August 06, 2019

Multi-Input 3D Photos

Anonymous Anonymous

Multi-Input 3D Photos

Abstract

In a section of the present disclosure, a system for generating a 3D image from a 2D image and a depth map of the 2D image is disclosed. A server receives the 2D image from a client device. A minimum depth value is set for pixels at bottom of the 2D image. An object detection algorithm identifies presence of an object in the 2D image. A maximum depth value is set at a position of the object, if it is detected; otherwise, it is set at top of the 2D image. A continuous depth function varying from the minimum depth value to the maximum depth value is used to represent the depth map. A stereoscopic image is then generated utilizing the depth map of the 2D image. Thereafter, a parallax disparity in the stereoscopic image is generated to produce two views in the stereoscopic image. Finally, the two views in the stereoscopic image are used to generate the 3D image.

In another section of the present disclosure, a system combines at least two input images (at least one is a 3D image or has a depth map) into a new combined 3D image. A server receives the input images (for example, a first RGBD image, a second RGBD image and an RGB image) from a client device. A first depth map is extracted from the first RGBD image and a second depth map is extracted from the second RGBD image. Thereafter, a depth map is also generated for the RGB image, referred to as a third depth map. Finally, a combined depth map is generated from the first depth map, the second depth map and the third depth map. In a next step, weight masks are defined for the input images. Each of the weight masks represents a set of coefficients. The set of coefficients are calculated by performing a histogram alignment and by applying a mismatch bias based on the histogram alignment. The weight masks are applied to pixels in each of the input images by multiplying the set of coefficients with values of the pixels at each pixel location. Then the input images with the applied weight masks are combined to produce a combined RGB image. Thereupon, the server generates the new combined 3D image from the combined RGB image and the combined depth map.

Problem statement

It might not be adequate to construct a 3D scene with just a single RGBD image or a single 3D image of an original scene. In a constructed 3D scene, there might be some parts which are occluded due to lesser information. This might also lead to artifacts in the constructed 3D scene. The artifacts are features which appear in the constructed 3D scene but are not present in the original scene.

The present disclosure proposes a novel solution to overcome these challenges.

System and working

The present disclosure describes a system to perform following first and second functionalities:

(i) Generating a 3D image from a 2D image and a depth map of the 2D image

(ii) Combining at least two input images (at least one is a 3D image or has a depth map) to generate a new combined 3D image

Following section describes the first functionality performed by the present disclosure. The system deploys a client-server model for performing functionalities described herein.

(i) A server receives the 2D image from a client device and a continuous depth function is utilized for generating the depth map of the received 2D image. A scene at top of the 2D image is usually farther away from a viewer than a scene at bottom of the 2D image. The continuous depth function assumes a small or a zero-depth value (i.e. a minimum depth value) at the bottom of the 2D image and a large (negative) depth value (i.e. a maximum depth value) at the top of the 2D image. The continuous depth function may be a piece-wise linear or a non-linear function. Additionally, the continuous depth function may not be a continuously increasing function from the bottom of the 2D image to the top of the 2D image. The 2D image may include objects such as buildings, walls, trees, *etc*. Value of depth remains fixed on the objects present in the 2D image. The value of the depth results in generation of a knee point in the continuous depth function. The knee point in the continuous depth function is detected based on a position of the object in the 2D image as determined by an object detection algorithm. The object detection algorithm identifies presence of the object(s) in the 2D image. If no objects are detected, the maximum depth value is set for pixels at the top of the 2D image. If the object is detected, the maximum depth value is set at the position of the object. The continuous depth function varying from the minimum depth value to the maximum depth value is used to represent the depth map. Hence, the continuous depth function is the piece-wise linear function (having the knee point), if the object is detected (as shown in Figure 1); otherwise, it is the non-linear function.

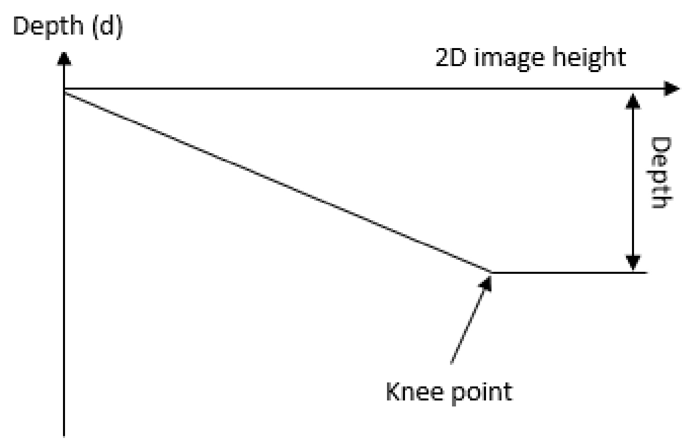

Figure 1: The piece-wise linear function represents the depth map of the 2D image

Then a stereoscopic image is generated utilizing the depth map of the 2D image. Thereafter, a parallax disparity (D) in the stereoscopic image is calculated using following equation:

$$D = tx * d/(S - d)$$

where d = object depth,

tx = a distance between a right eye and a left eye,

and S = a distance between eyes and a display screen

The parallax disparity makes the stereoscopic image appear 3D. The parallax disparity is calculated in order to produce two views in the stereoscopic image. The two views in the stereoscopic image are finally used to generate the 3D image.

Following section describes the second functionality performed by the present disclosure.

(ii) A server receives at least two input images from a client device (for example, a first RGBD image, a second RGBD image and an RGB image). The first RGBD image and the second RGBD image and the RGB image are three different images of an original scene. A first depth map is extracted from the first RGBD image and a second depth map is extracted from the second RGBD image. The server includes an analysis engine for constructing a combined depth map.

The input images, the first depth map and the second depth map are given as input to the analysis engine.

Constructing the combined depth map

The analysis engine determines pixel values of the first RGBD image and pixel values of the second RGBD image that are occluded. Depth values of the second depth map are used instead of depth values of the first depth map corresponding to the pixel values of the first RGBD image that are occluded to construct a combined depth map. Similarly, the depth values of the first depth map are used instead of the depth values of the second depth map corresponding to the pixel values of the second RGBD image that are occluded to construct the combined depth map. Then, for the pixel values of the first RGBD image that are not present in the second RGBD image, the depth values in the first depth map are used to construct the combined depth map. Similarly, for the pixel values of the second RGBD image that are not present in the first RGBD image, the depth values in the second depth map are used to construct the combined depth map. After choosing the depth values corresponding to the pixel values that are mutually exclusive, the analysis engine selects the depth values corresponding to pixel values that are common in the first RGBD image and the second RGBD image. Then the analysis engine compares the first depth map with the second depth map at the pixel values that are common to both the first RGBD image and the second RGBD image. For the pixel values that are common, the depth values of the first depth map are used to construct the combined depth map, if these are closer to depth values in neighborhood than the depth values of the second depth map. This ensures depth consistency over neighboring views. Thereafter, a depth map for the RGB image is constructed (referred to as a third depth map) as explained in section (i) previously. For pixel values of the RGB image that are not present in both the first RGBD image and the second RGBD image, depth values in the third depth map are used to construct the combined depth map. This way, the combined depth map is generated.

Constructing a combined RGB image

One of the input images is selected as a reference image. The input image with best exposure or best color intensities may be chosen as the reference image. After this, weight masks are defined for the input images. For example, a weight mask 1, a weight mask 2 and a weight mask 3 are defined for the first RGBD image, the second RGBD image and the RGB image respectively. Each of the weight masks represents a set of coefficients to be applied to pixels within at least one of the input images. The set of coefficients for each of the weight masks is calculated by performing a histogram alignment between the reference image and each of the input images other than the reference image. In the histogram alignment, pixel values of the reference image are compared to pixel values of each of the input images other than the reference image. Then, a mismatch bias is calculated, if the input images are not well-matched based on the histogram alignment. The mismatch bias is calculated based on how much the input images differ (e.g., proportional to a mismatch). Thereafter, the mismatch bias is calculated with respect to the reference image. The mismatch bias is applied to the set of coefficients in each of the weight masks, eliminating regions with local misalignment in the input images. Then the weight masks are applied to the pixels in each of the input images by multiplying the set of coefficients with the pixel values at each pixel location. Finally, the input images with the applied weight masks are combined to produce a combined RGB image. Thereupon, the server generates the new combined 3D image from the combined RGB image, and the combined depth map as explained in the section (i) previously.

Additional embodiments

In an additional embodiment, the system may be utilized in eliminating inpainting artifacts. Inpainting is a process of reconstructing lost or deteriorated parts of an image. Several features may be introduced in the image which are not present in an original image due to inpainting. These are known as inpainting artifacts. As the present disclosure utilizes multiple input images, this allows a user to provide more detail/information at different depths. This helps in better reconstruction of the image with less artifacts.

In a yet another embodiment, the user may specify bitmap masks to only use certain regions of the input images.

Conclusion

In today's world, there is an enormous demand for 3D content in applications such as computer graphics, virtual reality and communication. With the advent in technology, researchers have come up with ways to bridge this gap. Numerous models are being developed for generating the 3D content as per the requirements. But there have been several challenges such as deterioration, loss of information, *etc.,* while constructing 3D scenes. With the solution provided in the present disclosure, we can overcome these challenges to quite an extent. Using multiple RGBD images of an original scene provide additional information at different depths to construct the 3D scene in a better way. This additional information also ensures that the artifacts in the 3D scene are minimized to a great extent.