(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 November 2016 (17.11.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/183239 A1

(51) International Patent Classification:
G09G 3/36 (2006.01) *G06K 9/00* (2006.01)

(21) International Application Number:
PCT/US2016/031925

(22) International Filing Date:
11 May 2016 (11.05.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

62/160,353	12 May 2015 (12.05.2015)	US
15171033.2	8 June 2015 (08.06.2015)	EP
62/193,678	17 July 2015 (17.07.2015)	US
62/259,139	24 November 2015 (24.11.2015)	US

(71) Applicant: **DOLBY LABORATORIES LICENSING CORPORATION** [US/US]; 1275 Market Street, San Francisco, California 94103 (US).

(72) Inventors: **ATKINS, Robin**; c/o Dolby Laboratories, Inc., 1275 Market Street, San Francisco, California 94103 (US). **WARD, Gregory John**; c/o Dolby Laboratories, Inc., 1275 Market Street, San Francisco, California 94103 (US). **HULYALKAR, Samir N.**; c/o Dolby Laboratories, Inc., 1275 Market Street, San Francisco, California 94103 (US). **SU, Guan-Ming**; c/o Dolby Laboratories, Inc., 1275 Market Street, San Francisco, California 94103 (US). **QU, Sheng**; c/o Dolby Laboratories, Inc., 1275 Market Street, San Francisco, California 94103 (US).

(74) Agents: **DOLBY LABORATORIES, INC.** et al.; Intellectual Property Group, 1275 Market Street, San Francisco, California 94103 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,

[Continued on next page]

(54) Title: METADATA FILTERING FOR DISPLAY MAPPING FOR HIGH DYNAMIC RANGE IMAGES

(57) Abstract: Systems and methods are disclosed for filtering metadata to be used in display management. Given an input video stream and input metadata comprising at least one of minimum, average, or maximum luminance values of the video frames in the video stream, values of a function of the input metadata are filtered using a temporal filter to generate filtered metadata, wherein the filtering is based only on metadata for input frames in the same scene. Methods for temporal filtering based on an exponential moving average filter or a look-ahead sliding window filter are presented, including methods for scene-change detection using the input metadata.

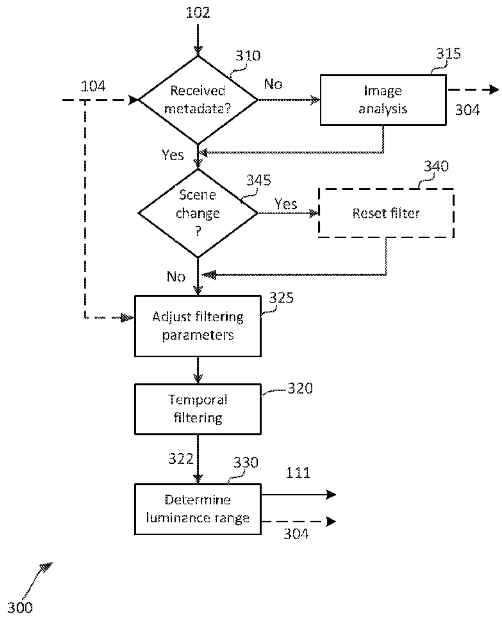

FIG. 3

WO 2016/183239 A1

SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

METADATA FILTERING FOR DISPLAY MAPPING FOR HIGH DYNAMIC RANGE IMAGES

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to United States Provisional Patent Application No. 62/160,353, filed on May 12, 2015, United States Provisional Patent Application No. 62/193,678, filed on July 17, 2015, United States Provisional Patent Application No. 62/259,139, filed on November 24, 2015 and European Patent Application No. 15171033.2, filed on June 8, 2015, each of which is incorporated herein by reference in its entirety.

TECHNOLOGY

[0002] The present invention relates generally to images. More particularly, an embodiment of the present invention relates to metadata filtering for display mapping for images and video sequences with high dynamic range.

BACKGROUND

[0003] As used herein, the term 'dynamic range' (DR) may relate to a capability of the human visual system (HVS) to perceive a range of intensity (e.g., luminance, luma) in an image, e.g., from darkest blacks (blacks) to brightest whites (highlights). In this sense, DR relates to a 'scene-referred' intensity. DR may also relate to the ability of a display device to adequately or approximately render an intensity range of a particular breadth. In this sense, DR relates to a 'display-referred' intensity. Unless a particular sense is explicitly specified to have particular significance at any point in the description herein, it should be inferred that the term may be used in either sense, e.g. interchangeably.

[0004] As used herein, the term high dynamic range (HDR) relates to a DR breadth that spans the some 14-15 orders of magnitude of the human visual system (HVS). In practice, the DR over which a human may simultaneously perceive an extensive breadth in intensity range may be somewhat truncated, in relation to HDR. As used herein, the terms enhanced dynamic range (EDR) or visual dynamic range (VDR) may individually or interchangeably relate to the DR that is perceivable within a scene or image by a human visual system (HVS) that includes eye movements, allowing for some light adaptation changes across the scene or image. As used herein, EDR may relate to a DR that spans 5 to 6 orders of magnitude. Thus while perhaps somewhat narrower in relation to true scene referred HDR, EDR nonetheless represents a wide DR breadth and may also be referred to as HDR.

[0005] In practice, images comprise one or more color components (e.g., luma Y and chroma Cb and Cr) wherein each color component is represented by a precision of n-bits per pixel (e.g., n=8). While SDR images can typically be encoded with 8-10 bits per color component, EDR and HDR images typically require more than 8 bits (e.g., 10-12 bits, or more). EDR and HDR images may also be stored and distributed using high-precision (e.g., 16-bit) floating-point formats, such as the OpenEXR file format developed by Industrial Light and Magic.

[0006] A reference electro-optical transfer function (EOTF) for a given display characterizes the relationship between color values (e.g., luminance) of an input video signal to output screen color values (e.g., screen luminance) produced by the display. For example, ITU Rec. ITU-R BT. 1886, "Reference electro-optical transfer function for flat panel displays used in HDTV studio production," (03/2011), which is included herein by reference in its entity, defines the reference EOTF for flat panel displays based on measured characteristics of the Cathode Ray Tube (CRT). Given a video stream, any ancillary information is typically embedded in the bit stream as metadata. As used herein, the term "metadata" relates to any auxiliary information that is transmitted as part of the coded bitstream and assists a decoder to render a decoded image. Such metadata may include, but are not limited to, color space or gamut information, reference display parameters, and auxiliary signal parameters, as those described herein.

[0007] Most consumer HDTVs range from 300 to 500 nits with new models reaching 1000 nits (cd/m^2). As the availability of HDR content grows due to advances in both capture equipment (e.g., cameras) and displays (e.g., the PRM-4200 professional reference monitor from Dolby Laboratories), HDR content may be color graded and displayed on displays that support higher dynamic ranges (e.g., from 1,000 nits to 5,000 nits or more). Such displays may be defined using alternative EOTFs that support high luminance capability (e.g., 0 to 10,000 nits). An example of such an EOTF is defined in SMPTE ST 2084:2014 "High Dynamic Range EOTF of Mastering Reference Displays," which is incorporated herein by reference in its entirety. In general, without limitation, the methods of the present disclosure relate to any dynamic range higher than SDR. As appreciated by the inventors here, improved techniques for the display of high-dynamic range images are desired.

[0008] The approaches described in this section are approaches that could be pursued, but not necessarily approaches that have been previously conceived or pursued. Therefore, unless otherwise indicated, it should not be assumed that any of the approaches described in this section qualify as prior art merely by virtue of their inclusion in this section. Similarly, issues identified with respect to one or more approaches should not assume to have been recognized in any prior art on the basis of this section, unless otherwise indicated.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] An embodiment of the present invention is illustrated by way of example, and not in way by limitation, in the figures of the accompanying drawings and in which like reference numerals refer to similar elements and in which:

[00010] FIG. 1 depicts an example process for backlight control and display management according to an embodiment of this invention;

[00011] FIG. 2 depicts an example relationship between a backlight setting and the instantaneous dynamic range for a display with dynamic range control;

[00012] FIG. 3 depicts an example process for adaptive luminance range mapping according to an embodiment;

[00013] FIG. 4A and FIG. 4B depict examples of applying moving average filtering to metadata values of an input sequence according to embodiments of this invention; and

[00014] FIG. 5 depicts an example of a look-ahead sliding window for filtering metadata according to an embodiment.

DESCRIPTION OF EXAMPLE EMBODIMENTS

[00015] Techniques for metadata filtering and display management or mapping of high dynamic range (HDR) images are described herein. In the following description, for the purposes of explanation, numerous specific details are set forth in order to provide a thorough understanding of the present invention. It will be apparent, however, that the present invention may be practiced without these specific details. In other instances, well-known structures and devices are not described in exhaustive detail, in order to avoid unnecessarily occluding, obscuring, or obfuscating the present invention.

OVERVIEW

[00016] Example embodiments described herein relate to metadata filtering for the display management of HDR images. In an embodiment, given an input video stream and associated input metadata comprising at least one of minimum (*min*), average (*mid*), or maximum (*max*) luminance values of the video frames in the video stream, the input metadata of the input frames are filtered to generate output metadata, wherein the filtering is based only on input metadata from input frames belonging to the same scene. Methods for temporal filtering based on exponential moving average filters or a look-ahead sliding window filter are presented, including methods for scene-change detection using the input metadata.

EXAMPLE DISPLAY CONTROL AND DISPLAY MANAGEMENT

[00017] FIG. 1 depicts an example process (100) for display control and display management according to an embodiment. Input signal (102) is to be displayed on display (120). Input signal may represent a single image frame, a collection of images, or a video signal. Image signal (102) represents a desired image on some source display typically defined by a signal EOTF, such as ITU-R BT. 1886 or SMPTE ST 2084, which describes the relationship between color values (e.g., luminance) of the input video signal to output screen color values (e.g., screen luminance) produced by the target display (120). The display may be a movie projector, a television set, a monitor, and the like, or may be part of another device, such as a tablet or a smart phone.

[00018] Process (100) may be part of the functionality of a receiver or media player connected to a display (e.g., a cinema projector, a television set, a set-top box, a tablet, a smart-phone, a gaming console, and the like), where content is consumed, or it may be part of a content-creation system, where, for example, input (102) is mapped from one color grade and dynamic range to a target dynamic range suitable for a target family of displays (e.g., televisions with standard or high dynamic range, movie theater projectors, and the like).

[00019] In some embodiments, input signal (102) may also include metadata (104). These can be signal metadata, characterizing properties of the signal itself, or source metadata, characterizing properties of the environment used to color grade and process the input signal (e.g., source display properties, ambient light, coding metadata, and the like).

[00020] In some embodiments (e.g., during content creation), process (100) may also generate metadata which are embedded into the generated tone-mapped output signal. A target display (120) may have a different EOTF than the source display. A receiver needs to account for the EOTF differences between the source and target displays to accurate display the input image. Display management (115) is the process that maps the input image into the target display (120) by taking into account the two EOTFs as well as the fact that the source and target displays may have different capabilities (e.g., in terms of dynamic range.)

[00021] As used herein, the terms “display management” or “display mapping” denote the processing (e.g., tone and gamut mapping) required to map an input video signal of a first dynamic range (e.g., 1000 nits) to a display of a second dynamic range (e.g., 500 nits). Examples of display management processes can be found in United States Provisional Patent

Application Ser. No. 62/105,139 (to be referred to as the '139 Application), filed on Jan. 19, 2015, titled "Display management for high dynamic range images," also filed as PCT Application PCT/US2016/013352, on Jan 14, 2016, each of which is incorporated herein by reference in its entirety.

[00022] In some embodiments, the dynamic range of the input (102) may be lower than the dynamic range of the display (120). For example, an input with maximum brightness of 100 nits in a Rec. 709 format may need to be color graded and displayed on a display with maximum brightness of 1,000 nits. In other embodiments, the dynamic range of input (102) may be the same or higher than the dynamic range of the display. For example, input (102) may be color graded at a maximum brightness of 5,000 nits while the target display (120) may have a maximum brightness of 1,500 nits.

[00023] In an embodiment, display (120) is controlled by display controller (130). Display controller (130) provides display-related data (134) to the display mapping process (115) (such as: minimum and maximum brightness of the display, color gamut information, and the like) and control data (132) for the display, such as control signals to modulate the backlight or other parameters of the display for either global or local dimming. An example of a display controller for dual modulation display systems is described in US Patent 8,493,313, "Temporal filtering of video signals," by G. Damberg and H. Seetzen, which is incorporated herein by reference in its entirety. Another example is described in PCT Application Ser. No. PCT/US2014/012568 (WO 2014/116715A1), filed on Jan. 22, 2014, "Global display management based light modulation," by T. Kunkel, which is incorporated herein by reference in its entirety.

[00024] Displays using global or local backlight modulation techniques adjust the backlight based on information from input frames of the image content and/or information received by local ambient light sensors. For example, for relatively dark images, the display controller (130) may dim the backlight of the display to enhance the blacks. Similarly, for relatively bright images, the display controller may increase the backlight of the display to enhance the highlights of the image. For example, FIG. 2 depicts a display with a full dynamic range of $Max_{\mathrm{BL}}/Min_{\mathrm{BL}}$, where Max_{BL} denotes the maximum possible luminance of the display when the backlight is full on (e.g., 1,000 nits), and Min_{BL} denotes the minimum possible luminance of the display when the backlight is full off (e.g., 0.001 nits). Note that, for illustration purposes, the Y axis of FIG. 2 depicts log luminance values. As used herein, the term "instantaneous dynamic range" denotes a simultaneous (or static) dynamic range, that is the range of black (e.g., $Min_{\mathrm{T}}(K)$) to white (e.g., $Max_{\mathrm{T}}(K)$) that can be shown simultaneously on a display for a given backlight level (K) at some instance of time. Let $Max_T(K) / Min_T(K)$ (e.g., 1,000:1) denote the instantaneous (also referred to as static) dynamic range of the display (120) for a specific level K (205) of backlight between full off and full on, then display management (115) maps the dynamic range of the input (102) into the range ($Min_T(K)$, $Max_T(K)$) of the display. The instantaneous dynamic range ratio $Max_T(K) / Min_T(K)$ may be fixed or approximately the same for all values of K. For example, for an instantaneous dynamic range of 1,000:1, $w = log(1{,}000)/2 = 1.5$.

[00025] In an embodiment, the display mapping (115) and display control (130) processes are enhanced by suitable image analysis (105) and image processing (110) operations as will be described herein.

IMAGE ANALYSIS

[00026] In an embodiment, unless specified already by the source metadata (104), for each input frame in signal (102) the image analysis (105) block may compute its minimum (*min*), maximum (*max*), and median (*mid*) (or average gray) luminance value. These values may be computed for the whole frame or part of a frame. In some embodiments, *min*, *mid*, and *max* luminance values may represent approximate values of the true values. For example, computed *min* and *max* values may represent 90% of the true *min* and *max* values in the input signal so as to be more robust to single pixel outliers.

[00027] In some embodiment, *min*, *mid*, and *max* luminance signal values may also be computed or received as metadata for a whole scene. As used herein, the terms ‘scene’ or ‘shot’ denote a series of sequentially-in-time captured sequence frames that may share the same overall color or brightness characteristics. Authored content, such as films and pre-recorded video can be edited in such a way that image statistics may be computed over a cohesive set of frames, such as a scene or a “cut,” which may prevent temporal artifacts; however, in computer games and live broadcast, there might not be enough information to have pre-determined scene cuts, so better adaptation techniques are required.

[00028] Scene cuts may be determined automatically or they may be denoted in the bitstream using metadata. Automatic scene change detection is a challenging and well-studied problem. Embodiments of this invention can easily tolerate missed scene cuts or false detected scene cuts, hence the exact method of scene-cut detection is not particularly important; nevertheless, without limitation, a variety of scene cut detection mechanisms are suggested herein.

[00029] For example, let Mid_{S1} and Mid_{S2} denote respectively the mid luminance values for two consecutive scenes $S1$ and $S2$, then, in an embodiment a scene cut may determined if:

$$Mid_{S1} - Mid_{S2} \neq 0, \qquad (1)$$

or

$$|Mid_{S1} - Mid_{S2}| > T_F,$$

where T_F is a predefined threshold (e.g., $T_F = 0.1$).

METADATA FILTERING

[00030] As discussed earlier, display management tone-mapping methods require information about the input video in order to produce a stable, pleasing result, on a target device. For example, such information may include, for each frame, a triplet of values indicating the minimum (*min*), average (*mid*), and peak (*max*) luminance values.

[00031] Given *min*, *mid*, and *max* luminance source data (107 or 104), image processing block (110) may compute the display parameters (e.g., Min_T and Max_T, or the level K of backlight) that allow for the best possible environment for displaying the input video. Due to brightness fluctuations even within frames in the same scene, treating each frame independently may lead to flickering and other unwanted visual artifacts. In an embodiment, a temporal filter is applied to a sequence of sequential frames in the scene to determine the best luminance mapping (e.g., Min_T and Max_T). In one embodiment, luminance mapping employs a temporal filter based on an exponential moving average (EMA) filter; however, other FIR or IIR temporal filters, as will be discussed later on, could be applied as well. In some embodiments, temporal filtering and other aspects of luminance range mapping (110) may applied at the source display, and the filter output data may be passed to the target display as metadata. This allows for fewer computations at the target display and additional creative control by the content provider. For example, the content creator (e.g., a director or a color grader) may decide to override the results of the filter output (110) to manually adjust how the image is displayed.

[00032] Let $L_{F(t)}$ denote a function of *min*, *mid*, and *max* luminance values in a frame at time t in a scene. In an embodiment $L_{F(t)}$ may be simply the *mid* luminance value of a frame at time t in a scene (e.g., $L_{F(t)} = Mid_{F(t)}$). In other embodiments, $L_{F(t)}$ may represent the *min* or *max* values, or a weighted combination of the *min*, *mid*, and *max* values. Then, in an embodiment, EMA filtering in a scene may be expressed as:

$$S_0 = L_{F(0)}, \text{ for } t = 0,$$

$$S_t = \alpha * L_{F(t)} + \beta * S_{t-1}, \text{ for } t > 0 \quad , \qquad (2)$$

where α (alpha) and β (beta) denote weight factors, and $t = 0$ denotes the beginning of the current scene.

[00033] In an embodiment,

$$\beta = 1 - \alpha.$$

In some embodiments, the weights may be fixed (e.g., $\alpha = 0.25$, $\beta = 0.75$). In some embodiments the weights may be selected from a fixed list of possible weights. For example, for $L_{F(t)} = Mid_{F(t)}$ the alpha value may be fixed (e.g. $\alpha = 0.25$), but for $L_{F(t)} = Max_{F(t)}$ and $L_{F(t)} = Min_{F(t)}$ the value of alpha may switch between two or more values, say $\alpha_1 = 0.175$ and $\alpha_2 = 0.475$. This will be referred to as asymmetric alpha. For example, in an embodiment that uses two asymmetric alpha values, if $S_t > S_{t-1}$, then for the next data point $\alpha = \alpha_2$, otherwise $\alpha = \alpha_1$. This allows tone-mapping operations to adapt quicker to new increased highlights or lower darks in the input image sequences.

[00034] FIG. 4A depicts examples of EMA filtering for the *min* (405-a), *mid* (410-a), and *max* (415-a) luminance values for about 200 frames in a video sequence. Filtering of the *mid* value uses a fixed alpha, while filtering of the min and max values uses asymmetric alpha values, as described earlier.

[00035] In some embodiments the weights may be a function of the delivery frame rate. Such an implementation is especially important for video streaming applications where the frame rate may change dynamically based on either computational resources or available bandwidth. For example, if α_M denotes a weight factor optimized for a delivery at M frames per second (e.g., $M = 24$), and R denotes the actual delivery rate (e.g., $R = 30$), then using a linear conversion:

$$\alpha = a_M * \frac{M}{R},$$

which allows alpha values to decrease when the actual frame rate increases.

[00036] In some embodiments β may be defined to be a function of time. For example, in an embodiment:

$$\beta = 0, \text{ for } t = 0$$
$$\beta = clip3(0, \frac{1-\alpha}{t}, (1-\alpha), \text{ for } 0 < t \leq m, \qquad (3)$$
$$\beta = clip3(0, \frac{1-\alpha}{m}, (1-\alpha)), \text{ for } t > m,$$

where $m > 0$ is a predetermined time instant and clip3(a,$f(x)$,c) denotes that the output of $f(x)$ is always clipped to be within the values of a and c, where a and c are included.

[00037] In some embodiments, the alpha value of the EMA filter may be reset or adjusted when a new scene cut or scene change is detected. For example, in an embodiment:

$$\alpha = min(1, SceneCut * |S_{t-1} - L_{F(t)}| * \alpha_{scene} + \alpha_{base}), \qquad (4a)$$

where *SceneCut* is in the range (0, 1) and denotes the confidence (or probability) in detecting a scene cut. For example, *SceneCut* =1 may specify there is a new scene with full confidence. Parameters α_{scene} and α_{base} denote predefined filter parameters that control how fast the filter adapts. In an embodiment, without limitation, typical ranges for these variables include $\alpha_{scene} = (2.0, 5.0)$ and α_{base} = (0.02, 0.2) (e.g., $\alpha_{scene} = 3.0$ and α_{base} =0.05). Hence, when a new scene is detected, the value of α may be increased proportionally to the change of the scene-related metadata (e.g., the average scene luminance) to make smoother the transition between the adjustment in mid brightness values. In some embodiments, in equation (4), S_{t-1} may also be substituted with $L_{F(t-1)}$. From equations (2)-(4), when a new scene is detected, α is getting very close to one and the value of β is close to zero, hence, the current $L_{F(t)}$ values are weighted more than past filtered values. In addition, when a new scene cut is detected, t may be reset to 0, and all of the previous S_t values may be cleared from the memory. In other words, optionally, the memory of the temporal filter may be reset to zero every time there is a scene cut.

[00038] As an example, FIG. 4B depicts examples of EMA filtering for the *min* (405-b), *mid* (410-b), and *max* (415-b) luminance values for about 200 frames in a video sequence (same as the one depicted in FIG. 4A) when taking into consideration when a scene-cut is detected.

[00039] As depicted in FIG. 4B, scenes (e.g., S1 to S4) in the video sequence may also be characterized and labeled in terms of their statistical features. For example, scene S4 may be determined to be statistically equivalent to scene S2 in terms of average *min*, *mid*, and *max* values or of a function of them (e.g., both are labeled as scene "B"). For example, from an implementation point of view, the image analysis unit (315) may keep a scene history log and associated statistics (e.g., averages of *min*, *mid*, *max*, and S_t values). This type of characterization may be used for a far more effective initialization of the EMA filter in equation (2) than always starting with the default (e.g., $S_0 = L_{F(0)}$, for $t = 0$). For example, one may initiate the EMA filter based on statistical data from a previous scene that bears the strongest resemblance to the new scene. For example, for the sequence of scenes in FIG. 4B, at the start of scene S4, one may have $S_0 = L_{S2}$, for $t = 0$, where L_{S2} represents a suitable value computed based on the statistics of scene S2, which is considered to be the closest to scene S4.

[00040] In another embodiment, α may be defined as a function of time (α_t). For example,

$$\alpha_t = a_{def} + (\alpha_{max} - a_{def})SceneCut(t), \tag{4b}$$

where a_{def} denotes a default value, α_{max} denotes a maximum value, and as before, *SceneCut*(*t*) denotes a probability of the frame at time *t* to belong to a new scene. This allows again for faster, but smoother, adaptation of the EMA filter to scene cuts or sudden changes to the luminance values of an input picture. If a frame has low probability to belong to a scene cut, then the default alpha parameter is being used, otherwise, for definite scene cuts, an alpha value closer to the α_{max} value is being used.

[00041] Let Min_S, Mid_S, and Max_S denote the brightness characteristics of a source or reference display, and let Min_T , Mid_T, and Max_T denote the brightness characteristics of the target display (120), then, as described by A. Ballestad et al., in U.S. Patent 8,593,480, titled "Method and apparatus for image data transformation," which is incorporated herein by reference in its entirety, these values may define the anchor points of a sigmoid-like, tone-mapping function, which together with other tone-mapping operations (e.g., as described in the '139 Application) enable the display management process (115) to generate a tone-mapped output (117) to be displayed on the target display (120).

[00042] In an embodiment, given the results of the temporal filter (e.g., equation (2)), the preferred instantaneous luminance range for the target display (120) may be computed as

$$Max_T = clip3(Min_{BL}, f_{max}(S_t), Max_{BL}) , \tag{5}$$
$$Min_T = clip3(Min_{BL}, f_{min}(S_t), Max_{BL}),$$

where $f_{max}(S_t)$, and $f_{min}(S_t)$, denote functions to determine the *max* and *min* values of the preferred instantaneous dynamic range of the target display based on one or more limit luminance values for the target display (e.g., Min_{BL}, Max_{BL}). For example, without limitation, assuming all display luminance values and S_t are in expressed in a linear domain (shown with an overbar) (e.g., $\overline{Max_{BL}} = 10^{Max_{BL}}$ when Max_{BL} is in log10 domain), let $\overline{w}$ denote one half of the instantaneous dynamic range in the linear domain (e.g., see FIG. 2). Then, if

$$S_t = clip3(\overline{Min_{BL}} * \overline{w}, S_t, \frac{\overline{Max_{BL}}}{\overline{w}}) ,$$

then

$$\overline{Max_T} = \overline{f_{max}}(S_t) = S_t * \overline{w} , \tag{6}$$

$$\overline{Min_T} = \overline{f_{min}}(S_t) = S_t/\overline{w}\,.$$

[00043] If S_t values are computed in a gamma or other perceptually-quantized luminance space, then they may have to be linearized first. Alternatively, equations (5)-(6) may also be computed in a logarithmic domain. For example, assuming all luminance values are expressed in logarithmic space, let w in denote one half of the instantaneous dynamic range in the logarithmic domain. Then if

$$log(S_t) = clip3(Min_{BL} + w, log(\,S_t), Max_{BL} - w),$$

then

$$Max_T = f_{max}(S_t) = log(S_t) + w, \tag{7}$$
$$Min_T = f_{min}(S_t) = log(S_t) - w.$$

For example, let a display have $\overline{Max_{BL}}$ = 1,000 nits, $\overline{Min_{BL}}$ = 0.001 nits, and an instantaneous dynamic range of 1,000:1. Then Max_{BL} = 3, Min_{BL} = -3, w = log(1,000)/2 = 1.5, and $\overline{w} = \sqrt{1000} = 31.62$.

[00044] Given the Min_T and Max_T values (111) computed by equations (6) or (7), the display controller (130) may then apply a look-up table or other internal-control processes to determine the appropriate level *K* for controlling the display's backlight. Alternatively, in a content-creation environment, S_t–related values or one or more of the computed Min_T and Max_T values or a function of these values (e.g., Mid_T) may be embedded as metadata in the tone-mapped bitstream to be delivered downstream to content consumers. Hence, a receiver with low computational resources, such as a tablet or a smartphone, may use directly these values to determine the best display setup.

[00045] FIG. 3 depicts an example process for adjusting the instantaneous luminance range mapping according to an embodiment. Given input (102) and associated metadata (104), step (310) determines whether luminance-related metadata values (e.g., *min*, *mid*, and *max*) for each frame are available. If there are no relevant metadata, then these values may be computed in step (315). In some embodiments, computing luminance-related metadata in (315) may also require a luminance linearization step, where a gamma or other non-linear mapping (as specified by the source's EOTF) is removed to generate data with linear luminance. In some embodiments, the results of image analysis (315) may also be stored and/or passed downstream as signal-related metadata (304). Next, in step (345), it is determined whether the current frame starts a new scene. If there is a scene change, optionally, in step (340), the temporal filter may also be fully reset (e.g., the memory of all past S_t filter output values is cleared and t is set to zero). Next, in step (325), the parameters of the temporal filter are adjusted as needed (e.g., equations (3) and (4)) and the process continues with the temporal filtering (320) (e.g., equation (2)) of values of the $L_{F(t)}$ function of these metadata values. In step (330), the results (322) of the temporal filtering (e.g., S_t) are used to determine the proper instantaneous luminance range (111) (e.g., Min_T and Max_T) to display the input data (102) on the target display. Appropriate data (111) are passed to the display controller (130) and the display management process (115). Optionally, related metadata (304) (e.g., the filtered $L_{F(t)}$ values for each frame) may be saved to be stored together with the modified (tone-mapped) by the display management bitstream to be passed to other downstream components.

[00046] While example embodiments have been presented for optimizing the display of images (either of standard dynamic range (SDR) or high dynamic range) on high-dynamic range displays, the same techniques may also be applied to improve SDR displays. For example, when viewing a display under high ambient light (e.g., a tablet or smartphone in day light), the techniques may be used to compensate for the low dynamic range caused by the high ambient light and the display's reflectivity parameters.

PREVIEW FRAMES AND FORWARD AND REVERSE EMA FILTERING

[00047] If one can have a preview of the next N frames of the source material, then the image analysis unit (315) may apply this information to improve scene-change detection performance, statistics gathering, and filter initialization. In an embodiment, one may compute the moving average of equation (2) in both time-forward and time-reverse order on the upcoming frames, facilitating detection of when a significant change in scene content takes place. For example, one may compute:

$$S_0^f = L_{F(0)}, \text{ and}$$

$$S_0^r = L_{F(N)}, \text{ for } t = 0,$$

$$S_t^f = \alpha * L_{F(t)} + (1-\alpha) * S_{t-1}^f \quad, \text{ and}$$

$$S_t^r = \alpha * L_{F(N-t)} + (1-\alpha) * S_{t-1}^r \quad, \text{ for } t > 0 . \qquad (8)$$

[00048] That is, in the time-reverse EMA filter (S_t^r), future preview frames are added one-by-one, starting from the most future one (N) and working backwards towards the current frame. If the time-reverse moving average (S_t^r) is getting closer to the time-forward moving average (S_t^f), then one can determine that there is no scene change moment. Likewise, if the two moving averages are within a distance threshold of each other, then one may continue normal in-scene progression; however, when one detects a large difference between the forward and reverse moving average metrics, then one with high confidence can determine there is a scene cut between the current frame and the N-th future frame, typically occurring at the maximum of their distance (e.g. $|S_t^f - S_t^r|$ and the like).

[00049] When such a scene change is detected, say at frame k, as described earlier, one can start the EMA filter for the new scene based on the current frame (e.g., $S_0^f = L_{F(k)}$). Alternatively, one may choose to initialize the EMA filter based on the already computed EMA values of the preview frames (e.g., $S_0^f = S_{N-k}^r$). This approach represents an improvement over starting fresh with only one frame's statistics, and in most cases outperforms the cached history approach described earlier.

METADATA FILTERING USING REAL-TIME, LOOK-AHEAD, SLIDING WINDOW

[00050] FIG. 5 depicts an example of a look-ahead sliding window (510) for filtering metadata related to display management. Let frames $f(0)$ to $f(F\text{-}1)$ all belong to the current scene. Let $f(j)$ denote the current frame. In an embodiment, without limitation, the sliding filtering window may include a total of up to $P+A+1$ frames, where A (e.g., $A=10$) denotes the number of future frames, or frames ahead of frame j (e.g., up to frame $f(j+A)$), and P (e.g., $P=30$) denotes the number of frames prior to frame j (e.g., up to frame $f(j\text{-}P)$). In practice, at the beginning of a scene (e.g., close to $f(0)$), there might not be enough previous frames, and towards the end of the scene (e.g., close to $f(F\text{-}1)$), there might not be enough future frames. For $j=0$ to $F\text{-}1$, let

$$S = max(j\text{-}P, 0) \quad , \qquad (9)$$
$$E = min(F\text{-}1, j+A),$$

denote the starting and ending frame indexes of the filtering sliding window (510) within a scene in between frames 0 and F-1. Let $N=E\text{-}S+1$ denote the length of the filter. Let $L_{F(j)}$ denote input metadata as a function of the *min*, *mid*, and *max* luminance values for frame j, then the filtered metadata may be derived as

$$S_j = \frac{1}{N}\sum_{i=S}^{E} w(N, i-S) * L_{F(i)}, \qquad j=0, 1, \ldots, F\text{-}1 \qquad (10)$$

where $w(N,k)$, for $k=0, 1, \ldots, N\text{-}1$ denote weighting factors that depend on the effective length (N) of the filter, and

$$\frac{1}{N}\sum_{i=S}^{E} w(N, i-S) = 1.$$

For example, for $w(N, k) = 1$, for all k and all N, corresponds to applying a sliding, moving-average filter. From equation (9), assuming $A < P$, the length of the filter (N), may range from $N=A+1$ to $N=A+P+1$.

[00051] Compared to the EMA filter or a scene-based average filter, the slide-window filter in equation (10) is much simpler to implement and behaves much better during fade-ins or other fast-changing scenes. The use of future frames allows the filter to take into consideration future statistics and hence respond faster. In addition, by a simple adjustment of the number of forward (A) and past frames (P), the filter can easily by optimized to satisfy any real-time processing requirements, say, for broadcasting or gaming applications.

IMPROVED SCENE-CHANGE DETECTION

[00052] In another embodiment, a scene-cut detection technique is based on statistical differences between the current frame (or a collection of future preview frames) and an existing moving average characteristic. For example, the following statistics may be computed for each input frame:

a) A 1-D log-luminance histogram (say, h_i^Y) representing a histogram of log-luminance values (e.g., log(Y) of a frame in the YCbCr domain)

b) A 2-D CIE (u´v´) chromaticity histogram (say, h_j^{uv}), and

c) An edge strength histogram (say, h_k^e)

[00053] The log-luminance histogram serves both as a measure of scene content and as a resource for tone-mapping in display management. The two-dimensional CIE (u´v´) chromaticity histogram is less common, but serves here to summarize the color content of a scene. The edge strength histogram represents a representation of a frame in terms of its edge content. An example of computing edge-strength histograms is described in Lee, Seong-Whan, Young-Min Kim, and Sung Woo Choi. "*Fast scene change detection using direct feature extraction from MPEG compressed videos*," in Multimedia, IEEE Transactions" on 2.4 (2000): 240-254, which is incorporated herein by reference in its entirety. In an embodiment, an edge-strength histogram may be computed as follows:

a) Take the luminance plane of the input picture (say, Y)

b) Subsample it, to reduce its resolution, say, by a factor of *m* in each dimension, e.g., (*m*=6) to generate *Yr*. These two steps allow one to reduce the computational complexity of this algorithmic step at the expense of some loss in accuracy.

c) Apply an edge kernel filter to *Yr* (say, the Sobel edge filter), to generate an edge map (say, Yr^e)

d) Compute a histogram of the edge map (say, h_k^e)

[00054] In an embodiment, a weighted Euclidian distance (say, Δ_t) between each of these histograms from the current frame (or preview frames) and the moving average of the previous frames is computed to determine whether a scene cut is appropriate at a given point in time. For example, in an embodiment, let

$$\Delta_t = w_1\Delta_t^Y + w_2\Delta_t^{uv} + w_3\Delta_t^e, \tag{11}$$

where w_i , for *i*=1, 2, and 3, denotes the weighting factors (e.g., $w_1 = w_2 = 0.35$ and $w_3 =$ 0.3), Δ_t^Y denotes a measure of the distance (e.g., L1, L2, and the like) between the luminance histograms of the current frame (e.g., at time *t*) and a previous frame (e.g., at time *t*-1) (e.g., using the L2 distance, $\Delta_t^Y = \sum_i (h(t)_i^Y - h(t-1)_i^Y)^2$), Δ_t^{uv} denotes a corresponding distance of the chroma histograms, and Δ_t^e denotes the distance between the corresponding edge histograms. In some embodiments, histogram values in each histogram are filtered by a low – pass filter before computing distance values to improve scene-cut detection.

[00055] Given Δ_t from equation (11), a scene cut may be determined if Δ_t is larger than a predetermined threshold. Alternatively, the probability of a scene cut may be determined as a function of Δ_t. For example, the variable *SceneCut* in equations (4a) and (4b) may be determined as

$$SceneCut = \frac{\Delta_t}{\Delta_t + C}, \tag{12}$$

where *c* is a tunable constant.

EXAMPLE COMPUTER SYSTEM IMPLEMENTATION

[00056] Embodiments of the present invention may be implemented with a computer system, systems configured in electronic circuitry and components, an integrated circuit (IC) device such as a microcontroller, a field programmable gate array (FPGA), or another configurable or programmable logic device (PLD), a discrete time or digital signal processor (DSP), an application specific IC (ASIC), and/or apparatus that includes one or more of such systems, devices or components. The computer and/or IC may perform, control, or execute instructions relating to metadata filtering and display mapping processes, such as those described herein. The computer and/or IC may compute any of a variety of parameters or values that relate to metadata filtering and display mapping processes described herein. The image and video embodiments may be implemented in hardware, software, firmware and various combinations thereof.

[00057] Certain implementations of the invention comprise computer processors which execute software instructions which cause the processors to perform a method of the invention. For example, one or more processors in a display, an encoder, a set top box, a transcoder or the like may implement methods related to metadata filtering and display mapping processes as described above by executing software instructions in a program memory accessible to the processors. The invention may also be provided in the form of a program product. The program product may comprise any non-transitory medium which carries a set of computer-readable signals comprising instructions which, when executed by a data processor, cause the data processor to execute a method of the invention. Program products according to the invention may be in any of a wide variety of forms. The program product may comprise, for example, physical media such as magnetic data storage media including floppy diskettes, hard disk drives, optical data storage media including CD ROMs, DVDs, electronic data storage media including ROMs, flash RAM, or the like. The computer-readable signals on the program product may optionally be compressed or encrypted.

[00058] Where a component (e.g. a software module, processor, assembly, device, circuit, etc.) is referred to above, unless otherwise indicated, reference to that component (including a reference to a "means") should be interpreted as including as equivalents of that component any component which performs the function of the described component (e.g., that is functionally equivalent), including components which are not structurally equivalent to the disclosed structure which performs the function in the illustrated example embodiments of the invention.

EQUIVALENTS, EXTENSIONS, ALTERNATIVES AND MISCELLANEOUS

[00059] Example embodiments that relate to efficient metadata filtering and display mapping processes are thus described. In the foregoing specification, embodiments of the present invention have been described with reference to numerous specific details that may vary from implementation to implementation. Thus, the sole and exclusive indicator of what is the invention, and is intended by the applicants to be the invention, is the set of claims that issue from this application, in the specific form in which such claims issue, including any subsequent correction. Any definitions expressly set forth herein for terms contained in such claims shall govern the meaning of such terms as used in the claims. Hence, no limitation, element, property, feature, advantage or attribute that is not expressly recited in a claim should limit the scope of such claim in any way. The specification and drawings are, accordingly, to be regarded in an illustrative rather than a restrictive sense.

CLAIMS

What is claimed is:

1. A method for metadata filtering with a processor, the method comprising:

accessing with a processor an input video stream (102) comprising video frames;

accessing input metadata for the input video stream, the input metadata comprising at least one of minimum (*min*), average (*mid*), or maximum (*max*) luminance values of the video frames in the input video stream; and

for an input frame in the input video stream:

testing whether the input frame belongs to a current scene or a new scene; and upon determining the input frame belongs to the current scene:

filtering the input metadata to generate output metadata for the input frame, wherein the filtering is based only on metadata from input frames belonging to the current scene.

2. The method of claim 1, wherein the at least one of *min*, *mid*, or *max* luminance values are generated by an image analysis processor (105) based on luminance values of the video frames in the input video stream.

3. The method of claim 1, wherein metadata filtering comprises computing an output S_t with an exponential moving average filter as:

$S_0 = L_{F(0)}$, for $t = 0$,

$S_t = \alpha_k * L_{F(t)} + \beta * S_{t-1}$, for $t > 0$,

where α_k (alpha) and β (beta) denote weight factors, $L_{F(t)}$ denotes a function of the at least one of *min*, *mid*, or *max* luminance values for the input frame at time t, and t=0 denotes the start of the current scene, wherein the alpha (α_k) value comprises two or more asymmetric values, wherein the alpha value increases when filter output values are expanding and the alpha value decrease when the filter output values are contracting.

4. The method of claim 3, wherein $\beta = 1 - \alpha_k$.

5. The method of claim 3, wherein an alpha weight value is computed as a function of a delivery frame rate, wherein the alpha weight value decreases when the delivery frame rate increases.

6. The method of claim 3, wherein instead of initializing the EMA filter based on the luminance values of a frame starting the current scene, the EMA filter is initialized based on the luminance values of frames in a past scene, wherein the past scene is statistically similar to the current scene.

7. The method of claim 3, wherein an alpha weight value is generated as a function of time.

8. The method of claim 7, wherein generating the alpha weight value comprises computing

$$\alpha_t = a_{def} + (\alpha_{max} - a_{def}) SceneCut(t), \tag{4b}$$

where a_{def} denotes a default alpha weight value, α_{max} denotes a maximum alpha weight value, and *SceneCut*(*t*) comprises a value in the range (0, 1) denoting a measure of confidence the input frame *t* does not belong in the current scene.

9. The method of claim 1, wherein determining whether the input frame belongs to the current scene comprises:

computing luminance, chromaticity, and edge-strength histograms for the video frames in the input video stream;

computing a weighted Euclidian distance between histogram values of the input frame and a prior frame; and

determining that the weighted Euclidian distance does not exceed a threshold value.

10. The method of claim 9, further comprising determining a probability measure of whether the input frame belongs to the current scene, the method comprising computing

$$SceneCut = \frac{\Delta_t}{\Delta_t + C},$$

where *c* is a tunable constant, $SceneCut$ denotes the probability measure, and Δ_t denotes the weighted Euclidian distance between histogram values of the input frame and the prior frame.

11. The method of claim 1, wherein filtering comprises an exponential moving average (EMA) filter and the output metadata values comprise a time-forward moving average value

based on the input frame and past frames and a time-reverse moving average value based on the input frame and future preview frames, and determining whether the input frame belongs to the current scene is based on a measure of a distance between the time-forward moving average value and the time-reverse forward moving average value.

12. The method of claim 11, wherein upon detecting the input frame belongs to a new scene, initialization of the EMA filter to compute forward moving average values for the new scene is based on the reverse moving average value for the previous scene.

13. The method of claim 1, wherein metadata filtering comprises a look-ahead sliding window comprising L frames in the current scene prior to the input frame and A frames in the current scene subsequent to the input frame in the current scene.

14. The method of claim 13, wherein metadata filtering comprises computing

$$S_j = \frac{1}{N}\sum_{i=S}^{E} w(N, i-S) * L_{F(i)},\ j=0, 1, \ldots, F-1,$$

where, $L_{F(i)}$ denotes an input metadata value for input frame i, $w(N, k)$, for k=0, 1, …, N-1, denotes filter weights, and S_j denotes the output metadata value for frame j, and wherein given the start of the current scene at $j=0$ and the end of the current scene at $j=F-1$, $S = max(j-P,0)$, $E=min(F-1, j+A)$, and $N=E-S+1$.

15. The method of claim 1, further comprising generating using the output metadata an output tone-mapped frame by mapping the dynamic range of the input frame to the dynamic range of a target display.

16. An apparatus comprising a processor and configured to perform any one of the methods recited in claims 1-15.

17. A non-transitory computer-readable storage medium having stored thereon computer-executable instruction for executing a method in accordance with any one of the claims 1-15.

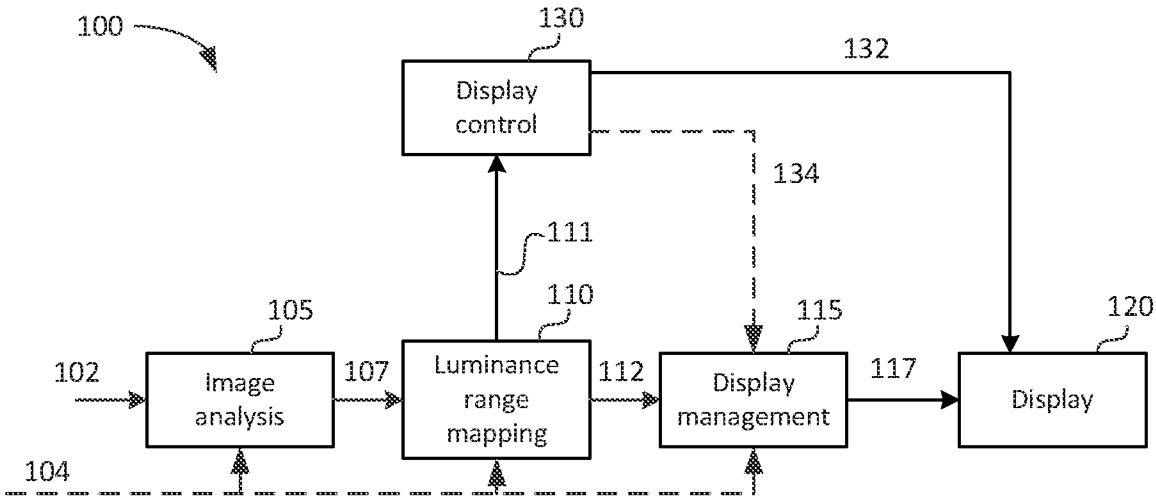
100
130
Display control
132
134
111
110
105
115
120
102
Image analysis
107
Luminance range mapping
112
Display management
117
Display
104

FIG. 1

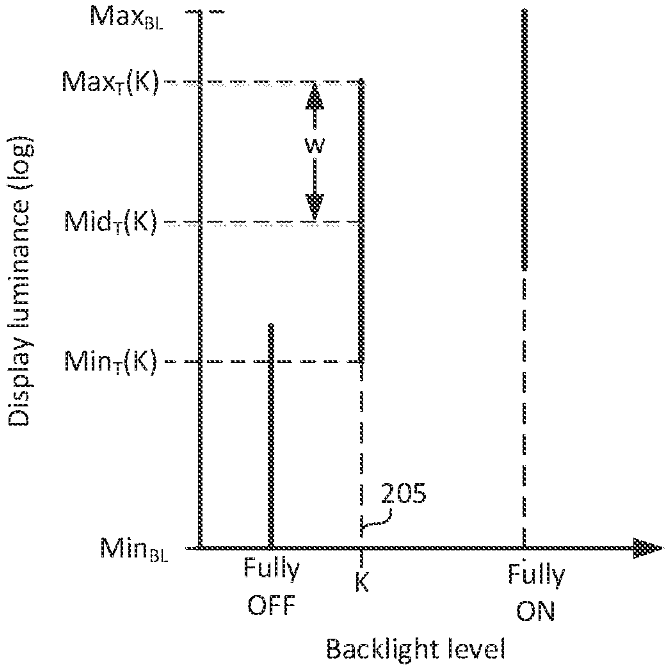
Max_BL
Max_T(K)
w
Mid_T(K)
Min_T(K)
Display luminance (log)
205
Min_BL
Fully OFF
K
Fully ON
Backlight level

FIG. 2

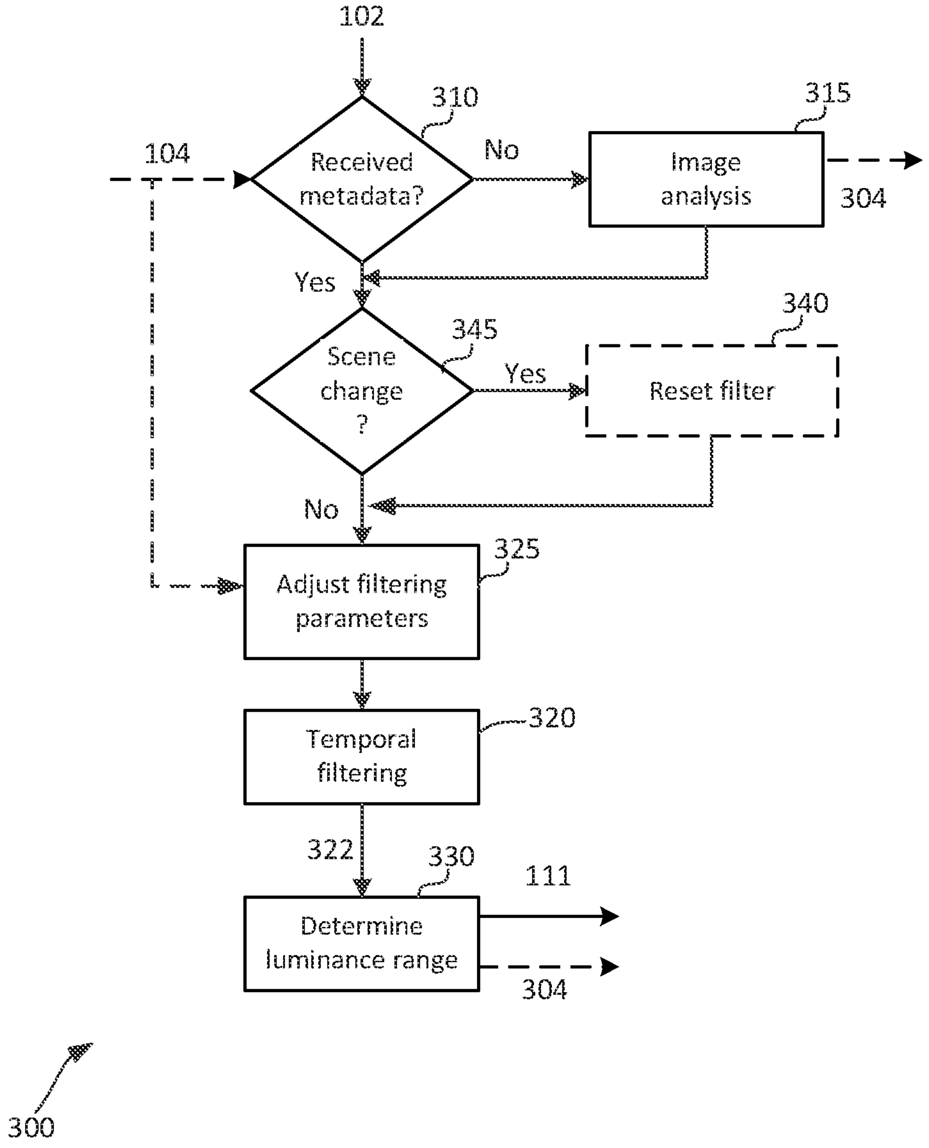
102
310
Received metadata?
No
315
Image analysis
304
104
Yes
345
Scene change ?
Yes
340
Reset filter
No
325
Adjust filtering parameters
Temporal filtering
320
322
330
Determine luminance range
111
304
300

FIG. 3

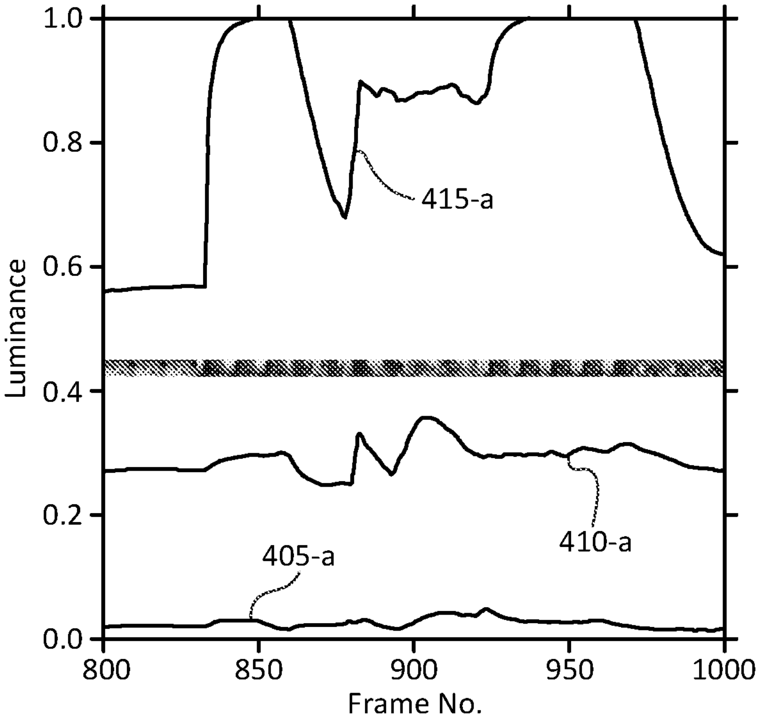
1.0
0.8
0.6
0.4
0.2
0.0
Luminance
800
850
900
950
1000
Frame No.
415-a
410-a
405-a

FIG. 4A

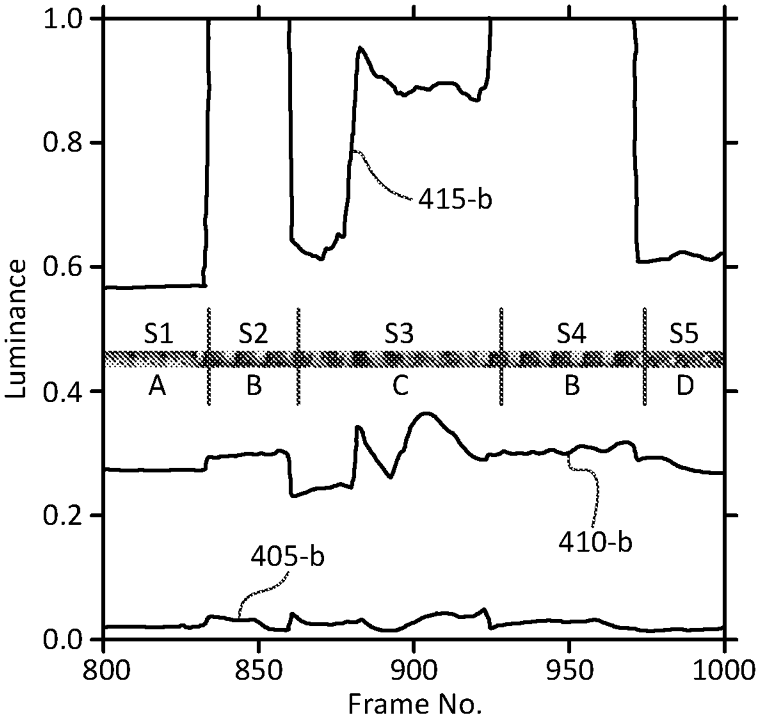
1.0
0.8
0.6
0.4
0.2
0.0
Luminance
800
850
900
950
1000
Frame No.
415-b
S1
S2
S3
S4
S5
A
B
C
B
D
410-b
405-b

FIG. 4B

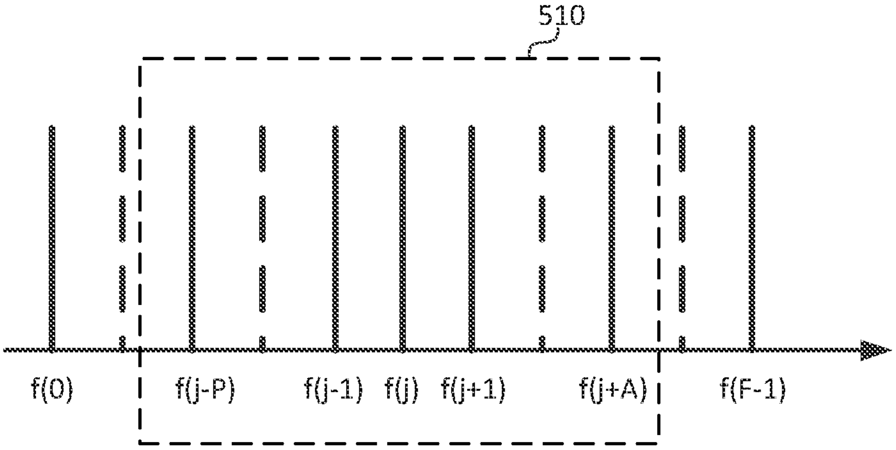

FIG. 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2016/031925

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

1-12, 15-17

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☒ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2016/031925

A. CLASSIFICATION OF SUBJECT MATTER
INV. G09G3/36 G06K9/00
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G09G H04N G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y A	US 2009/141178 A1 (KEROFSKY LOUIS J [US] ET AL) 4 June 2009 (2009-06-04) paragraphs [0019], [0120], [0201], [0473], [0490] - [0518]; figures 17,85,88-93	1-4,6-8, 15-17 9 5
X	WO 2015/017314 A1 (DOLBY LAB LICENSING CORP [US]) 5 February 2015 (2015-02-05) paragraphs [0069] - [0074]; figure 4	1,2, 15-17
X	US 2014/078165 A1 (MESSMER NEIL W [CA] ET AL) 20 March 2014 (2014-03-20) paragraphs [0005], [0011], [0042], [0068], [0100], [0124]; figure 2B	1,2, 15-17
A	WO 2014/176019 A1 (DOLBY LAB LICENSING CORP [US]) 30 October 2014 (2014-10-30) paragraphs [0017], [0028]; figure 2	1

-/--

☒ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search 12 October 2016	Date of mailing of the international search report 20/10/2016
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer Pichon, Jean-Michel

2

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2016/031925

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2013/130478 A1 (DOLBY LAB LICENSING CORP [US]) 6 September 2013 (2013-09-06) pages 5-6,9; figures 1,2 -----	1
A	US 6 683 657 B1 (MIYAWAKI MAMORU [JP]) 27 January 2004 (2004-01-27) column 10; figures 6-7 -----	3
Y	US 2007/201746 A1 (KIM YONG S [KR]) 30 August 2007 (2007-08-30) paragraphs [0001], [0044] - [0053] -----	9
A	JIAN ZHOU ET AL: "A web-enabled video indexing system", PROCEEDINGS OF THE 6TH. ACM SIGMM INTERNATIONAL WORKSHOP ON MULTIMEDIA INFORMATION RETRIEVAL. MIR. NEW YORK, NY, OCT. 15 - 16, 2004; [PROCEEDINGS OF THE ACM SIGMM INTERNATIONAL WORKSHOP ON MULTIMEDIA INFORMATION RETRIEVAL. MIR], NEW YORK, NY : ACM, U, 15 October 2004 (2004-10-15), pages 307-314, XP058138508, DOI: 10.1145/1026711.1026761 ISBN: 978-1-58113-940-2 pages 1-3 -----	9
A	XUE LING ET AL: "A Method for Fast Shot Boundary Detection Based on SVM", IMAGE AND SIGNAL PROCESSING, 2008. CISP '08. CONGRESS ON, IEEE, PISCATAWAY, NJ, USA, 27 May 2008 (2008-05-27), pages 445-449, XP031286735, ISBN: 978-0-7695-3119-9 pages 447-448 -----	9
A	TIAN-TSONG NG ET AL: "Identifying and prefiltering images", IEEE SIGNAL PROCESSING MAGAZINE, IEEE SERVICE CENTER, PISCATAWAY, NJ, US, vol. 26, no. 2, 1 March 2009 (2009-03-01), pages 49-58, XP011268065, ISSN: 1053-5888, DOI: 10.1109/MSP.2008.931077 page 53 -----	9

2

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2016/031925

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2009141178 A1	04-06-2009	CN 101816037 A	25-08-2010
		EP 2227798 A1	15-09-2010
		JP 5432889 B2	05-03-2014
		JP 2010536052 A	25-11-2010
		RU 2010105853 A	27-08-2011
		US 2009141178 A1	04-06-2009
		WO 2009069810 A1	04-06-2009
WO 2015017314 A1	05-02-2015	CN 105409203 A	16-03-2016
		EP 3028446 A1	08-06-2016
		JP 2016528826 A	15-09-2016
		KR 20160027975 A	10-03-2016
		TW 201521435 A	01-06-2015
		US 2016254028 A1	01-09-2016
		WO 2015017314 A1	05-02-2015
US 2014078165 A1	20-03-2014	CN 103563347 A	05-02-2014
		CN 103873738 A	18-06-2014
		CN 105933573 A	07-09-2016
		EP 2716028 A2	09-04-2014
		JP 5833233 B2	16-12-2015
		JP 2014523661 A	11-09-2014
		JP 2016054496 A	14-04-2016
		KR 20140017645 A	11-02-2014
		KR 20140035517 A	21-03-2014
		KR 20140110071 A	16-09-2014
		KR 20160054039 A	13-05-2016
		US 2014078165 A1	20-03-2014
		US 2015281707 A1	01-10-2015
		WO 2012166382 A2	06-12-2012
WO 2014176019 A1	30-10-2014	CN 105144729 A	09-12-2015
		EP 2989798 A1	02-03-2016
		HK 1214051 A1	15-07-2016
		US 2016105695 A1	14-04-2016
		WO 2014176019 A1	30-10-2014
WO 2013130478 A1	06-09-2013	CN 104126307 A	29-10-2014
		EP 2820848 A1	07-01-2015
		JP 2015517233 A	18-06-2015
		US 2015007243 A1	01-01-2015
		WO 2013130478 A1	06-09-2013
US 6683657 B1	27-01-2004	JP 2001100699 A	13-04-2001
		US 6683657 B1	27-01-2004
US 2007201746 A1	30-08-2007	AU 2002306116 A1	02-12-2003
		EP 1509882 A1	02-03-2005
		US 2007201746 A1	30-08-2007
		WO 03098549 A1	27-11-2003

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-8, 15-17

   two or more asymmetric alpha values

   ---

2. claims: 9-12

   a scene-cut determination using histograms and weighted Euclidian distances or using an EMA filter

   ---

3. claims: 13, 14

   a look-ahead sliding window
