Journal of Engineering and Applied Sciences 13 (Special Issue 5): 4699-4704, 2018
ISSN: 1816-949X

Inserting Virtual Movie Object into Video with Illumination Consistency

Israa Hadi Ali and Roa'a M. Al-Airaji
Information Technology College, Babylon University, 51001 Babylon, 00964 Iraq

Abstract: This study presents a novel approach to integrating virtual objects into a video of real objects with illumination consistency to enrich a video environment with virtual information. In order to achieve consistent composite results and make virtual and real objects appear as it coexists in the same environment, three key issues in achieving mixed reality need to be processed are geometry, illumination and time. This study focuses on the second issue and proposes several techniques for interactively projecting shadows of the "real video onto the virtual objects. The proposed system can be dividing into three phase: moving object detection phase, tracking phase and mixing phase. Moving object detection phase, include background subtraction and segmentation technique. Tracking phase include features extracting such as edge, area and center points to each existing object and then tracking these objects. Thus, the object-tracking technique uses center points of the certain object from each frame of a movie file for extracting trajectory. Mixing phase, include merge virtual objects back to the real video scene and illumination consistency to match shade and shadow in both videos to achieve a natural merge.

Key words: Mixed reality, background model, virtual and real object, video editing, HSV color space, video tracking

INTRODUCTION

Mixed reality refers to the merging of real and virtual environments and produces new environments where real and virtual objects co-exist and interact normally in real time (Zhang *et al*., 2011). Mixed reality includes both augmented reality and augmented virtuality where augmented reality is technology that superimposes and displays computer-generated worlds in the real world and augmented virtuality is technology that enhances the virtual world using real world data (Naemura *et al*., 2002; Bui *et al*., 2016).

In recent years, mixed reality have been widely studied and attract a great deal of attention in media industries such as film making, image/video editing, television broadcasting, advertising, virtual reality, production of images and animated sequences for cinema (Sato *et al*., 2005; Da Silva *et al*., 2013). However, the seamless integration of virtual objects into real video scenes remains a challenging problem in computer graphics. Comparing to embedding objects into images, three main challenges exist in inserting virtual objects into video are Abad *et al*. (2003) and Park *et al*. (2005):

- Consistency of geometry, virtual static objects must be located at a desired and correct location in the video scene (Idris *et al*., 2010)
- Consistency of illumination, the virtual object must cast a correct shadow and match to that of other objects in the real video scene
- Consistency of time, moving of virtual and video objects has to be correctly coordinated

In the mixed reality, achieving illumination consistency is a difficult task because of the lighting conditions in the real environment of the live video are usually unknown. In outdoor scenes, the sunlight is the only direct light source. As the orbit of the Sun is fixed and its position is deterministic at any time during a day, the main reason to calculate sunlight easily is light conditions. However, during cloudy days the skylight plays an important role for object illumination. Moreover, cloud motion causes changing of lighting conditions dramatically (Liu *et al*., 2009). There have been several approaches to estimate light sources including physical model-based approaches, measurement-based approaches and image estimation-based approaches. In addition, image-based relighting and inverse rendering also address a similar problem (Jacobs and Loscos, 2006).

Literature review: Rehg *et al*. (2000) presented method to video editing they segment video into multiple layers are foreground and background then segment and tracking foreground video. Then composition of the layers based on image-based rendering.

Corresponding Author: Israa Hadi Ali, Information Technology College, Babylon University, 51001 Babylon, 00964 Iraq

Naemura *et al*. (2002) proposed concepts virtual shadows in mixed reality environment using flashlight-like devices. The divide the concept of virtual shadow into four categories are real to virtual shadow for rigid objects, real to virtual shadow for non-rigid objects, image-based virtual to virtual shadow and virtual to real shadow.

Liu *et al*. (2009) estimation light source of outdoor scenes where proposed method for evaluating the light conditions of a static outdoor scene without knowing its geometry, material or texture.

Ren *et al*. (2013) insert virtual pedestrians into a pedestrian real video. The researcher's studies technique to mix virtual pedestrians and video of real pedestrian groups. They proposed an approach to efficient dynamic path planning consistent to the motion path of pedestrians in real video then the estimate distribution of pedestrians, lastly insert virtual characters into real video after processing every frame in video.

Israa and Adil (2016) develop idea of accumulation histogram method to background estimation and detect moving objects trajectories in video. the researchers development idea of accumulation histogram by divide each frame in video to blocks with the same size and find a mean value for each block then using the mean value instead of image pixels to find background they found this method efficiency in terms of time and gives accurate results. After that, they apply segmentation and tracking to find the trajectory of foreground objects.

Sakkari *et al*. (2012) used hand as support to insert virtual object in augmented reality. The researcher based on the real component of the scene without introducing a marker, he used hand using hand as support to insert virtual object. The researcher first extracts skin regions then identify and detect the hand by neuron network.

MATERIALS AND METHODS

The proposed system can be dividing into three parts: moving object detection, objects tracking and mixing videos as displayed in Fig. 1. The proposed method consists of many steps as follows:

Input real and virtual videos: In this step, we open AVI structure for each of the real and virtual video file and get list movie (still images/sequence frames) then get information such as total of frame, width of the frame and height of the frame from AVI header then split the video into frames (still image).

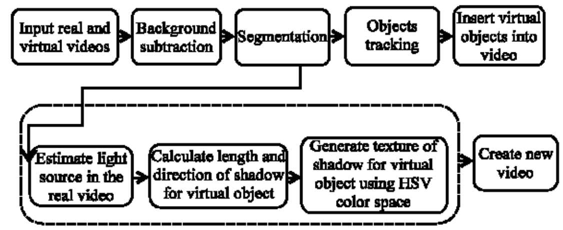

Fig. 1: Block diagram of the proposed method

Background subtraction: Background subtraction used to extract the motion part from a video stream. It handles separation of moving objects from the background scene. There are two important phases in a background subtraction algorithm are background model and foreground detection

Background model: Background model is an efficient way to get the background of the video and obtain foreground objects. There are several methods to get background of the video this study improved method of accumulation histogram. The general color histogram shows a representation of colors distribution in image, it is the feature to find an occurring probability of different colors in the image. In general color histogram the x-axis represents a color value and y-axis represents color occurring frequency in the image while in accumulation histogram the x-axis represents color value and y-axis represents color accumulation occurring frequency in the image.

This study improved accumulation histogram by dividing image into blocks and calculate median of colors for each block. The improved accumulation histogram work as follow:

Step 1: Take a certain number of frames (specific period of the video) and divide each frame into K blocks, each with U?V size.

Step 2: For each frame, calculate a median value for the first block and save these values in a temporal matrix.

Step 3: Calculate frequent of the median values that saved in a temporal matrix and find the most frequent of the median value, the most frequent of median value represents a number of frames that contain almost same color values in the first block.

Step 4: After that, the first block in the background model is reconstructing where values of blocks with the most frequent median value represent values of the first block in the background model.

Step 5: Advance to the next block and repeat steps 2-4. These steps continue until reach to the last block.

Foreground detection: Foreground detection is an approach for identifying the moving objects in a video sequence from the background. The simplest method to perform foreground detection is to compare frames obtained at time t with the background obtained from the previous step as explained in Eq. 1.

$$|F_t(x)\text{-}B(x)|>T \tag{1}$$

Where:

x = The pixel position intensity
F_t = Represent current frame at time t
B = Background model
T = The threshold value

The pixel value is background pixel if the difference between the pixel in current frame at time t and corresponding background pixel equal zero otherwise it is foreground pixel. However, pixel value may be background the pixel if the difference slightly greater than zero because of lighting change, so threshold value is used and it is the small value.

After applying background subtraction method on all frames in the video the output is a new video with 0's value for background and 1's value for the foreground.

Segmentation: Applying segmentation of the first frame in the video in this study region-growing segmentation is used to segment each existing object. the region grown method start with a number of the seed points, then the regions from these seed points are growing by allocating those neighboring pixels that have similar features with the seed point to the respective region, the features that used to measure the similarity are texture, color, intensity and gray level.

Objects tracking: After detecting each object in the real and virtual video, the next step is tracking of these objects based on some features such as size and center point of the object. The size of the object represents a total number of pixels for that object. The size of the object can be computing using Eq. 2:

$$\text{Size}=\sum_x \sum_y O(x,y) \tag{2}$$

Where, O(x, y) refers to the pixel object. The center of the object can be computing using Eq. 3:

$$\begin{aligned} X_c &= \left(\sum_x \sum_y (xO(x,y))\right)/A \\ Y_c &= \left(\sum_x \sum_y (yO(x,y))\right)/A \end{aligned} \tag{3}$$

Where:

X_c and Y_c = The centroid coordinates of an object
A = Refers to the object's area

Insert the virtual object into the video: The virtual objects are inserting to the real video with the same motion tracking in the virtual video. A trajectory of the virtual object obtained from the object tracking phase is necessary to determine move track of object in the real video. The process of inserting is done frame by frame.

Illumination consistency: To enhance the quality of new video that produced from the mixed of virtual and video objects and makes virtual objects appear as it co-exists with the real objects. It is necessary to extract illumination distribution of the real video order to render the virtual objects with this illumination. Illumination consistency consists of three steps:

Step 1: Estimate of light source, this study proposed method to estimate light source and the proposed method first convert image to the HSV color space. After that searches the image for points with high V, the points with the small difference between locations of it are grouping to the single area; the distance between locations of points is determined by the Euclidean distance.

Step 2: Reduce the number of selected points, eliminate a group of points lying in a close neighborhood to each other and belong to the same area but the size of the area is so small, the size of the area is a constant value stated experimentally.

Step 3: Determining groups of points belonging to a single area source of light and from these points select one point with the largest V which represents a center of the light source.

Calculate length and direction of shadow: The length and direction of the shadow mainly depend on the position and distance of the lighting from the object. To calculate the length of shadow first locates the point on the surface (plane) directly beneath the center point of light that obtained from the previous step. From this mark, draw a vertical line between these two points. After that, calculate the height of the object. Then calculate the length of shadow using Eq. 4:

$$L=\frac{x}{x\text{-}y}*y \tag{4}$$

Where:
L = A length of the shadow
x = A distance between center point of the light source and point on the surface (length of a vertical line)
h = Height of the object
y = A distance between the object and the vertical line

The direction of shadow determines by calculating the angle between the height of the object and length of shadow using Eq. 5:

$$\alpha = \frac{h}{L} \quad (5)$$

After that, the object is rotating by determining angle and flipping by 180°.

Generate texture of the shadow: This study used HSV color space to generate texture of shadow, the proposed method first convert RGB color for area of shadow on the image to the Hue Saturation Value (HSV) color space. After that, reduce Saturation (S) and Value (V) by specified number and converted back to RGB color to use as the shadow color.

RESULTS AND DISCUSSION

Example bellow explains results of the proposed method. Figure 2 explains the real and the virtual video as input to the system.

Figure 3a explains samples of frames which obtained from the process of converting the real video into sequence of frames (still images). Figure 3b explains samples of frames which obtained from the process of converting virtual video into a sequence of frames (still images).

Figure 4a explains background of the real video after performing proposed background model algorithm on the real video. Figure 4b explains background of the real video after performing proposed background model algorithm on the virtual video.

After background model, the next step is foreground detection, using Eq. 1 to find foreground of the real and virtual movie. Figure 5a explains samples of the real video frames after foreground detection with the threshold value equal eight. Figure 5b explains samples of the virtual video frames after foreground detection with the threshold value equal five.

After that, performing region growing segmentation algorithms on the first foreground frame in each of the real and virtual video and calculate center point for each object. For calculating the center point of the object must

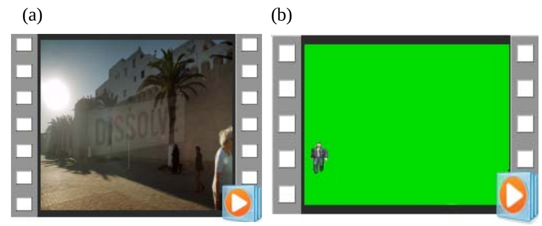

Fig. 2: a) Real video and b) Virtual video

Fig. 3: Samples of video frames: a) Real video and b) Virtual video

Fig. 4: a) Real video background and b) Virtual video background

calculate the size of an object. The equation 2 used to calculate size of an object. After that, the Eq. 3 used to calculate the location of an object. The center point of object that calculated in the first frame is saving and used as a seed to track object in the next frames through the movies. To tracking certain object, starting from the second frame and find the regions that represent object through the center which obtained from the object in the previous frame. Then calculate the new center for the object in current frame that using to find the objects in the next frame. In each time compute center must be saved because center points are useful and required to be used in the next step.

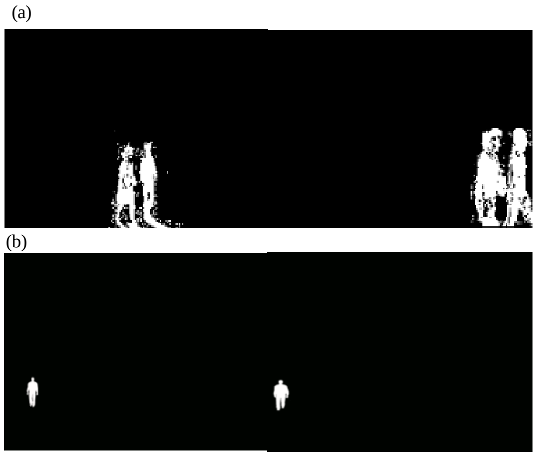

Fig. 5: Samples of video frames after foreground detection: a) Real video and b) Virtual video

Fig. 6: Samples of new video frames after inserting virtual object into real video

The result of object tracking algorithm is trajectory to each object in the real and virtual video. After object tracking, the next step is inserting virtual object into real video based on trajectories that obtained from the tracking process. Figure 6 explains samples of new video frames after inserting virtual object into real video.

To enhance the quality of new video that produced from the mixed of virtual and real objects and makes virtual objects appear as it co-exists with the real objects. It is necessary to extract illumination distribution of the real video order to render the virtual objects with this illumination. Illumination consistency consists of three steps: estimate of light source, calculate length and direction of shadow and generate texture of the shadow. Figure 7 explains center of the light source where the point with red color is center point of light source (sun) in real video.

Length of shadow calculates using Eq. 4 and direction of shadow calculate using Eq. 5 to calculate angle of direction then flipping the shadow by 180° to fall on the ground. Table 1 explains length and angle of direction of shadow for virtual object in two samples of frames in Fig. 6.

Fig. 7: Center point of light source

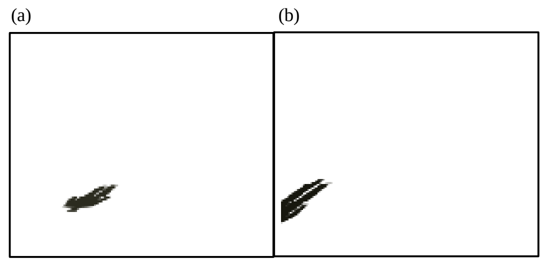

Fig. 8: Texture shadow of virtual object

Fig. 9: Samples of new video frames after texture shadow blending

Table 1: Length and angle of direction of shadow

Frame-number	Length of object	Length of shadow	Angle of direction
1	169	183	48°
2	174	202	45°

After calculate length and direction of shadow, the last step is generating texture of shadow by converting RGB color for area of shadow HSV color space. After that, reduce (S) by 25% and (V) by 15% then converted back to RGB color. Figure 8 and 9 explains texture shadow of virtual object after calculate length and direction of object shadow.

CONCLUSION

This study presented method for integrating virtual objects into the real video. The proposed method can achieve a high level of realism and using in any application that requires a realistic virtual object occlusion simulation such as design analysis, games and cinema. The cinema has been the driving force of this kind of application. The results demonstrate that the virtual object in the produced video appear as it is coexisting and interact normally with the real video objects in the real time.

REFERENCES

Abad, F., E. Camahort and R. Vivo, 2003. Integrating synthetic objects into real scenes. Comput. Graphics, 27: 5-17.

Bui, G., B. Morago, T. Le, K. Karsch and Z. Lu *et al*., 2016. Integrating videos with LIDAR scans for virtual reality. Proceedings of the 2016 IEEE Conference on Virtual Reality (VR), March 19-23, 2016, IEEE, Greenville, South Carolina, ISBN:978-1-5090-0836-0, pp: 161-162.

Da Silva, M.L.M., F.A. Caetano and R.L.D.S. Da Silva, 2013. A Semi-automatic approach to mix real and virtual objects for an augmented reality game. Proceedings of the SB Games 2013 Workshop on Virtual Augmented Reality and Games, October 16-18, 2013, Southern Baptist Convention, Sao Paulo, Brazil, pp: 1-4.

Idris, W.M.R.W., Y.M. Saman, A. Aziz and A.S.M. Noor, 2010. Managing virtual environments with tree-structured, Hierarchy-embedded virtual objects. Int. J. Comput. Sci. Network Secur., 10: 115-125.

Israa, H. and A.M. Adil, 2016. A method for background establishment using accumulate histogram for detection of object trajectory for video tracking applications. Q. Adjudicated J. Nat. Eng. Res. Stud., 3: 55-65.

Jacobs, K. and C. Loscos, 2006. Classification of illumination methods for mixed reality. Comput. Graphics Forum, 25: 29-51.

Liu, Y., X. Qin, S. Xu, E. Nakamae and Q. Peng, 2009. Light source estimation of outdoor scenes for mixed reality. Visual Comput., 25: 637-646.

Naemura, T., T. Nitta, A. Mimura and H. Harashima, 2002. Virtual shadows in mixed reality environment using flashlight-like devices. Trans. Virtual Reality Soc. JPN., 7: 227-237.

Park, J.S., M. Sung and S.R. Noh, 2005. Virtual object placement in video for augmented reality. Proceeding of the 6th Pacific Rim Conference on Advances in Multimedia Information Processing 2005, November 13-16, 2005, Springer, Jeju Island, Korea, pp: 13-24.

Rehg, J.M., S.B. Kang and T.J. Cham, 2000. Video editing using figure tracking and image-based rendering. Proceedings of the 2000 International Conference on Image Processing Vol. 2, September 10-13, 2000, IEEE, Vancouver, British Columbia, Canada, ISBN:0-7803-6297-7, pp: 17-20.

Ren, Z., W. Gai, F. Zhong, J. Pettre and Q. Peng, 2013. Inserting virtual pedestrians into pedestrian groups video with behavior consistency. Visual Comput., 29: 927-936.

Sakkari, M., M. Zaied and C.B. Amar, 2012. Using hand as support to insert virtual object in augmented reality applications. J. Data Process., 2: 10-24.

Sato, I., M. Hayashida, F. Kai, Y. Sato and K. Ikeuchi, 2005. Fast image synthesis of virtual objects in a real scene with natural shadings. Syst. Comput. Jpn., 36: 102-111.

Zhang, Y., J. Pettre, J. Ondrej, X. Qin and Q. Peng *et al*., 2011. Online inserting virtual characters into dynamic video scenes. Comput. Anim. Virtual Worlds, 22: 499-510.