February 2021

# RECOMMENDATION SYSTEM FOR HIDING SUBTITLES

Jan Ondras

Svein Gunnar Pettersen

# RECOMMENDATION SYSTEM FOR HIDING SUBTITLES

AUTHORS:
Jan Ondras
Svein Gunnar Pettersen

## ABSTRACT

Some videoconferencing/collaboration devices provide subtitles to users on-screen during meetings. However, when subtitles are permanently displayed on a videoconferencing device and a user is not reading/following the subtitles, the subtitles might become annoying to the user. As a result, the user is likely to be visually distracted. This proposal provides techniques to prevent users from being distracted by permanently displayed subtitles during videoconferences when users are not reading/following the subtitles. In particular, a system is presented herein that provides for the ability to either automatically hide videoconferencing subtitles when a system detects that a user is not reading/following the subtitles or, alternatively, provide a recommendation to the user to hide the subtitles when the system detects that the user is not reading/following the subtitles.

## DETAILED DESCRIPTION

As noted, subtitles that are permanently displayed on a videoconferencing device during a videoconference can annoy or distract a user when the user is not reading or following the subtitles. This proposal provides a novel recommendation system that can facilitate hiding subtitles on videoconferencing devices. The system can either automatically hide subtitles or offer a user an option to hide the subtitles if the system detects that the user is not reading the subtitles.

Figure 1, below, illustrates example details associated with the recommendation system in which eye gaze tracking techniques can be utilized to facilitate various operations of the system.

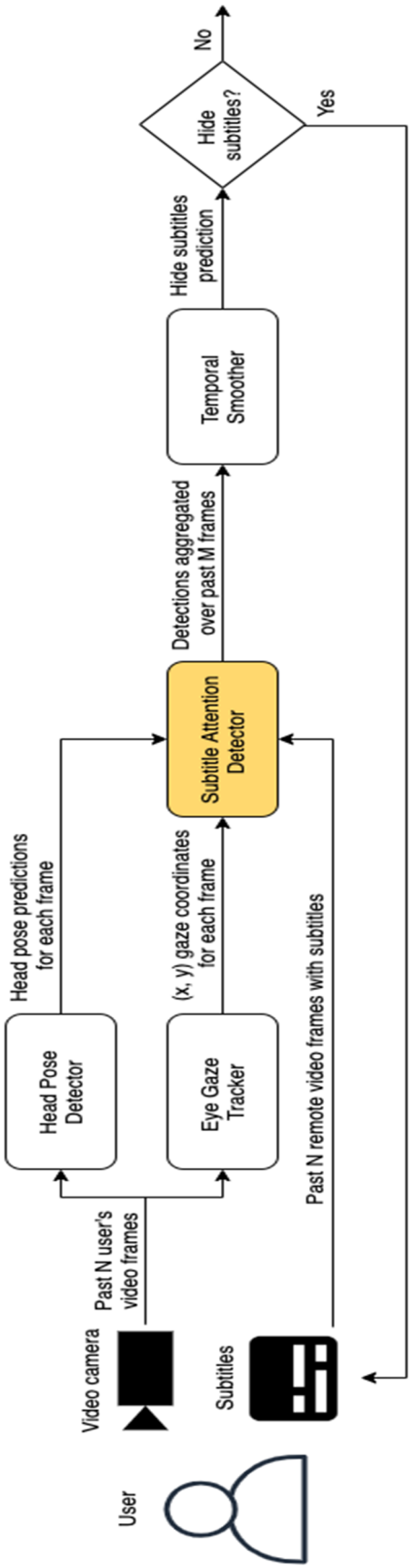


*Figure 1: Recommendation System for Hiding Videoconference Subtitles*

Consider an operational example involving the recommendation system as illustrated in Figure 1 that can facilitate hiding subtitles during a videoconference involving a user. During the videoconference, the last N video frames from the user's video camera are processed by an Eye Gaze Tracker to obtain (x, y) gaze coordinates for each video frame. Additionally, the last N video frames are fed into a Head Pose Detector predicting the user's head pose (e.g., frontal or left profile) for each frame. Thereafter, the Subtitle Attention Detector uses the eye gaze locations and head pose predictions, incorporates information about subtitles displayed over the same time frame, and predicts whether the user was reading the subtitles during the past N video frames. The last M such predictions are then aggregated by a Temporal Smoother and smoothed in time to generate a final prediction indicating whether or not to hide subtitles for the videoconference.

Additional details regarding various components of the recommendation system illustrated in Figure 1 follow. For example, the Eye Gaze Tracker can utilize an existing state of the art eye gaze tracking library that obtains inputs from a conventional camera (e.g., not relying on a special eye gaze tracking hardware). The Eye Gaze Tracker will process the input frames in real-time using computer vision and machine learning algorithms and predict two-dimensional (2D) gaze coordinates in the form of (x, y) gaze coordinates for each frame. Next, the Head Pose Detector can utilize existing speaker tracking functionality to detect heads and predict head poses (e.g., frontal or left profile) in real-time for each frame.

The Subtitle Attention Detector is a central component of the recommendation system in which the Subtitle Attention Detector can generate an ordered sequence of eye gaze fixation locations and times based on the 2D frame-by-frame gaze coordinates obtained from the Eye Gaze Tracker. The Subtitle Attention Detector will also input the past N video frames with subtitles and extract locations of individual subtitle words for each frame. Extracting the locations of individual subtitle words provides for the ability to disregard time intervals when no subtitles are displayed and, most importantly, relate the eye gaze behavior to the displayed subtitles for certain time intervals. Additionally, the Subtitle Attention Detector incorporates head pose predictions obtained from another source (e.g., the Head Pose Detector) that provides information indicating whether or not

the user's head is directed towards the screen or not, thus enabling the Subtitle Attention Detector to make more robust final subtitle attention predictions.

Several techniques may be utilized by the Subtitle Attention Detector to determine whether or a not the user is reading/following the subtitles. In one example, a correlation of locations technique can be utilized in which the locations of eye gaze fixations can be compared with the locations of words at a particular point in time. If the correlation exceeds a certain threshold value, the algorithm will conclude that the subtitles are being read. Another possibility is a correlation of dynamics technique in which the dynamics of eye gaze fixations can be compared with the dynamics of incoming subtitles' words. If the correlation exceeds a certain threshold value, the algorithm will conclude that the subtitles are being read. In yet another example, head pose predictions can be utilized. For this technique, for instance, if the number of frames with frontal head pose predictions is less than a certain threshold value, the algorithm will conclude that the subtitles are not being read. Yet another technique may involve the use of a machine-learning model that can be trained on time sequences of gaze coordinates, subtitles' word locations, and possibly also head pose predictions. A prediction from the model can be directly used to conclude whether the subtitles are being read. In various implementations, one or more of these techniques can be utilized for a final subtitle attention prediction.

To avoid unstable predictions over time and thus flickering of the final hide subtitle prediction, the Temporal Smoother can be utilized to smooth the past M attention predictions using a sliding window approach and a median filter.

Use cases involving shared use of the same videoconferencing device by multiple users can also be supported. For example, since the Eye Gaze Tracker and Head Pose Detector can provide outputs for each person in a video stream, the Subtitle Attention Detector can also generate a subtitle attention prediction for each of multiple users that may be viewing the same videoconferencing device. For such use cases involving multiple users, the system may be enhanced to hide or recommend to hide the subtitles only if the system determines that none of the users are reading the subtitles.

In general, subtitles can be hidden via one of the following *subtitle hiding modes*:

- **Explicit** (user requested) - In this mode, a user will be able to hide the subtitles via a user interface (UI) (e.g., by pressing a button, etc.) or, if available, using a Voice Assistant (e.g., using a voice command "Ok, system. Hide subtitles."). This mode will be always available for the system.

- **Semi-automatic** (user prompted) - In this mode, a recommendation to hide subtitles will be delivered to a user either via a UI (e.g., a pop-up window displayed on the screen) or, if available, using a Voice Assistant (e.g., using a voice prompt, such as, "Do you wish to hide subtitles?"). In both cases, the user will have an option to accept or reject the recommendation with an option to disable such recommendations in future (e.g., a "Don't ask again" option) in which case the system will transition into the explicit subtitle hiding mode.

- **Fully automatic** (user notified) - In this mode, subtitles will be hidden automatically. A user will be notified about the action and will be offered an immediate possibility to revert the action either via a UI (e.g., a pop-up window displayed on the screen) or, if available, using a Voice Assistant (e.g., using a voice prompt, such as, "We noticed you were not reading the subtitles, so we hid them. Do you wish to bring them back?"). In either case, if the user decides to revert it (i.e., bring the subtitles back on the screen) the system will transition into the semi-automatic subtitle hiding mode. If the system infers that subtitles are to be hidden, either the semi-automatic or the fully automatic subtitle hiding mode will be used.

Once removed, subtitles can be brought back onto a screen (restored) via one of the following *subtitle restoration modes*:

- **Explicit** (user requested) - In this mode, the user will be able to display/restore subtitles via a UI (e.g., by pressing a button) or, if available, using a Voice Assistant (e.g., using a voice command, such as, "Ok, system. Show subtitles."). This mode will be always available for the system.

- **Semi-automatic** (user prompted) - For this mode, if, within a certain amount of time after the subtitles were hidden, the user is still looking at the area where the subtitles were, then a recommendation to restore the subtitles will be delivered to the user either via a UI (e.g., a pop-up window displayed on the screen) or, if available, using a Voice Assistant (e.g., using a voice prompt, such as, "Do you wish to restore the subtitles?"). In both cases, the user will have an option to accept or reject the recommendation with an option to disable such recommendations in the future (e.g., a "Don't ask again" option) in which case the system will transition into the explicit subtitle restoration mode.

- **Fully-automatic** (user notified) - For this mode, if, within a certain amount of time after the subtitles were hidden, the user is still looking at the area where the subtitles were, then the subtitles will be restored automatically and the user will be notified about this action and offered an immediate possibility to revert it either via a UI (e.g., a pop-up window displayed on the screen) or, if available, using a Voice Assistant (e.g., using a voice prompt, such as, "We noticed you were looking for subtitles, so we restored them for you. Do you wish to hide them?"). In either case, if the user decides to revert restoration of the subtitles (i.e., hide the subtitles) the system will transition into the semi-automatic subtitle restoration mode.

The subtitle hiding and restoration modes are illustrated via state transition diagrams, shown below in Figure 2. The system can be initialized in any subtitle hiding mode and any subtitle restoration mode with the explicit subtitle hiding and restoration modes always being available.

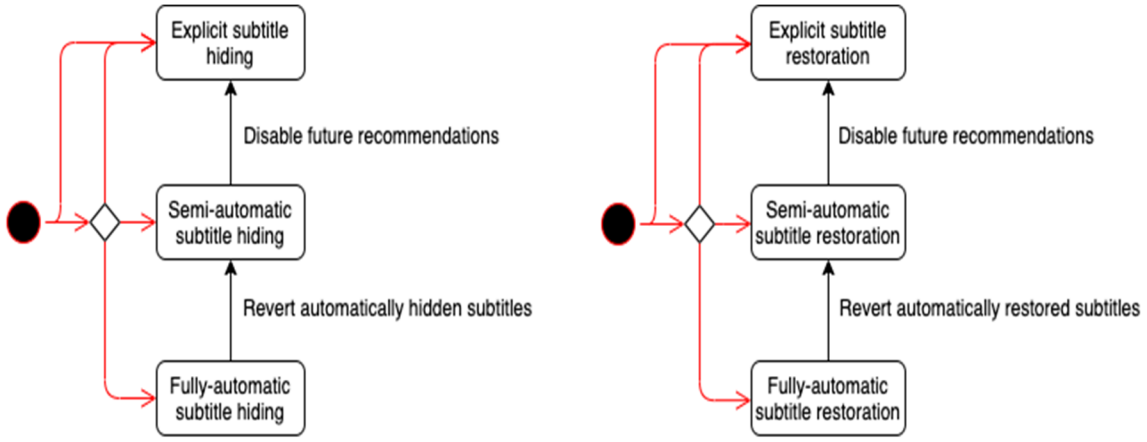


*Figure 2: Subtitle Hiding Modes (left) and Subtitle Restoration Modes (right)*

In some instances, the system may be enhanced to provide considerations for a presentation mode involving multiple screens in order to handle cases in which the screen that a user is following might not be in line with a video camera capturing the user. For example, the system can be enhanced to handle such use cases based on determining whether a presentation mode for multiple screens is on or off. Based on determining such information, the system can enable only the explicit (or possibly the semi-automatic) subtitle hiding mode so that the subtitles will not be automatically hidden. An additional enhancement may include providing a capability to distinguish transcripts from various speakers in order to facilitate hiding subtitles for certain speakers, as desired.

Multiple features for the recommendation system presented herein can be distinguished from current eye gaze tracking and subtitle systems. To begin, current systems do not provide for the ability to predict whether or not subtitles should be hidden. Further, current systems do not consider a combination of input information, including eye gaze coordinates, subtitles words' locations and timings, and head pose predictions. Additionally, current systems typically perform an offline analysis rather than an online (on-the-fly) analysis as provided by the recommendation system presented herein.

In summary, the recommendation system architecture including the Subtitle Attention Detector provide a novel solution to provide for the ability to either automatically hide subtitles or provide recommendations to hide subtitles when the system detects that a user is not reading/following the subtitles during a videoconference.