October 2020

SMART CONTENT COMPOSER

Hsing-Sheng Lin

IWei Lin

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

SMART CONTENT COMPOSER

ABSTRACT

This paper describes techniques related to a smart content composer that may be implemented on a client media player/platform and/or a content distribution platform. The smart composer creates a superior viewing experience to watch any content on any type of display device, even when the aspect ratios of the generated content and the display device are different, such that the original aspect ratio of the generated content is maintained for display in one or more identified fields of view on any given display device. In addition, the smart composer may be enabled to create a tailored and potentially new content presentation each time, for the same personal and/or shared media content (e.g., images and videos), such that the user is not provided with the same presentation of the same content over and over. Instead, the presentation of this content may be dynamically composed according to one or more factors, such as a user's context, and may in certain cases be optimized for specific sentiments of the user. Based on the customized output, the smart content composer may in some cases also be configured to automatically generate personal or special topic digests of, e.g., videos or trailers that are based on the customized presentation of content, which may be shared with other users. As a result, the techniques described herein may provide an innovative viewing experience for users, and they may also provide improved content creation and sharing mechanisms.

DESCRIPTION

Users typically access or view content on different types of display devices (e.g., display devices of smartphones, personal computers, televisions, game consoles, or other infotainment devices) in their daily life. These users expect to be able to view or consume content on any type

of display device that is convenient to them at any given time. In some cases, display devices are configured to output content in a landscape or horizontal orientation, particularly if such display devices (e.g., televisions) have screens with larger horizontal dimensions. However, an increasing number of display devices, such as display devices of mobile computing devices or smartphones, are configured to output content in a portrait or vertical orientation. Users of such devices, though, may periodically have a bad viewing experience when there is a mismatch of the aspect ratio between the display device and that of the actual generated content (e.g., video and/or image content). The mismatch of these different aspect ratios between the display devices and the content can degrade a user's viewing experience.

In addition to this potential mismatch issue, users can often get easily bored when watching the same content over and over. For instance, users may consume certain types of content repeatedly over time, such as personal videos or pictures, music videos, movies, advertisements, and the like. Currently, in these situations, users are typically presented with the content in exactly the same way in each instance, and thus the users may get bored with such content over time. This may especially be the case for certain types of content, such as advertisements, when users may become annoyed because they have less control over whether or not to consume such content when it is presented. Even if users have some control over the presentation of such content, they still may have a reduced interest in watching such content if it is presented in the same way each and every time. As a result, current approaches may result in missed opportunities to create new value for existing content.

Currently, there are certain existing mitigation tactics to address the issue of presenting landscape-oriented or horizontal content (e.g., videos) on mobile devices having a portrait-oriented or vertical display device, or presenting portrait-oriented content for landscape-oriented

display devices (e.g., television or personal computer displays). Certain mitigation techniques include letterboxing, pillarboxing, cropping, and scaling. However, these existing techniques have limitations. For example, letterboxing and pillarboxing techniques may degrade the "full screen" experience for a user; cropping may hide the entirety of the content for display to the user; and scaling may destroy the aspect ratio of the content.

To address these issues and limitations, the techniques described herein provide a smart content composer. This smart composer creates a superior viewing experience to watch any content on any type of display device, even when the aspect ratios of the generated content and the display device are different. In addition, the smart composer may be enabled, in various cases, to create a tailored and potentially new content prospect or presentation every time for the same content. The presentation of this content may be dynamically composed according to one or more factors, such as a user's context, and may in certain cases be optimized for specific sentiments of the user. As a result, the techniques described herein may provide an innovative viewing experience for a user, and may provide improved content creation and sharing mechanisms.

As described herein, the storage of any user data, such as sentiment data or identification information associated with a user, may only occur, in various examples, in response to a computing device receiving an affirmative consent from the user. The computing device, for example, may provide an explicit request to the user regarding the storage and subsequent use of such data, and the computing device may proceed to store the data only upon receiving an affirmative response and consent from the user. Furthermore, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs, or features described herein may enable collection and/storage of user information (e.g., sentiment

data, user preference data, user location data, user identification data), and/or if and when systems, programs, or features described herein may enable transmission of content or communications between devices. In addition, certain data may be treated in one or more ways before it is stored or used, so that identifiable information is removed. For example, a user's identity may be treated so that no identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is stored and/or used, and what information is provided to the user.

System Architecture

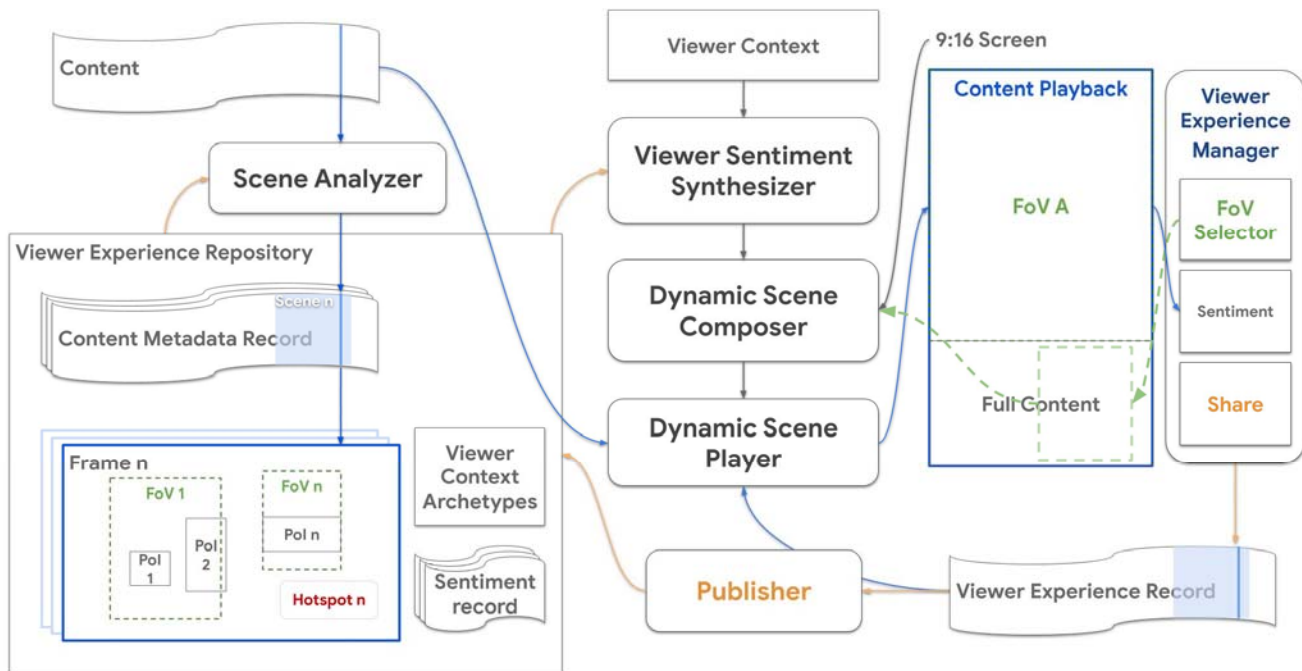Figure 1 shown below illustrates an example system architecture for implementing the disclosed techniques.

Figure 1: System Architecture

5

As shown in Figure 1, the example system architecture includes various indicated components, including the following: a scene analyzer, a viewer experience repository, a viewer sentiment synthesizer, a dynamic scene composer, a dynamic scene player, and a viewer experience manager, which are described in further detail below.

Scene Analyzer

As illustrated in Figure 1, the scene analyzer is configured to pre-analyze content (e.g., video content), on a frame-by-frame basis, to generate a Content Metadata Record that is associated with the content. The Content Metadata Record may include metadata information for points-of-interest candidates (e.g., objects, people, faces, etc.) associated with the content, field-of-view candidates, hotspot candidates, and indications of the start and end for each frame of a distinguished scene, as described in further detail below. The scene analyzer may, in some examples, also enhance the information included in the Content Metadata Record for each candidate with one or more corresponding Sentiment Record datasets. Although the system automatically generates the Content Metadata Record, a user may also manually annotate the Content Metadata Record based on user input or preferences, and the system may also fine tune the Content Metadata Record with respect to initial baseline values or to enhancements that are later incorporated. In some cases, a machine learning engine may utilize a machine learning model to identify or predict points-of-interest candidates, field-of-view candidates, and hotspot candidates identified in the Content Metadata Record. The machine learning model may be trained and updated based on Viewer Experience Records and/or other related viewer information that is captured over time, as described in further detail below.

Viewer Experience Repository

As also shown in Figure 1, the viewer experience repository aggregates Viewer Experience Records. Each Viewer Experience Record (VER) contains a Content Metadata Record, a Sentiment Record, and a Viewer Context (described below).

<u>Content Metadata Record</u>

A Content Metadata Record (CMR) is a data structure that includes key features of particular content. A CMR may include the following for each piece of content:

- a content identifier (ID);
- for each scene of the content:
    - a scene ID; and
    - for each frame of the respective scene:
        - a frame ID;
        - x Points of Interest:
            - o Point of Interest (PoI) ID for each PoI;
            - o position [X,Y] for each PoI; and
            - o size [W,H] for each PoI;
        - y Fields of View (which each may include one or more PoI's):
            - o Field of View (FoV) ID for each FoV;
            - o position [X,Y] for each FoV; and
            - o size [W,H] for each FoV;
        - z Hotspots (e.g., hotspots or regions of interest that are, e.g., identified by a user or a machine-learning engine):

- o   hotspot ID for each hotspot;

- o   Position [X,Y] for each hotspot; and

- o   Size [W,H] for each hotspot.

Sentiment Record

A Sentiment Record (SR) includes sentiment features (e.g., sentiment keywords) of a viewing experience associated with a given viewer context. Sentiment keywords may include any keywords provided by a user to indicate the user's sentiment for a particular target or portion of content. For instance, a SR is a data structure that may include the following:

- a SR (ID);

- a Content Metadata Record ID that is associated with the SR ID;

- a Viewer Context (VC);

- a content ID; and

- for each target:

  - a target (e.g., content, scene, frame, PoI, FoV, hotspot) ID;

  - sentiment keywords; and

  - a score.

Viewer Context and Archetypes

A Viewer Context (VC) includes viewer and context features. A Viewer Context and Archetypes (VCA) can be created by grouping together VC's for a specific goal dynamically. A VC may comprise a data structure that includes the following elements:

- key viewer features:

  - target viewer persona (e.g., features or characteristics of a target viewer of the content);

  - anti-viewer persona (e.g., features or characteristics of a non-targeted or non-expected viewer of the content);

- key context features

  - time of the day, day of the week, month, year, etc.;

  - special occasions associated with viewer or content (e.g., festival, holiday, football or other sporting event, etc.);

  - viewer mode (e.g., focused watching, multi-tasking, primary or secondary content);

  - aspect ratio of the screen on which content is viewed;

  - place/location (e.g., workplace, home, kitchen, living room, study room, on the road) associated with viewed content;

  - activities performed (e.g., walking, running, riding, driving) while viewing content

Viewer Sentiment Synthesizer

As also shown in Figure 1, the system further includes a Viewer Sentiment Synthesizer (VSS). The VSS dynamically scores and ranks all visual components (e.g., PoI's, FoV's, hotspots) by synthesizing a current viewer context with the viewer experience datasets from the Viewer Experience Repository. At the beginning of a particular scene, a candidate list of components is created and ranked from scratch. In the middle of a scene, the positions and sizes

of the top candidates are adjusted accordingly on each frame. In various examples, the VSS may interact with a machine-learning engine to rank and score the visual components based on the information from the Viewer Experience Repository and/or the model.

Dynamic Scene Composer

Figure 1 also shows a Dynamic Scene Composer (DSC). The DSC determines the actual Fields of Views (FOV's) and composition of top Viewer Contexts at playback time, frame by frame, according to the screen aspect ratio of the Viewer Context. The DSC can either be performed locally for local playback, or in the backend for streaming.

Vertical Content to Landscape Screens

Smartphones are typically used in a portrait orientation to output content, and content is often produced for that orientation when displayed at the display devices of these smartphones. In many cases, this vertically oriented content is generated with, e.g., an aspect ratio of 9:16 or a similar ratio. However, viewers may also wish to view this vertically oriented content on other types of display devices that have a landscape configuration, such as flat-screen televisions, which provide different aspect ratios (e.g., a 16:9 aspect ratio) than that of the vertical content. In order to provide such a viewing experience, existing or traditional systems and players may need to utilize a pillarboxing approach where the vertical content is displayed on only a portion of the landscape screen.

For example, as shown in Figure 2(A) below, a common pillarboxing approach may be used to display vertical content on a landscape screen. The vertical content is

generated for, e.g., a smartphone and has, e.g., a 9:16 aspect ratio. This content has a

particular field of view "FoV A" that is located substantially in the middle of the screen,

as indicated by the green rectangle in Figure 2(A). However, since the landscape screen

has a different aspect ratio (e.g., 16:9) than the aspect ratio of the vertical content, the

pillarbox technique displays the vertical content in the center of the display, surrounded

by black regions on both the left and right sides, as shown in Figure 2(A).
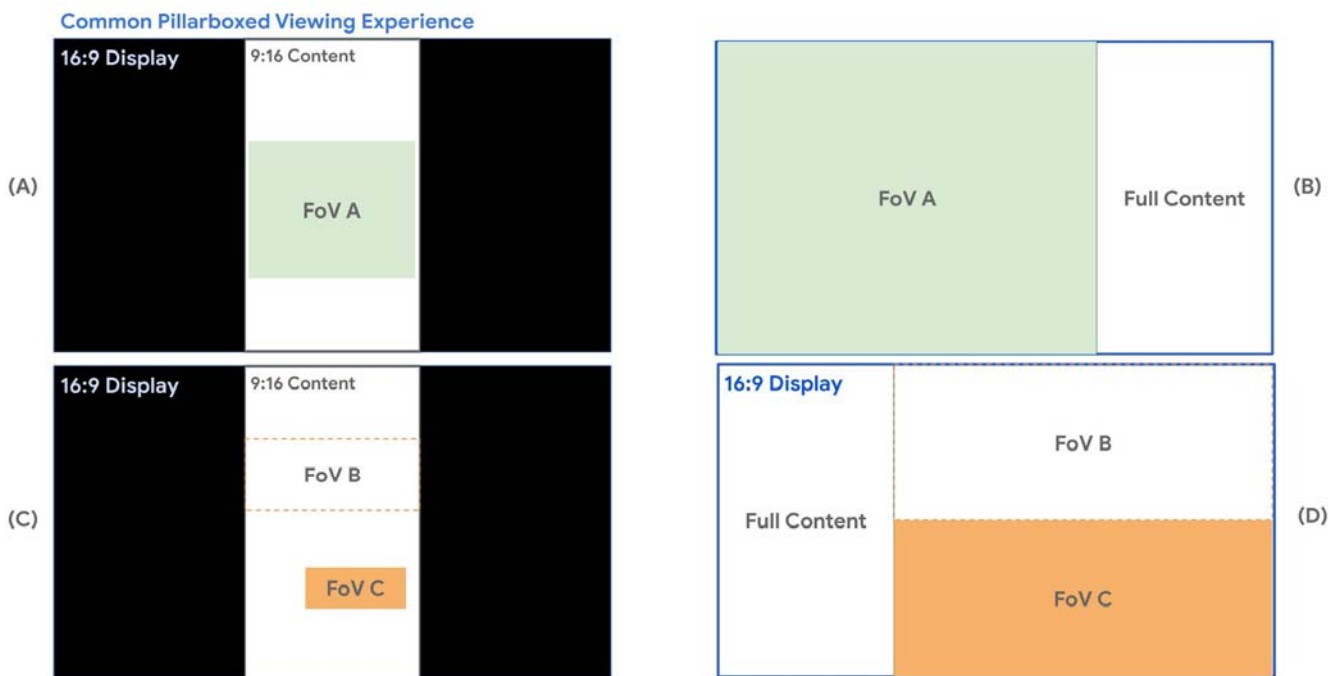
Figures 2(A) – 2(D): Displaying Vertical Content on Landscape Screens

However, utilizing the techniques disclosed herein, the system of Figure 1

provides a great new viewing experience on these landscape screens, such as shown in

Figure 2(B). The system of the present disclosure may output the entire or full vertical

content on a portion of the display, such as shown on the right-hand side of Figure 2(B),

which has a content aspect ratio of 9:16 (i.e., the aspect ratio of the originally generated vertical content). In addition, the system may output a scaled rendering of the particular field of view "FoV A," as shown in green on the left-hand side of Figure 2(B). The system renders the content included in this field of view using the same aspect ratio of 9:16, thus maintaining the original content aspect ratio of the vertical content when scaling information included in this particular field of view.

Figure 2(C) illustrates another example in which the vertical content includes two identified fields of view, namely "FoV B" and "FoV C." The user and/or system may have identified these fields of view. Using the traditional pillarboxing approach shown in Figure 2(C), the vertical content having a 9:16 aspect ratio is displayed substantially in the middle of the screen, including the two fields of view "FoV B" and "FoV C." Since the landscape screen has a different aspect ratio (e.g., 16:9) than the aspect ratio of the vertical content, similar to that of Figure 2(A), the pillarbox technique once again displays the vertical content in the center of the display, surrounded by black regions on both the left and right sides, as indicated in Figure 2(C).

However, utilizing the techniques disclosed herein, the system of Figure 1 can provide an improved viewing experience over that shown in Figure 2(C). For instance, as shown in Figure 2(D), the system of the present disclosure may output the entire or full vertical content on a portion of the display, such as shown on the left-hand side of Figure 2(D), which has a content aspect ratio of 9:16. In addition, the system may output a scaled rendering of the particular fields of view "FoV B" and "FoV C," as shown on the right-hand side of Figure 2(D). The system renders the content included in these fields of

view using the same aspect ratio of 9:16, thus maintaining the original content aspect ratio of the vertical content when scaling information included in these fields of view.

As a result, using the horizontal displays shown in Figures 2(B) and 2(D) (e.g., displays providing a 16:9 aspect ratio for displaying content), the user can view not only the original and full vertical content in its original aspect ratio of 9:16, but also the identified and/or specific fields of view included in the content, where these fields of view are also output using the same aspect ratio of 9:16 as the original content. As such, if desired, the viewer can see a copy of the entire original content, as well as enlarged versions of just the specifically identified fields of view (e.g., those identified by the dynamic scene composer in the system of Figure 1.)

In some cases, users or the machine learning engine may also dynamically adjust the positions and/or sizes of the fields of views (e.g., for one or more frames). In some cases, the system of Figure 1 may record or share the metadata (e.g., metadata included in the Content Metadata Records) with other systems or users. The system (e.g., using the machine learning engine) may also use this metadata for dynamically fine turning the Viewer Context.

Horizontal Content to Portrait Screens

In various cases, content is also created for display on various display devices configured in a horizontal or landscape orientation, such as high-definition television (HDTV) screens having a 16:9 aspect ratio, or legacy devices (e.g., personal computers, televisions) having a 4:3 aspect ratio. Viewers may also wish to view content generated for such horizontally oriented screens, which may be referred to as horizontal content, on

other types of screens that have a portrait configuration or orientation in certain modes, such as smartphones. For instance, a smartphone having a portrait orientation or configuration may display content having a 9:16 aspect ratio, which is different than the aspect ratio for the horizontal content that is generated for, e.g., a flat-screen television. In order to provide a viewing experience of such horizontal content on a smartphone operating in a portrait orientation, existing systems and players may need to utilize a letterboxing approach where the horizontal content is displayed on only a portion of the smartphone screen.

For example, as shown in Figure 3(A) and 3(B) below, horizontal content having an aspect ratio of 16:9, which is generated for a flat-screen television providing a similar aspect ratio, may have one or more fields of view. Figure 3(A) illustrates a first example of content having one identified field of view ("FoV A"). Figure 3(B) illustrates a second example of content having two different, and partially overlapping, fields of view ("FoV B" and "FoV C"). A common letterboxing approach may be used to display horizontal content on a portrait screen of a smartphone, which provides an aspect ratio of 9:16, as shown in Figure 3(C).
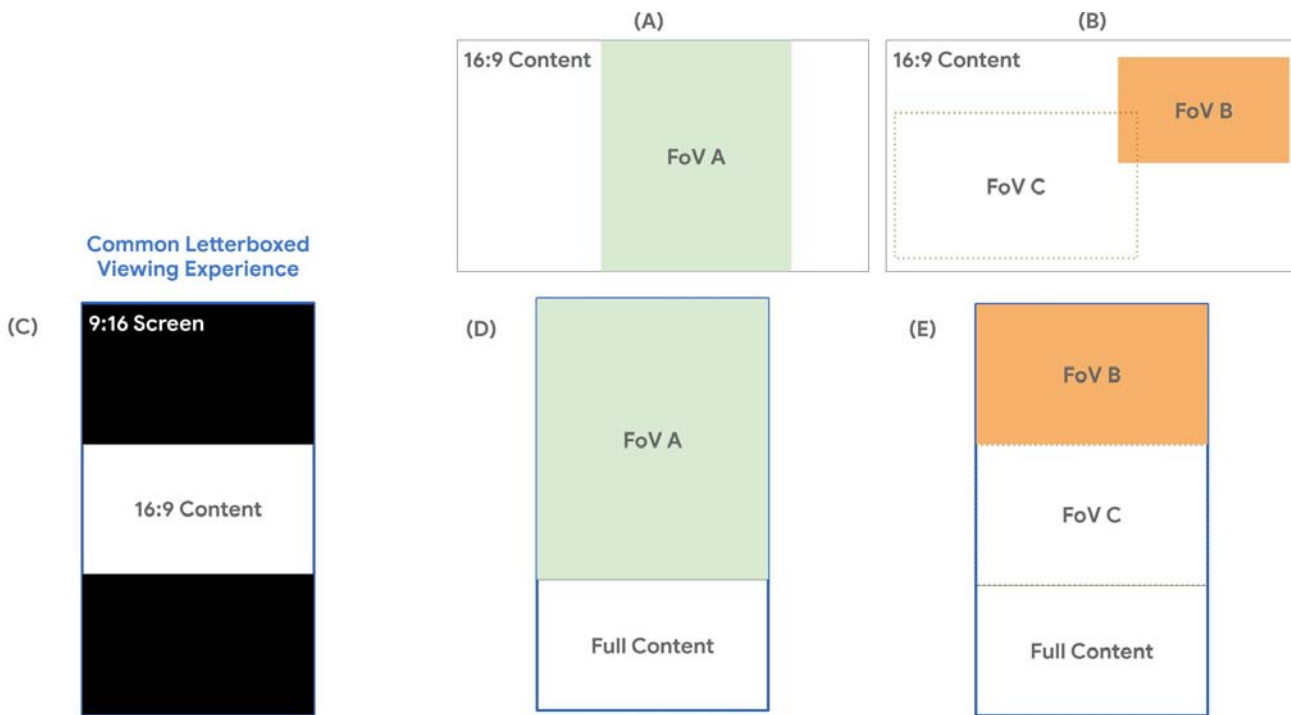
Figures 3(A) – 3(E): Displaying Horizontal Content on Portrait Screens

Using the traditional letterboxing approach shown in Figure 3(C), the horizontal

content generated with a 16:9 aspect ratio is displayed substantially in the middle of the

smartphone screen, such as in both cases in which the content includes one field of view

(e.g., "FoV A") or multiple fields of view (e.g., "FoV B" and "FoV C"). However, since

the portrait screen of the smartphone may provide a different aspect ratio (e.g., 9:16) than

the aspect ratio of the horizontal content, the letterbox technique displays the horizontal

content in the center of the display, surrounded by black regions on both the top and

bottom, as shown in Figure 3(C).

However, utilizing the techniques disclosed herein, the system of Figure 1

provides an improved viewing experience over that shown in Figure 3(C). For instance,

as shown in Figures 3(D) and 3(E), the system of the present disclosure may output the

entire or full horizontal content on a portion of the portrait screen, such as shown on the

bottom-portion of both Figure 3(D) and Figure 3(E), which has the originally generated

content aspect ratio of 16:9 for the horizontal content. In addition, the system may output

a scaled rendering of the particular fields of view "FoV A," "FoV B," and "FoV C."

Figure 3(D) shows the content included in "FoV A" above the full content, where the

content for "FoV A" is displayed at the aspect ratio of 16:9 for the originally generated

horizontal content. Similarly, Figure 3(E) shows the content included in "FoV B" and

"FoV C" displayed at the original aspect ratio of 16:9 for the originally generated

horizontal content. The system renders the content included in all of these fields of view

using the same aspect ratio of 16:9, thus maintaining the same content aspect ratio of the

original horizontal content when scaling content included in these particular fields of

view.

Dynamic Scene Player

As also illustrated in Figure 1, the system described herein further includes a Dynamic

Scene Player (DSP). The DSP can either play the content according to the output of the

Dynamic Scene Composer (DSC) or just according to the information from the Viewer

Experience Repository (VER). This content may be saved by the viewer or shared by others.

Viewer Experience Manager

Figure 1 also shows a Viewer Experience Manager (VEM). The VEM provides various

features at view time for a user. For instance, a user may interact with the VEM to add or adjust

positions and/or sizes of any of the displayed fields of view. The user may use the VEM to

identify fields of view and/or identify any particular objects or points of interest that are to be included in these fields of view. Once a field of view is identified, the user may interact with the VEM to change the position and/or size of this view.

In addition, the VEM may switch to a different composition or, based on input from a viewer or the machine learning engine, annotate sentiment keywords to fields of view, frames, scenes, and/or the content. Furthermore, based on input from the viewer or the machine learning engine, the VEM may provide feedback (e.g., 'like' or 'dislike' feedback) with respect to fields of view, frames, scenes, and/or the content. The VEM may also be configured to save viewer experience repositories, manage a viewer experience playlist, and/or share viewer experience repositories.

As indicated above, hotspots or regions of interest are, e.g., identified by a user or a machine-learning engine. However, in some cases, an eye tracking technology may be added as part of a hotspot analyzer into the system of Figure 1 to track the viewer's eye movement over time to identify hotspots. For example, if the eye tracking technology determines that the viewer's eyes look or stare for an extended period of time at a particular region having one or more point of interests, the hotspot analyzer may identify this region as a potential hotspot. The hotspot manager and eye tracking technology may, in some cases, be part of the VEM.

User Data & Privacy for Client and Content Distribution Platforms

The design of the system illustrated in Figure 1 is flexible to be implemented on various different client and/or content distribution platforms. For example, with respect to client platforms, the system may be implemented in a content player application on one or more platforms or operating systems that adhere to the following requirements:

- the client platform refrains from recording any personally identifiable information for a user;

- the client platform complies with the platform application programming interface (API), permission, and user data sharing policy of each available or collected user;

- the client platform acquires information regarding the user concerns when collecting any user signals; and

- the client platform explicitly obtains user consent before performing any user actions for content saving, sharing, and/or uploading.

With respect to content distribution platforms, the smart composer system can be integrated into any video or image distribution platform, which adhere to the following requirements:

- the platform will always maintain the anonymity of user signals, and will aggregate to such signals to a larger user group in general, rather than for an individual, unless a user explicitly agrees to maintain the user signals otherwise in association with a user account;

- if the platform enables a user to connect to a user account, users will have control as to which user signals are uploaded to the platform and/or when such signals are uploaded; and

- if such signals are uploaded, the platform's user data & privacy will be honored.

Conclusion

Thus, any number of client media players or platforms, as well content distribution platforms, may implement the disclosed techniques and systems for providing a smart content composer. The smart composer creates a superior viewing experience to watch any content on any type of display device, even when the aspect ratios of the generated content and the display device are different, such that the original aspect ratio of the generated content is maintained for display in one or more identified fields of view on any given display device. In addition, the smart composer may be enabled to create a tailored and potentially new content presentation each time, for the same personal and/or shared media content (e.g., images and videos). For example, the smart content composer may automatically tailor content to each viewer and corresponding viewer context. The smart composer may provide a new viewer-generated viewing experience by creating new prospects or remixing of content, such that the user is not always provided with the same presentation of the same content each time, but is instead provided with a tailored and customized presentation of material based on source content. In some cases, the user may interact with the smart composer to share such tailored or customized content with other users. Based on the customized output, the smart content composer may also be configured to automatically generate personal or special topic digests of, e.g., videos that are based on the customized presentation of content, which also may be shared with other users. As a result, the techniques described herein may provide an innovative viewing experience for users, and may also provide improved content creation and sharing mechanisms.

REFERENCES

1.   United States Patent Application Publication No. 2018-0005041.

2.   International Application Publication No. WO 2019/112642.