

System for dynamically maximizing the contrast between the foreground and background in images and/or image sequences

The invention relates to a system for dynamically maximizing the contrast between foreground and background in images and/or image sequences.

For the smooth integration of virtual objects into real scenes (also called augmented reality), it is known that there are two problems to solve. On one hand, the transformation between camera coordinate system and scene coordinate system must be determined and, on the other hand, occlusions of virtual objects, which are usually displayed in the background, must be determined by real objects, which usually are in the foreground.

In the application area "television transmission" or "broadcasting", one is interested in solving the two aforementioned problems preferably non-mechanically, i.e. without the additional use of markers or sensors. This allows for a shorter assembly time and reduces the complexity of a broadcast system on site to be assembled as a whole. Especially with live broadcasts, the majority of the equipment must be transported to the usage site and subsequently assembled and disassembled.

To solve the transformation problem, there are, on one hand, hardware-based solutions such as "encoded heads" (e.g. from vinten.com), and, on the other hand, software-based, visually working solutions. Both solutions provide a mapping $x=M(X)$ which allows to map a 3D coordinate X of a virtual object in the 3D space in a camera coordinate x in the current camera image.

Previously known segmentation methods require so-called keying instructions, such as a specific background color (e.g., in the case of chroma keying or infrared keying). However, these mechanical solutions have the above-mentioned disadvantages, since the keying instructions must be installed at the transmission site.

Object of the invention is to avoid the above disadvantages.

To solve this object, a system for dynamically maximizing the contrast between

foreground and background in images and/or image sequences is suggested, comprising

at least one image-recording device,

at least one analysis module connected to the image-recording device, which is configured to determine a transformation between a two-dimensional coordinate system of the image-recording device and a two-dimensional coordinate system of a display device, and carry out a segmentation in order to determine at least occlusions of background objects through foreground objects in an image,

at least one control module connected to the analysis module,

at least one display device, which is configured to display display information, in particular, video clips, wherein the image-recording device is configured to record the display information displayed on the display device,

wherein the analysis module is further configured to identify quality data on the segmentation and transmit same to the control module, and

wherein the control module is configured to adjust the display information displayed on the display device based on the quality data.

By taking into account quality data on the segmentation, it is possible to obtain information about whether foreground and background are in sufficient contrast to each other. In case of a non-sufficient contrast, the display information can be adjusted, so that in the case of an unchanged foreground, a high-contrast background can be generated.

The system can further include a database, which is connected to the control module and to the display device, wherein the database is configured to store a sequence of display information to be displayed on the display device one after the other. The database can, for example, be part of a media server. The media server transmits display information to be played, for example, video clips, to the display device.

In this connection, it is preferred that the control module is configured to change the stored sequence of the display information. In other words, the control module can instruct the media server to output or display a particular piece of display information on the display device. To achieve optimum control by means of the control module, the display information output from the media server on the display device (video clip)

can also be transmitted to the control module.

The analysis module can be configured to apply and minimize an energy function during the segmentation, wherein the energy function is based on probability distributions of foreground objects and background objects and/or on pixel-wise probability distributions. The segmentation into foreground and background (with intermediate values such as semi-occlusions or blurring being possible) is designated with Y .

Here segmentation methods are used that do not require keying instructions, but can separate arbitrary background and foreground objects. In this process, an energy function $E(Y)$ is defined, which is to be minimized. $E(Y)$ can correspond to the negative logarithm of the probability that an occlusion has been determined correctly. Methods of convex optimization permit to minimize the energy efficiently and optimally, taking into account each individual image point, i.e. to maximize the corresponding probability. However, for minimizing the energy function $E(Y)$, a priori knowledge of the appearance of the background and of the foreground is required, in particular, in the form of probability distributions F (foreground) and B (background) or pixel-wise distributions f_i and b_i , wherein i designates a specific image point/pixel.

This a priori knowledge needs to be updated due to possible changes in brightness or changes of the appearance, which causes drifting (the progressive deviation of a model from reality). Due to the above-described feedback mechanism, in which quality data on the segmentation are transmitted to the control module, and the control module adjusts the display information based on the quality data, the drift problem can be counteracted. This allows therefore the use of current segmentation methods with enhanced performance. The feedback takes, in particular, place via the display device, preferably an LED system as it, for example, exists for banner advertising in many sporting events.

The control module can be further configured to determine for remaining display information of the sequence, in which display information an expected probability distribution of the foreground and an expected probability distribution of the

background of the image-recording device differ the most.

The control module can also be configured to calculate a probability mapping for a respective image-recording device, wherein the probability mapping describes which probability distribution exists in an image of the respective image-recording device if the display device displays any one of the remaining pieces of display information.

The control module can be further configured to determine the probability mapping based on the time-varying 2D-2D mapping $x=F(d)$ between d in the coordinate system of the display device and x in the coordinate system of the image-recording device. This 2D-2D mapping F can, for example, be determined by applying $x=M(X)$, wherein X is the physical position X of the display point d .

The control module can be configured to determine the probability mapping based on a radiometric calibration between the display device and the respective image-recording device.

On the display device a plurality of arbitrarily designed calibration marks can be arranged, which are detectable by the image-recording device, in order to facilitate the 2D-2D transformation F between the coordinate system of the display device and the coordinate system of the image-recording device. In order that such calibration marks need not be regularly assembled and disassembled, the display devices can have such calibration marks permanently. The calibration marks can also be designed in such a manner that a radiometric calibration between recording device and display is possible.

The control module can be further configured to insert calibration information into the sequence of display information, wherein the display device is configured to display the calibration information. In this process, the calibration information can comprise at least a single monochrome image, wherein the image has a display duration of at least one or precisely one frame, and wherein the image preferably has a color, which has a high contrast to a foreground object. If a piece of calibration information is displayed as a monochrome image for the duration of a frame on the display device only, the calibration information is not visible to a viewer of the display device

with the naked eye. However, the calibration information displayed on the display device is recorded by the image-recording device, so that the respective frame can be used to update knowledge of foreground objects and background objects, as it is known for this frame, which color the background has.

The aforementioned problem is also solved by a method for dynamically maximizing the contrast between foreground and background in images and/or image sequences, comprising the steps of:

detecting at least one image by means of at least one image-recording device,
calculating a transformation between a two-dimensional coordinate system of the image-recording device and a two-dimensional coordinate system of the display device,
carrying out a segmentation in order to determine at least occlusions of background objects through foreground objects in an image,
displaying display information, in particular, video clips, on a display device, wherein the display information displayed on the display device is recorded by the image-recording device,
wherein, during segmentation, quality data on the segmentation are identified, and the display information displayed on the display device are adjusted based on the quality data.

This method can comprise further optional steps corresponding to the features described above for the system. Accordingly, further optional aspects of the method can be directly derived from these features.

In the following text, the invention is exemplarily and non-restrictingly described with reference to the accompanying figures.

Fig. 1 shows a schematic diagram of the system presented herein.

A system 10 for dynamically maximizing the contrast comprises a plurality of image-recording devices C1, C2, Ck (indicated by three black dots below C2) and a plurality of analysis modules A1, A2 to Ak (indicated by three black dots below A2). An analysis module A1, A2 is assigned to each camera C1, C2.

The cameras C_1, C_2, C_k pass on appropriate data S_{1k} to the respective analysis module A_1, A_2, A_k . The analysis modules A_1, A_2, A_k determine (1) the current 2D-2D mapping F_1, F_2, F_k between a 2D coordinate in the display coordinate system of L in the image coordinate system of the respective camera C_1, C_2, C_k and (2) image point-wise occlusion. The data S_{1k} are, depending on the analysis module, purely optical, but can also include sensor data from step sensors, acceleration sensors, etc. The mapping F_k may have been calibrated manually or with an automatic calibration procedure.

The output data S_{2k} of the analysis modules A_1, A_2, A_k describe the specific mappings F_k as well as meta data on the occlusion. These output data are transmitted to a control module K .

Meta information to enhance the segmentation flows in the other direction (indicated by the arrows TCP) from the control module K to the analysis modules A_k .

Furthermore, the control module K is connected, via a network, to a database, D , in which display information (video clip) is stored, which can be displayed on the LED system L . The control module has, in particular, control access over the network on a media server with a video clip database D . The media server sends video clips to be played or display information Z to the LED system L .

The control module K is configured to instruct the media server to output certain display information or video clips Z , via an image transport protocol BP, e.g. via HDMI, SDI, DisplayPort or SDI over TCP, to the display device L . For control, the currently output video image is returned again to K via BP.

In particular, the setup should be noted with which the control module K selects the video material (display information) that is to be displayed on the display device L , as well as the control loop with feedback to the camera C_1, C_2 . Usually, the video material (video clips) is organized a priori in a playlist, which is played during a broadcast of an event, such as a match in a specific sport, without the possibility of adaptation.

In the present case, the control module K is configured to dynamically adjust a sequence R of remaining display information or video clips Z, in order to obtain preferably good segmentation results.

For this purpose, a metric on the quality (quality data) of the segmentation is transmitted in the metadata S2k, for example, the segmentation energy $E_k(L)$ resulting after minimization. If this energy exceeds an acceptable level in one or more cameras C1, C2, Ck, the video clip order of the remaining sequence of video clips is dynamically adjusted. As already mentioned in the introduction, segmentation methods are used that do not require keying instructions, but can separate arbitrary background and foreground objects. In this process, an energy function $E(Y)$ is defined, which is to be minimized. $E(Y)$ can correspond to the negative logarithm of the probability that an occlusion has been determined correctly. Methods of convex optimization permit to minimize the energy efficiently and optimally, taking into account each individual image point, i.e. to maximize the corresponding probability. However, for minimizing the energy function $E(Y)$ a priori knowledge of the appearance of the background and of the foreground is required, in particular, in the form of probability distributions F (foreground) and B (background) or pixel-wise distributions f_i and b_i .

In the following equation, 1 designates the analysis module, where the remaining energy $E_k(Y_k)$ is the greatest, i.e. the segmentation is most likely not correct or most uncertain. This is an indication of a poor contrast between foreground and background.

$$1 = \arg_k \max E_k(Y_k)$$

Accordingly, the image $G1(Z)$ is, for example, determined for the selected camera C1, which describes which probability distribution (at which pixel) is present in the camera image of C1, when the LED system L plays the display information or the video clip Z. $G1$ considers the 2D-2D mapping $F1$ from the LED system L to the camera C1 and background information of a radiometric calibration from the LED system L to the camera C1.

For every potential video clip Z_m (display information) remaining in the sequence, the mapping $G_1(Z_m)$ is determined. That clip Z_n is played next where the probability distribution F_1 of the foreground of camera C_1 and the expected probability distribution G_1 of the background of camera C_1 differ the most:

$$n = \arg_m \max d(F_1, G_1(Z_m))$$

A metric d for the comparison of foreground and background distribution is, for example, the Kullback-Leibler divergence.

The system 10 described herein and associated method steps allow to periodically, i.e. when the remaining energy E_k increases due to inaccurate knowledge, update knowledge of foreground and background appearance in each camera image and thus actively prevent drifting (the progressive deviation of a model from reality).

The display device L described herein can be an LED system, for example, in the form of advertising banners. LED systems L internally have a control box, which receives appropriate prepared video signals via an image transport protocol, e.g. HDMI or the like. Furthermore, LED systems have a plurality of so-called cabinets, which can each constitute a small part of the entire image and are connected to the control box. In this process, the control box passes on an image section of the original video to each cabinet. For simplicity, it can be assumed that an LED system corresponds to a regular screen, except that the structure of an LED system can be significantly more flexibly designed, e.g., that no rectangular layout must be selected.

It should be noted that the system 10 described herein can also be part of a superordinate television broadcast system, as described in the application entitled "Television broadcast system for generating augmented images" filed by the same applicant on the same day. The control module described herein can be contained in the system described therein in the control module 18. The display device described herein can be contained in the LED system 20 described therein.

Claims

1. A system for dynamically maximizing the contrast between foreground and background in images and/or image sequences, comprising
 - at least one image-recording device (C1, C2),
 - at least one analysis module (A1, A2) connected to the image-recording device (C1, C2), which is configured to determine a transformation between a two-dimensional coordinate system of the image-recording device (C1, C2) and a two-dimensional coordinate system of a display device (L), and carry out a segmentation in order to determine at least occlusions of background objects through foreground objects in an image,
 - at least one control module (K) connected to the analysis module (A1, A2),
 - at least one display device (L), which is configured to display display information, in particular, video clips, wherein the image-recording device (C1, C2) is configured to record the display information displayed on the display device (L),
 - wherein the analysis module (A1, A2) is further configured to identify quality data on the segmentation and transmit same to the control module (K), and
 - wherein the control module (K) is configured to adjust the display information displayed on the display device (L) based on the quality data.
2. The system according to claim 1, characterized in that it further includes a database (D), which is connected to the control module (K) and to the display device (L), wherein the database (D) is configured to store a sequence of display information to be displayed on the display device (L) one after the other.
3. The system according to claim 2, characterized in that the control module (K) is configured to change the stored sequence of the display information.
4. The system according to any one of the preceding claims, characterized in that the analysis module (A1, A2) is configured to apply and minimize an energy function during segmentation, wherein the energy function is based on probability distributions of foreground objects and background objects and/or on pixel-wise probability distributions.

5. The system according to claims 2 and 4, characterized in that the control module (K) is configured to determine for remaining display information of the sequence, in which display information an expected probability distribution of the foreground and an expected probability distribution of the background of the image-recording device differ the most.
6. The system according to claim 5, characterized in that the control module (K) is configured to calculate a probability mapping for a respective image-recording device (C1, C2), wherein the probability mapping describes which probability distribution exists in an image of the respective image-recording device (C1, C2) if the display device (L) displays any one of the remaining pieces of display information.
7. The system according to claim 6, characterized in that the control module (K) is configured to determine the probability mapping based on a transformation between the coordinate system of the display device and the coordinate system of the image-recording device (C1, C2).
8. The system according to claim 7, characterized in that the control module (K) is configured to determine the probability mapping based on a radiometric calibration between the display device (L) and the respective image-recording device (C1, C2).
9. The system according to claim 7 or 8, characterized in that a plurality of calibration marks are arranged on the display device (L), which are detectable by the image-recording device (C1, C2) in order to facilitate the transformation between the coordinate system of the display device (L) and the coordinate system of the image-recording device (C1, C2).
10. The system according to any one of claims 5 to 9, characterized in that the control module is configured to insert calibration information into the sequence of display information, wherein the display device is configured to display the calibration information.
11. The system according to claim 10, characterized in that the calibration information comprises at least a single monochrome image, wherein the image has a

display duration of at least one or precisely one frame, and wherein the image preferably has a color, which has a high contrast to a foreground object.

12. A method for dynamically maximizing the contrast between the foreground and background in images and/or image sequences, comprising the steps of:

- detecting at least one image by means of at least one image-recording device,
- calculating a transformation between a two-dimensional coordinate system of the image-recording device and a two-dimensional coordinate system of a display device,

- carrying out a segmentation in order to determine at least occlusions of background objects through foreground objects in an image,

- displaying display information, in particular, video clips, on a display device, wherein the display information displayed on the display device is recorded by the image-recording device,

- wherein, during segmentation, quality data on the segmentation are identified, and the display information displayed on the display device are adjusted based on the quality data.