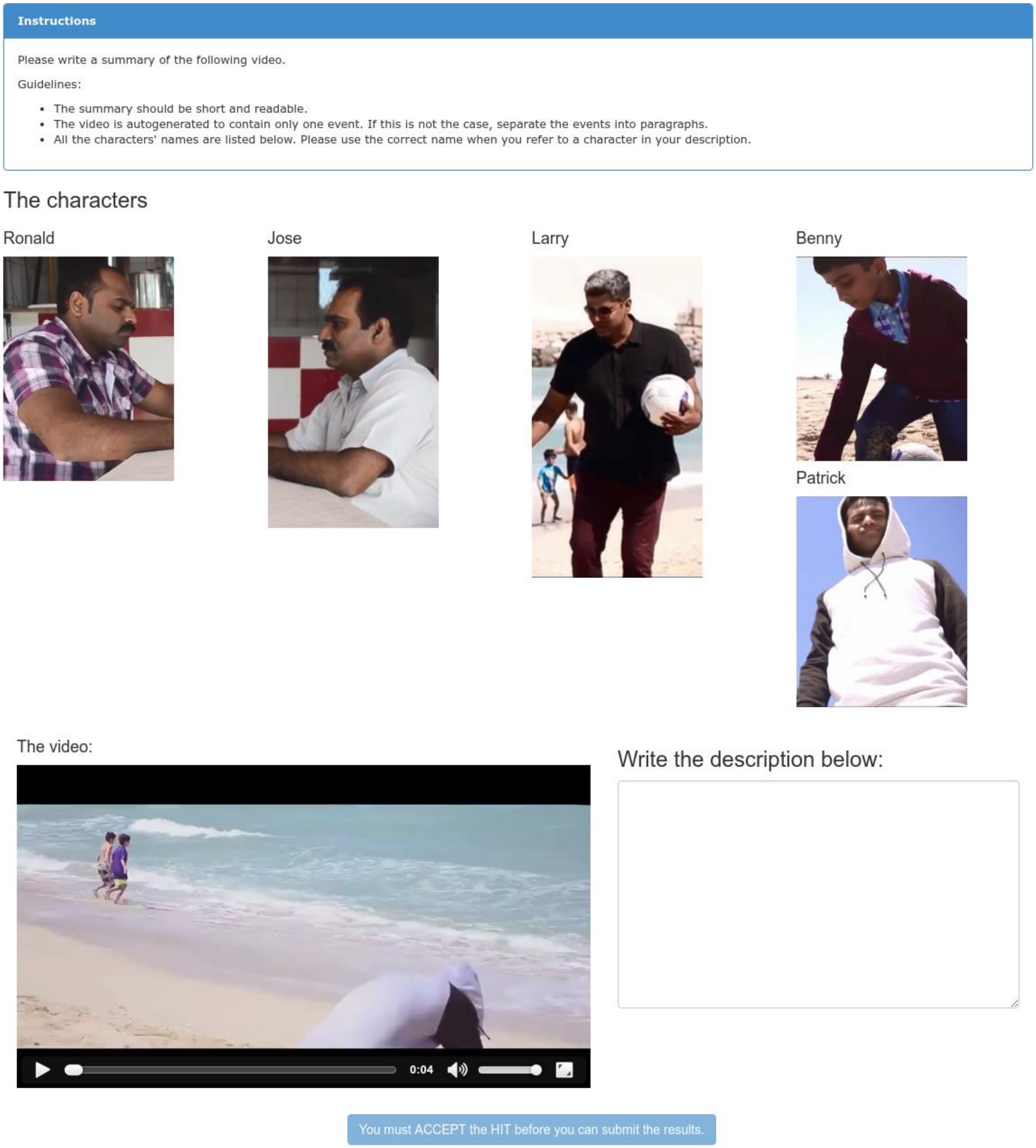
Innocence

Shot number	Shot description
0	Larry and Benny are sitting on the beach
1	Patrik walks on the beach. He notices a cell phone on the sand. He stopps, looks at it, and then looks to see if anyone is watching
2	Patrick picks up a cell phone, opens the power options menu and turns it off.
3	Patrick puts the phone he was holding in his hand in his pocket. The camera then zooms in on the restaurant Golden Chef. At this restaurant, Ronald is having breakfast with Jose.
4	Jose is reading the newspaper with somebody else.
5	Jose can't find his cell phone.
6	Jose talk to Ronald about losing his cellphone. Ronald give Jose a call from his phone, so he may attempt to find it but the Jose's phone was switch off.
7	Jose tries to make a phone call. Larry and benny plays on the beach with a ball. Patrik sits on the beach near them. When Patrik getets up, he forgets his wallet on the sand. Benny notices it - he picks it up and starts running towards Patrik
8	Benny finds Patrick's wallet on the beach and gives it to him.
9	Benny runs up and gives Larry a hug.
10	Patrick put on his hood and checks his phone at the beach.
11	Patrick is looking at his LG phone on the beach and he seems quite preoccupied with something that is distressing.
12	Jose gets a phone call. He asks the other side where he is, and says that he will come right away
13	Jose says happily to Ronald that he found his cell phone. Patrik hangs up and sits in the beach
14	Patrick is sitting on the beach.

Appendix: Algorithm HITs

Shot description (1 worker - 0.2 cents)

Used for translating the video shot into text



Instructions

Please write a summary of the following video.

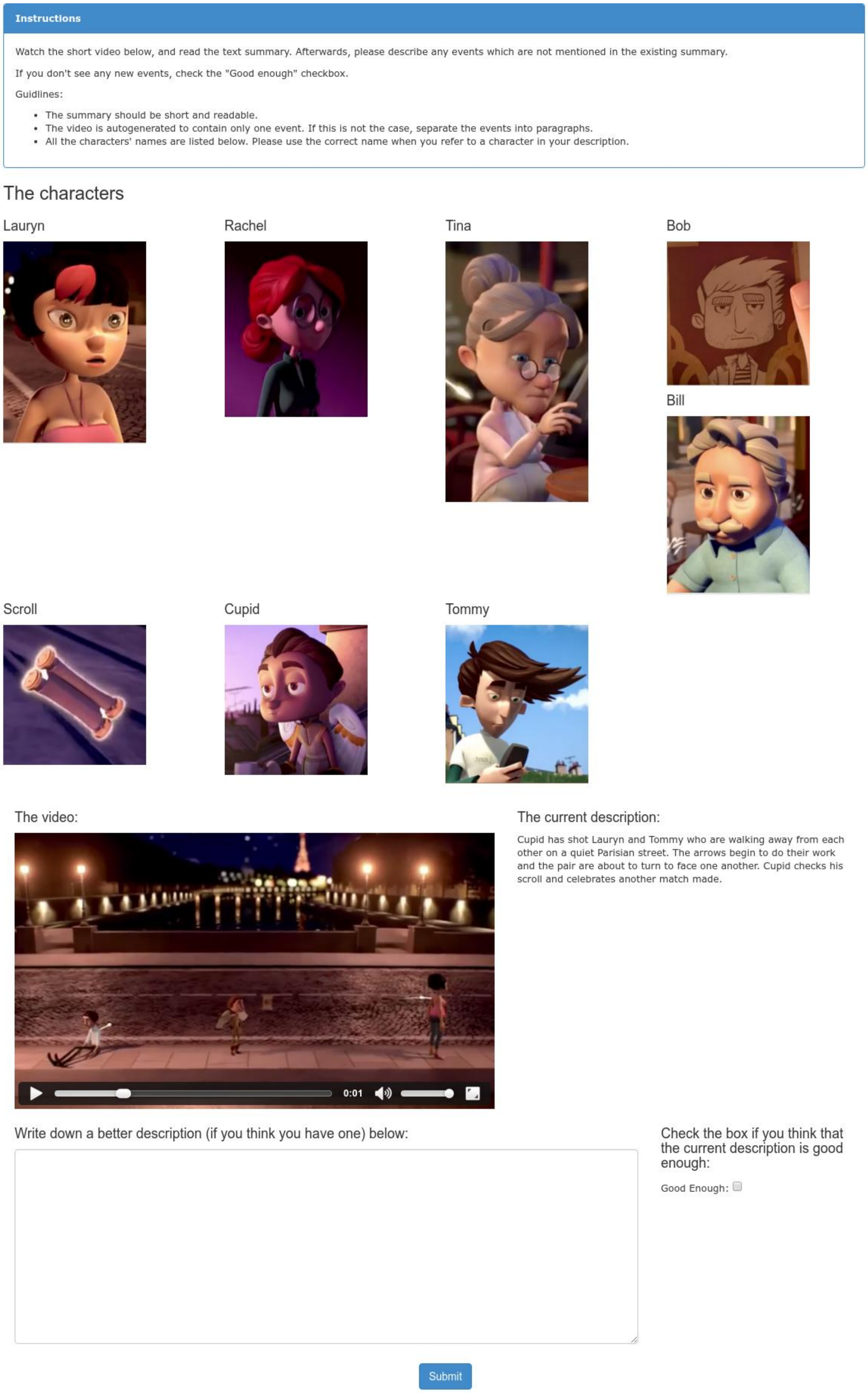
Guidelines:

- The summary should be short and readable.
- The video is autogenerated to contain only one event. If this is not the case, separate the events into paragraphs.
- All the characters' names are listed below. Please use the correct name when you refer to a character in your description.

The characters
Ronald
Jose
Larry
Benny
Patrick
The video:
0:04
Write the description below:
You must ACCEPT the HIT before you can submit the results.

Shot description improvement (0.15 cents)

Used to refine a shot description



Instructions

Watch the short video below, and read the text summary. Afterwards, please describe any events which are not mentioned in the existing summary.

If you don't see any new events, check the "Good enough" checkbox.

Guidlines:

- The summary should be short and readable.
- The video is autogenerated to contain only one event. If this is not the case, separate the events into paragraphs.
- All the characters' names are listed below. Please use the correct name when you refer to a character in your description.

The characters

Lauryn

Rachel

Tina

Bob

Bill

Scroll

Cupid

Tommy

The video:

The current description:

Cupid has shot Lauryn and Tommy who are walking away from each other on a quiet Parisian street. The arrows begin to do their work and the pair are about to turn to face one another. Cupid checks his scroll and celebrates another match made.

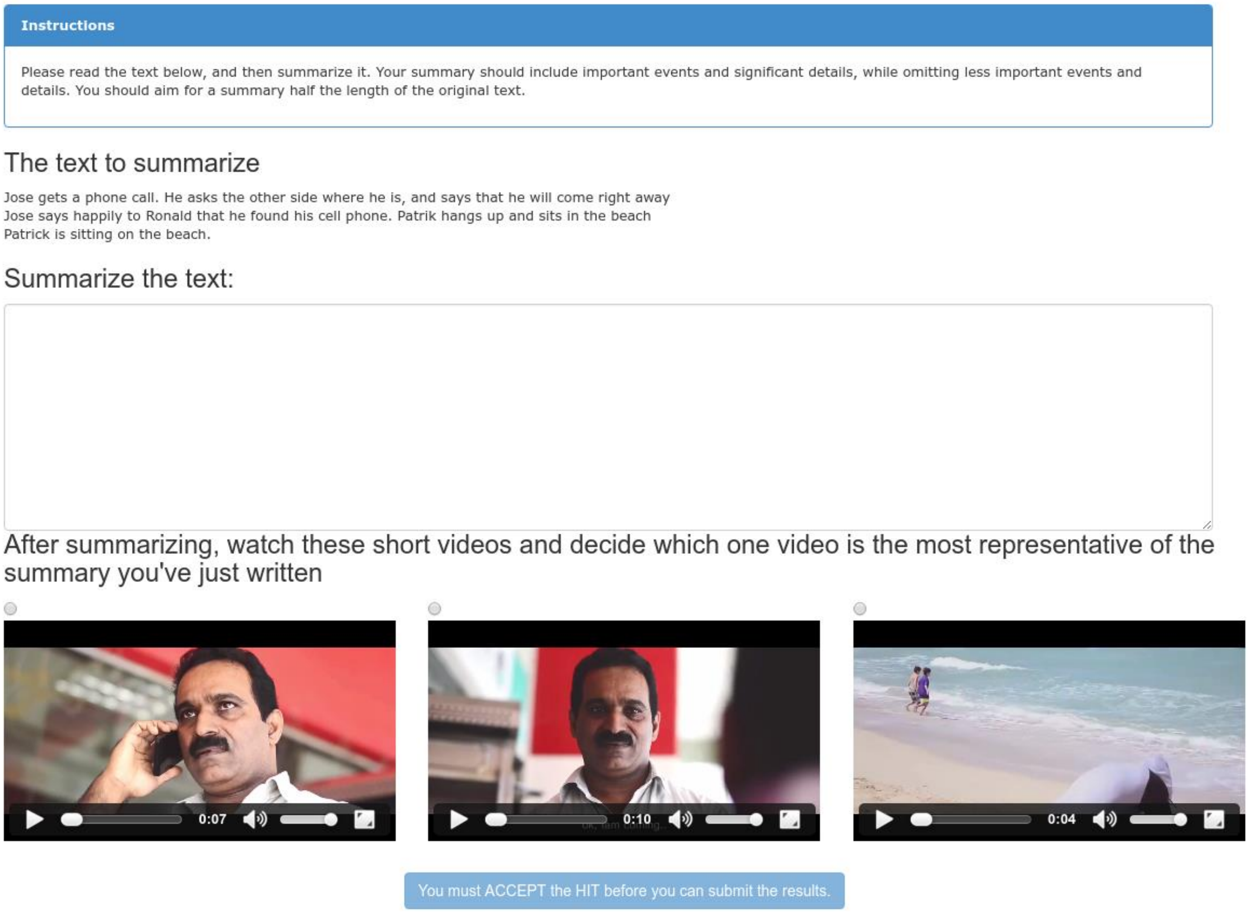
Write down a better description (if you think you have one) below:

Check the box if you think that the current description is good enough:

Good Enough: □

Context tree construction

Up phase - leaves (4 workers - 0.25 cents)
Used in the up-phase of the context trees algorithm from the shot descriptions to their parent (leaves level to create their parents)



Instructions

Please read the text below, and then summarize it. Your summary should include important events and significant details, while omitting less important events and details. You should aim for a summary half the length of the original text.

The text to summarize

Jose gets a phone call. He asks the other side where he is, and says that he will come right away
Jose says happily to Ronald that he found his cell phone. Patrik hangs up and sits in the beach
Patrick is sitting on the beach.

Summarize the text:

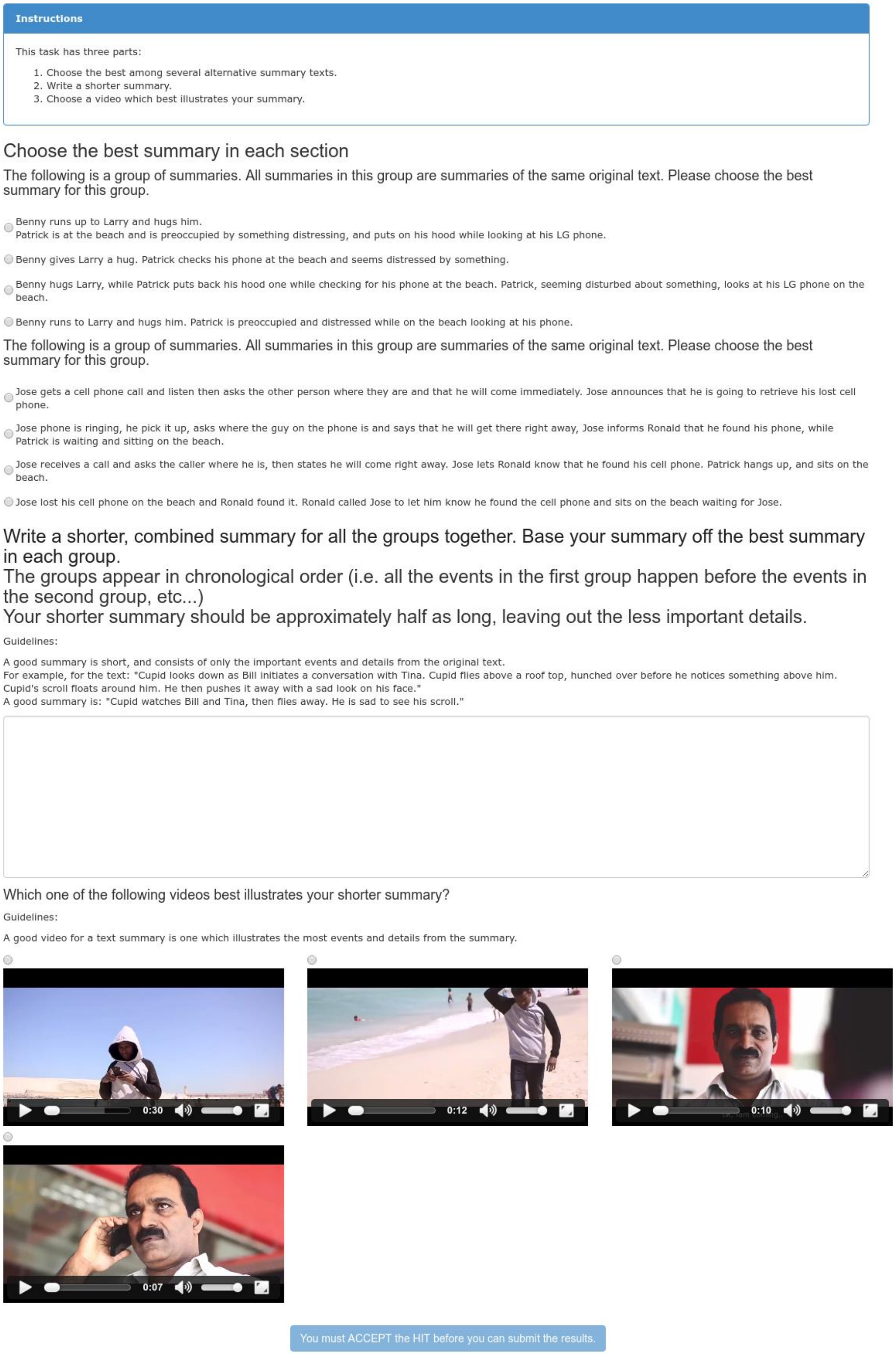
After summarizing, watch these short videos and decide which one video is the most representative of the summary you've just written

○ ○ ○

Context tree construction

Up phase - Inner nodes (4 workers - 0.3 cents)

Used in the up-phase of the context trees algorithm (every inner level, up until the root).



Instructions

This task has three parts:

1. Choose the best among several alternative summary texts.
2. Write a shorter summary.
3. Choose a video which best illustrates your summary.

Choose the best summary in each section

The following is a group of summaries. All summaries in this group are summaries of the same original text. Please choose the best summary for this group.

○ Benny runs up to Larry and hugs him.
Patrick is at the beach and is preoccupied by something distressing, and puts on his hood while looking at his LG phone.

○ Benny gives Larry a hug. Patrick checks his phone at the beach and seems distressed by something.

○ Benny hugs Larry, while Patrick puts back his hood one while checking for his phone at the beach. Patrick, seeming disturbed about something, looks at his LG phone on the beach.

○ Benny runs to Larry and hugs him. Patrick is preoccupied and distressed while on the beach looking at his phone.

The following is a group of summaries. All summaries in this group are summaries of the same original text. Please choose the best summary for this group.

○ Jose gets a cell phone call and listen then asks the other person where they are and that he will come immediately. Jose announces that he is going to retrieve his lost cell phone.

○ Jose phone is ringing, he pick it up, asks where the guy on the phone is and says that he will get there right away, Jose informs Ronald that he found his phone, while Patrick is waiting and sitting on the beach.

○ Jose receives a call and asks the caller where he is, then states he will come right away. Jose lets Ronald know that he found his cell phone. Patrick hangs up, and sits on the beach.

○ Jose lost his cell phone on the beach and Ronald found it. Ronald called Jose to let him know he found the cell phone and sits on the beach waiting for Jose.

Write a shorter, combined summary for all the groups together. Base your summary off the best summary in each group. The groups appear in chronological order (i.e. all the events in the first group happen before the events in the second group, etc...) Your shorter summary should be approximately half as long, leaving out the less important details.

Guidelines:

A good summary is short, and consists of only the important events and details from the original text.
For example, for the text: "Cupid looks down as Bill initiates a conversation with Tina. Cupid flies above a roof top, hunched over before he notices something above him. Cupid's scroll floats around him. He then pushes it away with a sad look on his face."
A good summary is: "Cupid watches Bill and Tina, then flies away. He is sad to see his scroll."

Which one of the following videos best illustrates your shorter summary?

Guidelines:

A good video for a text summary is one which illustrates the most events and details from the summary.

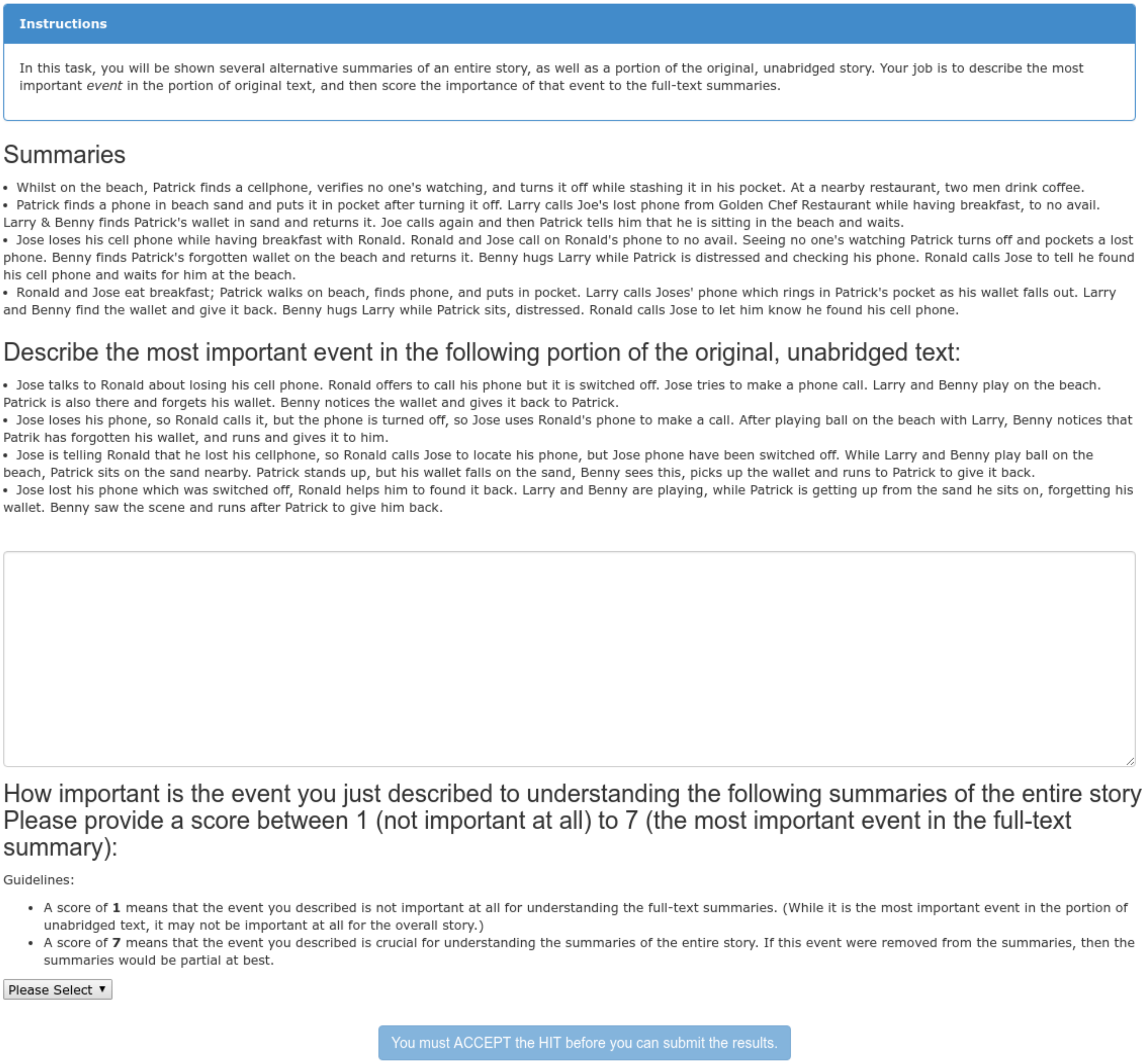
○ ○ ○ ○

You must ACCEPT the HIT before you can submit the results.

Context tree construction

Down phase - inner nodes (4 workers - 0.15 cents)

Used in the down-phase of the context tree algorithm (for every inner level, until the parents of the leaves).



Instructions

In this task, you will be shown several alternative summaries of an entire story, as well as a portion of the original, unabridged story. Your job is to describe the most important *event* in the portion of original text, and then score the importance of that event to the full-text summaries.

Summaries

- Whilst on the beach, Patrick finds a cellphone, verifies no one's watching, and turns it off while stashing it in his pocket. At a nearby restaurant, two men drink coffee.
- Patrick finds a phone in beach sand and puts it in pocket after turning it off. Larry calls Joe's lost phone from Golden Chef Restaurant while having breakfast, to no avail. Larry & Benny finds Patrick's wallet in sand and returns it. Joe calls again and then Patrick tells him that he is sitting in the beach and waits.
- Jose loses his cell phone while having breakfast with Ronald. Ronald and Jose call on Ronald's phone to no avail. Seeing no one's watching Patrick turns off and pockets a lost phone. Benny finds Patrick's forgotten wallet on the beach and returns it. Benny hugs Larry while Patrick is distressed and checking his phone. Ronald calls Jose to tell he found his cell phone and waits for him at the beach.
- Ronald and Jose eat breakfast; Patrick walks on beach, finds phone, and puts in pocket. Larry calls Joses' phone which rings in Patrick's pocket as his wallet falls out. Larry and Benny find the wallet and give it back. Benny hugs Larry while Patrick sits, distressed. Ronald calls Jose to let him know he found his cell phone.

Describe the most important event in the following portion of the original, unabridged text:

- Jose talks to Ronald about losing his cell phone. Ronald offers to call his phone but it is switched off. Jose tries to make a phone call. Larry and Benny play on the beach. Patrick is also there and forgets his wallet. Benny notices the wallet and gives it back to Patrick.
- Jose loses his phone, so Ronald calls it, but the phone is turned off, so Jose uses Ronald's phone to make a call. After playing ball on the beach with Larry, Benny notices that Patrik has forgotten his wallet, and runs and gives it to him.
- Jose is telling Ronald that he lost his cellphone, so Ronald calls Jose to locate his phone, but Jose phone have been switched off. While Larry and Benny play ball on the beach, Patrick sits on the sand nearby. Patrick stands up, but his wallet falls on the sand, Benny sees this, picks up the wallet and runs to Patrick to give it back.
- Jose lost his phone which was switched off, Ronald helps him to found it back. Larry and Benny are playing, while Patrick is getting up from the sand he sits on, forgetting his wallet. Benny saw the scene and runs after Patrick to give him back.

How important is the event you just described to understanding the following summaries of the entire story Please provide a score between 1 (not important at all) to 7 (the most important event in the full-text summary):

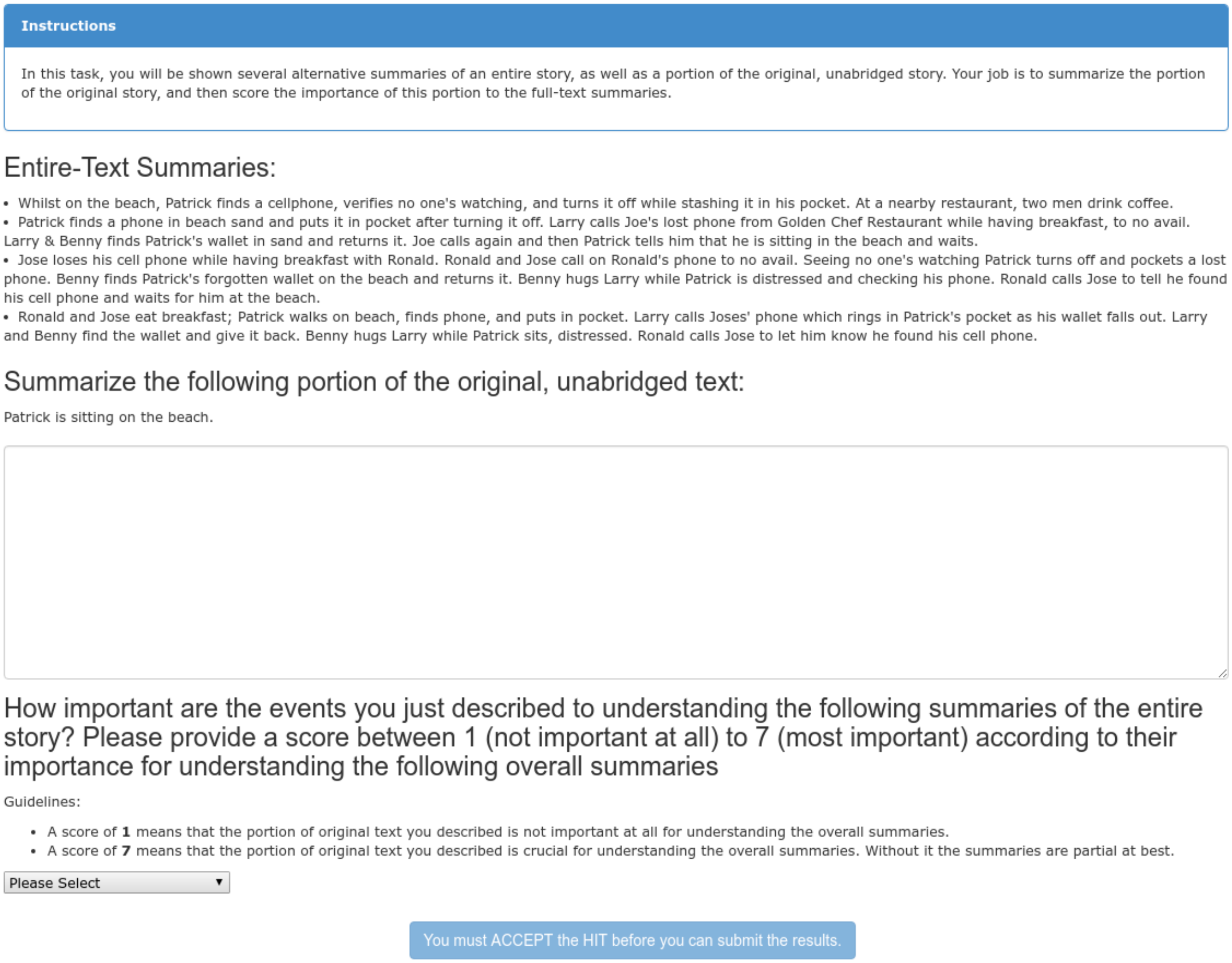
Guidelines:

- A score of **1** means that the event you described is not important at all for understanding the full-text summaries. (While it is the most important event in the portion of unabridged text, it may not be important at all for the overall story.)
- A score of **7** means that the event you described is crucial for understanding the summaries of the entire story. If this event were removed from the summaries, then the summaries would be partial at best.

Context tree construction

Down pahse - shot importance (4 workers - 0.15 cents)

Used in the down-phase of the context tree algorithm, moving from the parents of the leaves to the leaves.



Instructions

In this task, you will be shown several alternative summaries of an entire story, as well as a portion of the original, unabridged story. Your job is to summarize the portion of the original story, and then score the importance of this portion to the full-text summaries.

Entire-Text Summaries:

- Whilst on the beach, Patrick finds a cellphone, verifies no one's watching, and turns it off while stashing it in his pocket. At a nearby restaurant, two men drink coffee.
- Patrick finds a phone in beach sand and puts it in pocket after turning it off. Larry calls Joe's lost phone from Golden Chef Restaurant while having breakfast, to no avail. Larry & Benny finds Patrick's wallet in sand and returns it. Joe calls again and then Patrick tells him that he is sitting in the beach and waits.
- Jose loses his cell phone while having breakfast with Ronald. Ronald and Jose call on Ronald's phone to no avail. Seeing no one's watching Patrick turns off and pockets a lost phone. Benny finds Patrick's forgotten wallet on the beach and returns it. Benny hugs Larry while Patrick is distressed and checking his phone. Ronald calls Jose to tell he found his cell phone and waits for him at the beach.
- Ronald and Jose eat breakfast; Patrick walks on beach, finds phone, and puts in pocket. Larry calls Joses' phone which rings in Patrick's pocket as his wallet falls out. Larry and Benny find the wallet and give it back. Benny hugs Larry while Patrick sits, distressed. Ronald calls Jose to let him know he found his cell phone.

Summarize the following portion of the original, unabridged text:

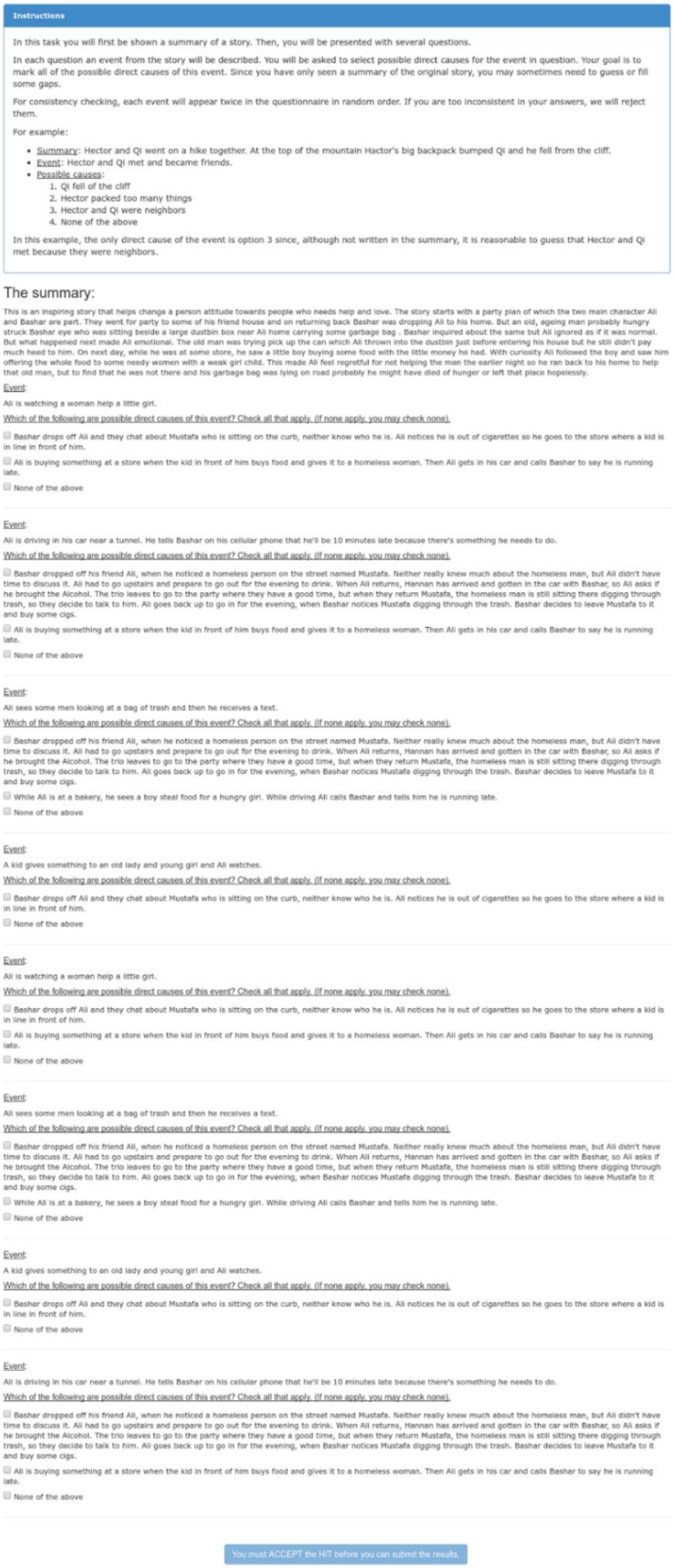
Patrick is sitting on the beach.

How important are the events you just described to understanding the following summaries of the entire story? Please provide a score between 1 (not important at all) to 7 (most important) according to their importance for understanding the following overall summaries

Guidelines:

- A score of **1** means that the portion of original text you described is not important at all for understanding the overall summaries.
- A score of **7** means that the portion of original text you described is crucial for understanding the overall summaries. Without it the summaries are partial at best.

Causality (5 workers - 0.25 cents)



Instructions

In this task you will first be shown a summary of a story. Then, you will be presented with several questions.

In each question an event from the story will be described. You will be asked to select possible direct causes for the event in question. Your goal is to mark all of the possible direct causes of this event. Since you have only seen a summary of the original story, you may sometimes need to guess or fill some gaps.

For consistency checking, each event will appear twice in the questionnaire in random order. If you are too inconsistent in your answers, we will reject them.

For example:

- Summary: Hector and Qi went on a hike together. At the top of the mountain Hactor's big backpack bumped Qi and he fell from the cliff.
- Event: Hector and Qi met and became friends.
- Possible causes:
  1. Qi fell of the cliff
  2. Hector packed too many things
  3. Hector and Qi were neighbors
  4. None of the above

In this example, the only direct cause of the event is option 3 since, although not written in the summary, it is reasonable to guess that Hector and Qi met because they were neighbors.

The summary:

This is an inspiring story that helps change a person attitude towards people who needs help and love. The story starts with a party plan of which the two main character Ali and Bashar are part. They went for party to some of his friend house and on returning back Bashar was dropping Ali to his home. But an old, ageing man probably hungry struck Bashar eye who was sitting beside a large dustbin box near Ali home carrying some garbage bag . Bashar inquired about the same but Ali ignored as if it was normal. But what happened next made Ali emotional. The old man was trying pick up the can which Ali thrown into the dustbin just before entering his house but he still didn't pay much heed to him. On next day, while he was at some store, he saw a little boy buying some food with the little money he had. With curiosity Ali followed the boy and saw him offering the whole food to some needy women with a weak girl child. This made Ali feel regretful for not helping the man the earlier night so he ran back to his home to help that old man, but to find that he was not there and his garbage bag was lying on road probably he might have died of hunger or left that place hopelessly.

Event:

Ali is watching a woman help a little girl.

Which of the following are possible direct causes of this event? Check all that apply. (If none apply, you may check none).

□ Bashar drops off Ali and they chat about Mustafa who is sitting on the curb, neither know who he is. Ali notices he is out of cigarettes so he goes to the store where a kid is in line in front of him.

□ Ali is buying something at a store when the kid in front of him buys food and gives it to a homeless woman. Then Ali gets in his car and calls Bashar to say he is running late.

□ None of the above

Event:

Ali is driving in his car near a tunnel. He tells Bashar on his cellular phone that he'll be 10 minutes late because there's something he needs to do.

Which of the following are possible direct causes of this event? Check all that apply. (If none apply, you may check none).

□ Bashar dropped off his friend Ali, when he noticed a homeless person on the street named Mustafa. Neither really knew much about the homeless man, but Ali didn't have time to discuss it. Ali had to go upstairs and prepare to go out for the evening to drink. When Ali returns, Hannan has arrived and gotten in the car with Bashar, so Ali asks if he brought the Alcohol. The trio leaves to go to the party where they have a good time, but when they return Mustafa, the homeless man is still sitting there digging through trash, so they decide to talk to him. Ali goes back up to go in for the evening, when Bashar notices Mustafa digging through the trash. Bashar decides to leave Mustafa to it and buy some cigs.

□ Ali is buying something at a store when the kid in front of him buys food and gives it to a homeless woman. Then Ali gets in his car and calls Bashar to say he is running late.

□ None of the above

Event:

Ali sees some men looking at a bag of trash and then he receives a text.

Which of the following are possible direct causes of this event? Check all that apply. (If none apply, you may check none).

□ Bashar dropped off his friend Ali, when he noticed a homeless person on the street named Mustafa. Neither really knew much about the homeless man, but Ali didn't have time to discuss it. Ali had to go upstairs and prepare to go out for the evening to drink. When Ali returns, Hannan has arrived and gotten in the car with Bashar, so Ali asks if he brought the Alcohol. The trio leaves to go to the party where they have a good time, but when they return Mustafa, the homeless man is still sitting there digging through trash, so they decide to talk to him. Ali goes back up to go in for the evening, when Bashar notices Mustafa digging through the trash. Bashar decides to leave Mustafa to it and buy some cigs.

□ While Ali is at a bakery, he sees a boy steal food for a hungry girl. While driving Ali calls Bashar and tells him he is running late.

□ None of the above

Event:

A kid gives something to an old lady and young girl and Ali watches.

Which of the following are possible direct causes of this event? Check all that apply. (If none apply, you may check none).

□ Bashar drops off Ali and they chat about Mustafa who is sitting on the curb, neither know who he is. Ali notices he is out of cigarettes so he goes to the store where a kid is in line in front of him.

□ None of the above

Event:

Ali is watching a woman help a little girl.

Which of the following are possible direct causes of this event? Check all that apply. (If none apply, you may check none).

□ Bashar drops off Ali and they chat about Mustafa who is sitting on the curb, neither know who he is. Ali notices he is out of cigarettes so he goes to the store where a kid is in line in front of him.

□ Ali is buying something at a store when the kid in front of him buys food and gives it to a homeless woman. Then Ali gets in his car and calls Bashar to say he is running late.

□ None of the above

Event:

Ali sees some men looking at a bag of trash and then he receives a text.

Which of the following are possible direct causes of this event? Check all that apply. (If none apply, you may check none).

□ Bashar dropped off his friend Ali, when he noticed a homeless person on the street named Mustafa. Neither really knew much about the homeless man, but Ali didn't have time to discuss it. Ali had to go upstairs and prepare to go out for the evening to drink. When Ali returns, Hannan has arrived and gotten in the car with Bashar, so Ali asks if he brought the Alcohol. The trio leaves to go to the party where they have a good time, but when they return Mustafa, the homeless man is still sitting there digging through trash, so they decide to talk to him. Ali goes back up to go in for the evening, when Bashar notices Mustafa digging through the trash. Bashar decides to leave Mustafa to it and buy some cigs.

□ While Ali is at a bakery, he sees a boy steal food for a hungry girl. While driving Ali calls Bashar and tells him he is running late.

□ None of the above

Event:

A kid gives something to an old lady and young girl and Ali watches.

Which of the following are possible direct causes of this event? Check all that apply. (If none apply, you may check none).

□ Bashar drops off Ali and they chat about Mustafa who is sitting on the curb, neither know who he is. Ali notices he is out of cigarettes so he goes to the store where a kid is in line in front of him.

□ None of the above

Event:

Ali is driving in his car near a tunnel. He tells Bashar on his cellular phone that he'll be 10 minutes late because there's something he needs to do.

Which of the following are possible direct causes of this event? Check all that apply. (If none apply, you may check none).

□ Bashar dropped off his friend Ali, when he noticed a homeless person on the street named Mustafa. Neither really knew much about the homeless man, but Ali didn't have time to discuss it. Ali had to go upstairs and prepare to go out for the evening to drink. When Ali returns, Hannan has arrived and gotten in the car with Bashar, so Ali asks if he brought the Alcohol. The trio leaves to go to the party where they have a good time, but when they return Mustafa, the homeless man is still sitting there digging through trash, so they decide to talk to him. Ali goes back up to go in for the evening, when Bashar notices Mustafa digging through the trash. Bashar decides to leave Mustafa to it and buy some cigs.

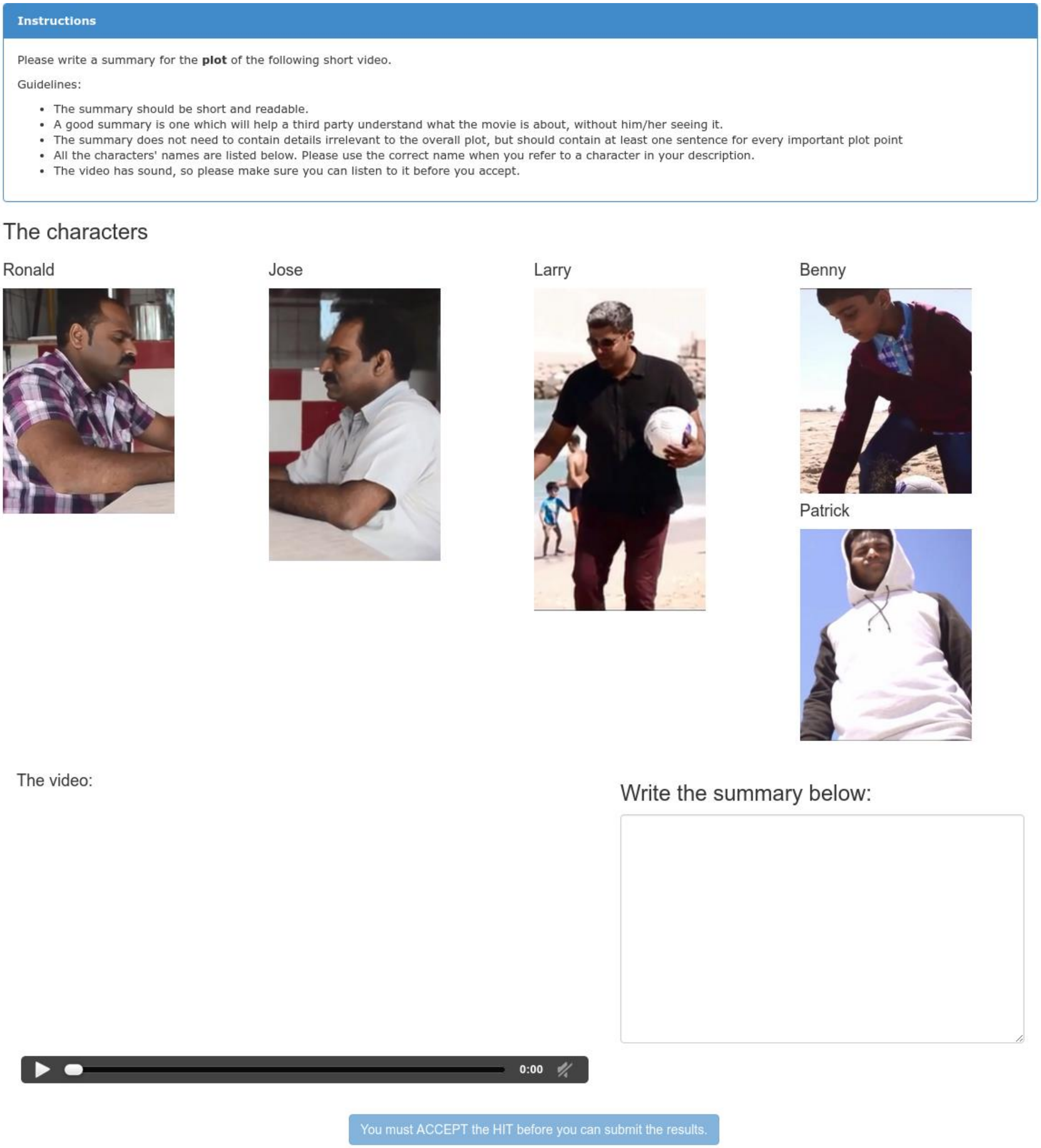
□ Ali is buying something at a store when the kid in front of him buys food and gives it to a homeless woman. Then Ali gets in his car and calls Bashar to say he is running late.

□ None of the above

Appendix: Evaluation HITs

Evaluations

Write a summary of a short video (5 workers - 0.25 cents)



Instructions

Please write a summary for the **plot** of the following short video.

Guidelines:

- The summary should be short and readable.
- A good summary is one which will help a third party understand what the movie is about, without him/her seeing it.
- The summary does not need to contain details irrelevant to the overall plot, but should contain at least one sentence for every important plot point
- All the characters' names are listed below. Please use the correct name when you refer to a character in your description.
- The video has sound, so please make sure you can listen to it before you accept.

The characters

Ronald

Jose

Larry

Benny

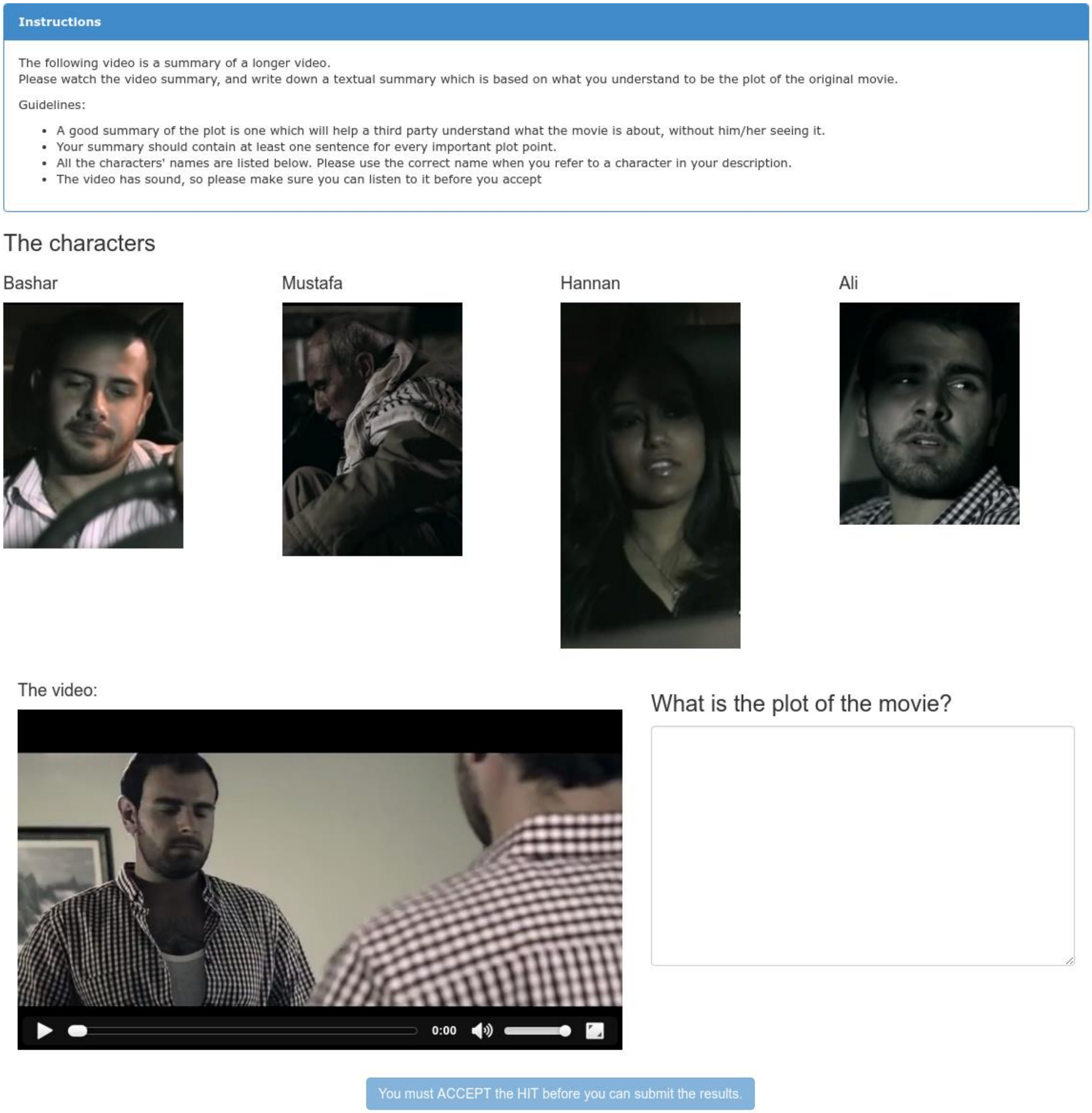
Patrick

The video:

Write the summary below:

Evaluations

Decide what is the plot based on the summary (5 workers - 0.20 cents)



Instructions

The following video is a summary of a longer video.
Please watch the video summary, and write down a textual summary which is based on what you understand to be the plot of the original movie.

Guidelines:

- A good summary of the plot is one which will help a third party understand what the movie is about, without him/her seeing it.
- Your summary should contain at least one sentence for every important plot point.
- All the characters' names are listed below. Please use the correct name when you refer to a character in your description.
- The video has sound, so please make sure you can listen to it before you accept

The characters

Bashar

Mustafa

Hannan

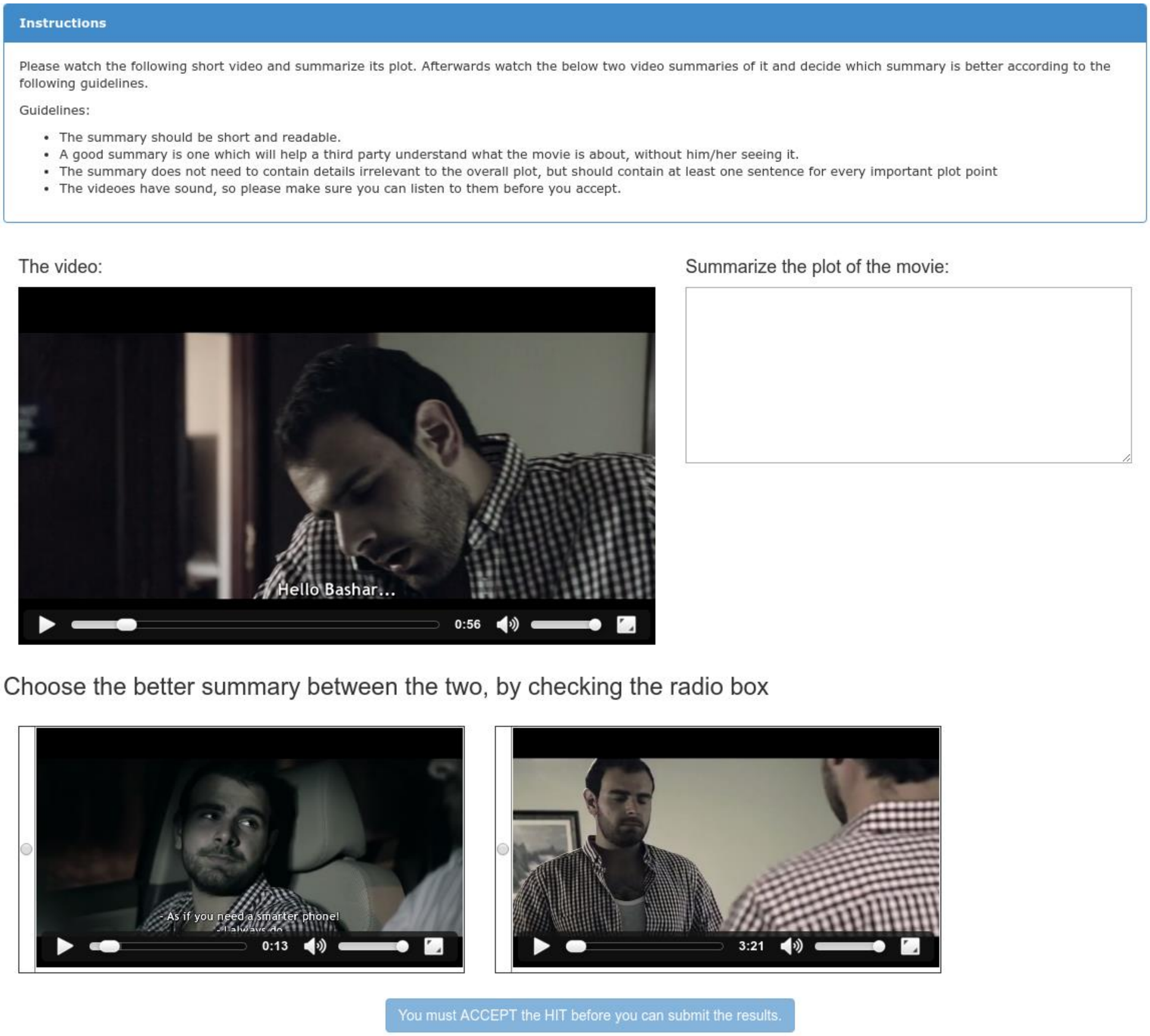
Ali

The video:

What is the plot of the movie?

Evaluations

Context tree Vs. Causality Graph (20 workers - 0.3 cents)



Instructions

Please watch the following short video and summarize its plot. Afterwards watch the below two video summaries of it and decide which summary is better according to the following guidelines.

Guidelines:

- The summary should be short and readable.
- A good summary is one which will help a third party understand what the movie is about, without him/her seeing it.
- The summary does not need to contain details irrelevant to the overall plot, but should contain at least one sentence for every important plot point
- The videoes have sound, so please make sure you can listen to them before you accept.

The video:

Summarize the plot of the movie:

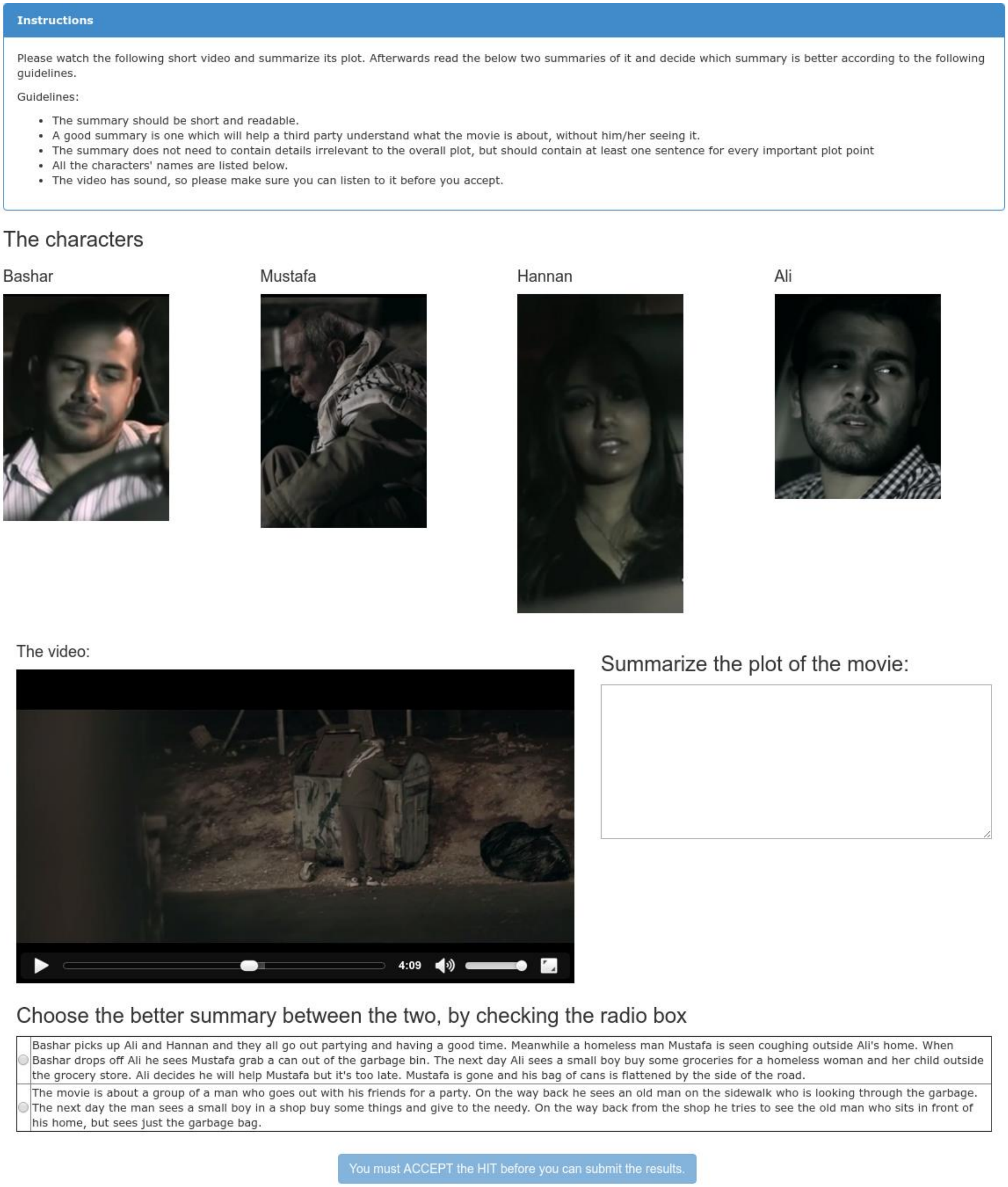
Choose the better summary between the two, by checking the radio box

○ ○

Hello Bashar...
0:56
- As if you need a smarter phone!
- I always do
0:13
3:21
You must ACCEPT the HIT before you can submit the results.

Evaluations

Full movie Vs. Causality graph (2 workers - 0.25 cents)



Instructions

Please watch the following short video and summarize its plot. Afterwards read the below two summaries of it and decide which summary is better according to the following guidelines.

Guidelines:

- The summary should be short and readable.
- A good summary is one which will help a third party understand what the movie is about, without him/her seeing it.
- The summary does not need to contain details irrelevant to the overall plot, but should contain at least one sentence for every important plot point
- All the characters' names are listed below.
- The video has sound, so please make sure you can listen to it before you accept.

The characters

Bashar

Mustafa

Hannan

Ali

The video:

Summarize the plot of the movie:

Choose the better summary between the two, by checking the radio box

○	Bashar picks up Ali and Hannan and they all go out partying and having a good time. Meanwhile a homeless man Mustafa is seen coughing outside Ali's home. When Bashar drops off Ali he sees Mustafa grab a can out of the garbage bin. The next day Ali sees a small boy buy some groceries for a homeless woman and her child outside the grocery store. Ali decides he will help Mustafa but it's too late. Mustafa is gone and his bag of cans is flattened by the side of the road.
○	The movie is about a group of a man who goes out with his friends for a party. On the way back he sees an old man on the sidewalk who is looking through the garbage. The next day the man sees a small boy in a shop buy some things and give to the needy. On the way back from the shop he tries to see the old man who sits in front of his home, but sees just the garbage bag.

Bibliography

[Assa, Caspi, and Cohen-Or 2005] Assa, J.; Caspi, Y.; and Cohen-Or, D. 2005. Action synopsis: pose selection and illustration. ACM Transactions on Graphics (TOG) 24(3):667–676.

[Bernstein et al. 2010] Bernstein, M. S.; Little, G.; Miller, R. C.; Hartmann, B.; Ackerman, M. S.; Karger, D. R.; Crowell, D.; and Panovich, K. 2010. Soylent: a word processor with a crowd inside. In User interface software and technology (UIST), 313–322. ACM.

[Bernstein et al. 2011] Bernstein, M. S.; Brandt, J.; Miller, R. C.; and Karger, D. R. 2011. Crowds in two seconds: Enabling realtime crowd-powered interfaces. In User interface software and technology (UIST), 33–42. ACM.

[Boreczky and Rowe 1996] Boreczky, J. S., and Rowe, L. A. 1996. Comparison of video shot boundary detection techniques. Journal of Electronic Imaging 5(2):122–128.

[Di Salvo, Giordano, and Kavasidis 2013] Di Salvo, R.; Giordano, D.; and Kavasidis, I. 2013. A crowdsourcing approach to support video annotation. In Proceedings of the International Workshop on Video and Image Ground Truth in Computer Vision Applications, 8. ACM.

[Gingold, Shamir, and Cohen-Or 2012] Gingold, Y.; Shamir, A.; and Cohen-Or, D. 2012. Micro perceptual human computation for visual tasks. ACM Transactions on Graphics (TOG) 31(5):119.

[Goldman and Brandt 2011] Goldman, D., and Brandt, J. 2011. Task decomposition and human computation in graphics and vision. In ACM CHI 2011 Workshop on Crowdsourcing and Human computation.

[Gupta et al. 2009] Gupta, A.; Srinivasan, P.; Shi, J.; and Davis, L. S. 2009. Understanding videos, constructing plots: learning a visually grounded storyline model from annotated videos. In Computer Vision and Pattern Recognition (CVPR), 2012–2019. IEEE.

[He et al. 1999] He, L.; Sanocki, E.; Gupta, A.; and Grudin, J. 1999. Auto-summarization of audio-video presentations. In Multimedia, 489–498. ACM.

[Heck, Wallick, and Gleicher 2007] Heck, R.; Wallick, M.; and Gleicher, M. 2007. Virtual videography. ACM Trans. Multimedia Comput. Commun. Appl. 3(1).

[Herlocker, Konstan, and Riedl 2002] Herlocker, J.; Konstan, J. A.; and Riedl, J. 2002. An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms. Information retrieval 5(4):287–310.

[Holm 1979] Holm, S. 1979. A simple sequentially rejective multiple test procedure. Scandinavian journal of statistics 65–70.

[Kelso, Weyhrauch, and Bates 1993] Kelso, M.T.; Weyhrauch, P.; and Bates, J. 1993. Dramatic presence. PRESENCE: Teleoperators & Virtual Environments 2(1):1–15.

[Kim et al. 2014] Kim, J.; Nguyen, P. T.; Weir, S.; Guo, P. J.; Miller, R. C.; and Gajos, K. Z. 2014. Crowdsourcing step-by- step information extraction to enhance existing how-to videos. In SIGCHI Conference on Human Factors in Computing Systems, 4017–4026. ACM.

[Li et al. 2013] Li, B.; Lee-Urban, S.; Johnston, G.; and Riedl, M. 2013. Story generation with crowdsourced plot graphs. In AAAI Conference on Artificial Intelligence.

[Little et al. 2010] Little, G.; Chilton, L. B.; Goldman, M.; and Miller, R. C. 2010. TurKit: human computation algorithms on mechanical turk. In User interface software and technology (UIST), 57–66. ACM.

[Pavel, Hartmann, and Agrawala 2014] Pavel, A.; Hartmann, B.; and Agrawala, M. 2014. Video digests: A browsable, skimmable format for informational lecture videos. In User interface software and technology (UIST), 573–582. ACM.

[Pritch, Rav-Acha, and Peleg 2008] Pritch, Y.; Rav-Acha, A.; and Peleg, S. 2008. Nonchronological video synopsis and indexing. Pattern Analysis and Machine Intelligence, IEEE Transactions on 30(11):1971–1984.

[Shin et al. 2015] Shin, H. V.; Berthouzoz, F.; Li, W.; and Durand, F. 2015. Visual transcripts: Lecture notes from blackboard-style lecture videos. ACM Trans. Graph. 34(6):240:1–240:10.

[Smith and Kanade 1998] Smith, M., and Kanade, T. 1998. Video skimming and characterization through the combination of image and language understanding. In IEEE International Workshop on Content-Based Access of Image and Video Database, 61–70. IEEE.

[Truong and Venkatesh 2007] Truong, B. T., and Venkatesh, S. 2007. Video abstraction: A systematic review and classification. ACM Trans. Multimedia Comput. Commun. ppl. 3(1).

[Uchihashi et al. 1999] Uchihashi, S.; Foote, J.; Girgensohn, A.; and Boreczky, J. 1999. Video manga: generating semantically meaningful video summaries. In Proceedings of the seventh ACM international conference on Multimedia (Part 1), 383–392. ACM.

[Verroios and Bernstein 2014] Verroios, V., and Bernstein, M. S. 2014. Context trees: Crowdsourcing global understanding from local views. In Second AAAI Conference on Human Computation and Crowdsourcing.

[Weir et al. 2015] Weir, S.; Kim, J.; Gajos, K. Z.; and Miller, R. C. 2015. Learnersourcing subgoal labels for how-to videos. In Computer Supported Cooperative Work & Social Computing (CSCW), 405–416. ACM.

[Weyhrauch 1997] Weyhrauch, P. 1997. Guiding Interactive Drama. Ph.D. Dissertation, Carnegie Mellon University.

[Wu, Thawonmas, and Chen 2011] Wu, S.-Y.; Thawonmas, R.; and Chen, K.-T. 2011. Video summarization via crowdsourcing. In CHI'11 Extended Abstracts on Human Factors in Computing Systems, 1531–1536. ACM.

[Young 1999] Young, R. M. 1999. Notes on the use of plan structures in the creation of interactive plot. In AAAI Fall Symposium on Narrative Intelligence.

[Zhuang et al. 1998] Zhuang, Y.; Rui, Y.; Huang, T. S.; and Mehrotra, S. 1998. Adaptive key frame extraction using unsupervised clustering. In Image Processing (ICIP), volume 1, 866–870. IEEE.

תקציר

סיכום וידאו זהו כלי שימושי ליישומים רבים, כמו למשל חיפוש בוידאו. אולם, סיכום אוטומאטי זוהי משימה קשה במיוחד עקב בעיות חישוביות כמו זיהוי גבולות סצינה, קביעת מי או מה קורה בכל סצינה, זיהוי מילים ופעולות, ומעל הכל – הבנה של משמעות, קשרי סיבה-ותוצאה והחשיבות של האירועים המוצגים. אלגוריתמיי ראייה חישובית יכולים לספק פתרונות לבעיות הראשונות, אך הבעיות האחרונות הן סמנטיות בטבען ולכן מאתגרות אלגורתמים המבוססים על עיבוד מחשב.

אנו מציגים פתרון אמין מבוסס "חוכמת ההמון" עבור סיכום של וידאו, המבוסס על משימות שנשלחות לקהל. פתרון זה בא לענות על אחת המשימות המאתגרות בהבנה של עלילה: זיהוי קשרים של סיבה ותוצאה. הגישה שלנו בתחילה מחלקת את הוידאו באופן אוטומאטי לשוטים פשוטים, והם נחשבים לאלמנטים הבסיסיים של האלגוריתם. כיוון שיותר טבעי לבני אדם להבין סמנטיקה בטקסט (להבדיל מוידאו), נקטנו בגישה של המרת כל אלמנט בסיסי לטקסט באמצעות הקהל. לאחר מכן אנו עושים שימוש באלגוריתם "עצי הקשר" כדי לקבל הבנה גלובאלית. לבסוף, האלגוריתם פותר את בעיית קשרי הנסיבתיות ע"י בנייה של "גרף סיבה ותוצאה" אשר מכיל את האלמנטים הבסיסיים.

ניסויים שערכנו לאמידת יכולת הסיכום של האלגוריתם מראים שהאינפורמציה מ"גרפי סיבה ותוצאה" מאפשרים לייצר סיכומים טובים יותר עבור הוידאו המקורי.

עבודה זו בוצעה בהדרכתו של פרופ' אריאל שמיר מבי"ס אפי ארזי למדעי המחשב, המרכז הבינתחומי, הרצליה.

הבינתחומי הרצליה

בית ספר אפי ארזי למדעי המחשב

המרכז הבינתחומי בהרצליה

בית-ספר אפי ארזי למדעי המחשב

התכנית לתואר שני (M.Sc.) - מסלול מחקרי

סיכום סמנטי של וידאו באמצעות גרפי סיבה ותוצאה

מאת

שי שינפלד

עבודת תיזה המוגשת כחלק מהדרישות לשם קבלת תואר מוסמך M.Sc. במסלול המחקרי
בבית ספר אפי ארזי למדעי המחשב, המרכז הבינתחומי הרצליה

יוני 2016