Cut off Your Arm: A Medium-Cost System for Integrating a 3D Object with a Real Actor

Mahmoud Afifi[1], Mostafa Korashy[1], Ebram K. William[1], Ali H. Ahmed[1], and Khaled F. Hussain[2]
[1]Department of Information Technology, Faculty of computers and information, Assiut University, Assiut, Egypt
[2]Department of Computer Science, Faculty of computers and information, Assiut University, Assiut, Egypt
Email: {m.afifi, Khussain}@aun.edu.eg, {m.korashy, ebramkw, ali.bekhit}@ fci.au.edu.eg

Abstract—In the film industry, many tricks have been employed using the integration of a 3D object with a real actor. Usually, attaching a 3D object with a real actor is a costly process because of the usage of an expensive motion capture system. This paper presents a system using a medium-cost motion capture system and a chroma-keying technique for generating a video footage of an actor with an integrated 3D object (e.g. amputated arm). The result of the proposed system shows the attaching process of different 3D objects with a real actor who is combined with a new background scene in the same viewpoint.

Index Terms—Video Compositing, Image Enhancement, Motion Capture System, Chroma-keying.

I. INTRODUCTION

The rapid pace of evolution in attaching 3D objects with a real actor has a variety of applications such as film industry and entertainment [1]. Many research techniques have been introduced to integrate a 3D object with a real actor. In order to attach a 3D object to a moving real object, the real object must be tracked for applying the same movement to the 3D part. Many efforts have been made in the academic scope to integrate a virtual object with a real one. In [2], 3D shoes and clothes are integrated with a real body, so customers can see the virtual shoes and clothes in the magic mirror before the payment process. Where in [3], a 3D dressed body is attached to a real head for generating virtual clothes for an actor. The actor's motion is captured using a Motion Capture System (MCS). The captured motion is applied to a 3D virtual body of the actor. In [4], Microsoft Kinect is used in the tracking process for the animation of a generated virtual hand. The tracking process is performed using the depth images that are generated from Microsoft Kinect. In [5], Image Metrics and the University of Southern California's Institute for Creative Technologies (USC ICT) generate a realistic animated digital face that is integrated with a real actress. The digital face is reconstructed using Light Stage system that uses 156 controllable white LED lights. The generated face is attached to the real actress using a manually guided match moving process. In [6], a 3D manipulable body is attached to a real actor using a given video of the actor and a database of 3D scanned human-bodies. The

database consists of 550 virtual objects. A motion capture scheme is used for fitting the appropriate 3D body to the real actor's body.

All the mentioned applications are in the academic scope. In the cinematic scope, many films and TV series use the tricks of the 3D object integration with the actors. The "Grey's Anatomy" and "The Walking Dead" television series contain many characters with either hand or neck is cut and compensated by 3D objects [7]. In "The Curious Case of Benjamin Button" movie, Benjamin's head is a 3D digital head that is attached to a real body [8]. In "Stolen" movie, a real actor's leg is replaced by a 3D metal one [9].

In the proposed system, a fabricated video is produced with the usage of a medium-cost Motion Capture System (MCS) for capturing the movement of the real actor in order to integrate a 3D object with the real actor.

The remainder of this paper is organized as follows: Section II presents the proposed system in details. The results are shown in section III. Finally, the paper is concluded in section IV.

II. THE PROPOSED TECHNIQUE

The proposed system uses a medium-cost Motion Capture System (MCS) and a chroma-keying technique for generating a video footage of an actor with an integrated 3D object (e.g. amputated arm). In this section, the proposed system is described in details. First of all, the actor is extracted from the original scene using layered background model for foreground extraction. The next step is cutting off the arm of the actor using a chorma-keying technique. In the next step, we introduce the overlapping problem which appears when the arm of the actor overlaps his body, so we complete the overlapped part using a proposed body completion technique. Next, the movement of the 3D object is aligned with the actor's movement. Finally, the attached 3D object with the real actor is combined with a new background scene in the same viewpoint, as shown in Fig. 1.

A. Foreground Extraction

The actor should be extracted from the original video in order to be placed in a new scene. The used MCS is restricted by a predetermined environment and cannot be used in the real scene of the movie. Many techniques
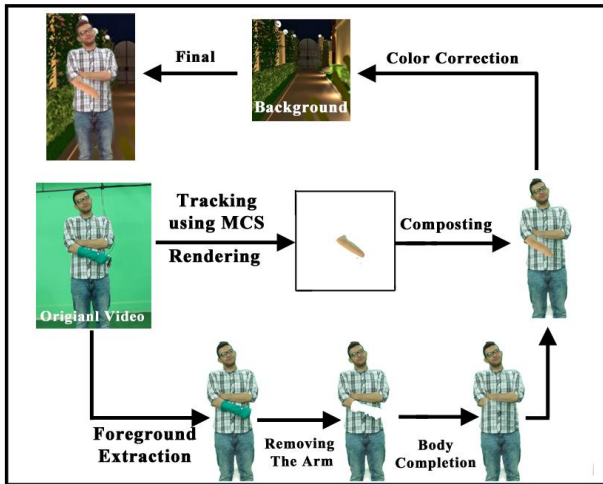
Fig. 1. Overview of the proposed technique.

have been developed to extract the foreground from still images and videos such as: background subtraction [10], Mixture of Gaussian (MOG) [11], and layered background model [12].

Even though, it is a simple and straightforward technique, background subtraction suffers from poor quality of results. It depends on the first frame, without any actor, as a background frame. This frame is subtracted from the incoming frames to obtain the actors. Lighting and shadow conditions are not the same in each frame as they were in the first one. This difference leads to poor results.

MOG model gives better results than frame differencing, because the background model changes in time to adapt the changes in the illumination conditions. The major limitation of MOG is the initialization of Gaussian and selecting values for the various parameters in the model. Although, the model adopts gradual changes in lighting conditions, any sudden and drastic changes are not adopted and produce inaccurate results.

We use the layered background model which produced the best results compared with the other techniques, as shown in Fig. 2. In the layered background model, a model containing five color values is constructed for each pixel. A pixel is considered a background pixel if it has a distance less than a specified threshold T. To obtain more accurate results, an additional step is added to the technique in which a color/brightness difference value D is computed from the following equation
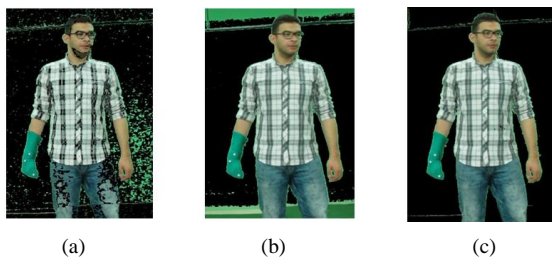

Fig. 2. The results of foreground extraction. (a) Background Subtraction. (b) Mixture of Gaussian. (c) Layered Background Model.
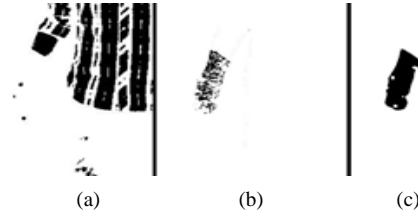

Fig. 3. A comparison between binary masks that are generated using (a) luma-key, (b) chroma-key, and (c) color difference technique using the inverted green channel

$$D = 18 \times D_{gb} + \left| \log\left(\frac{V_{BG}}{V_{Curr}} \right) \right| \qquad (1)$$

Where D_{gb} is the Euclidean metric between the intensity of the current pixel and the intensity of the most common background pixel, V_{BG} is the brightness in the background, and V_{Curr} is the brightness of the current pixel.

The color/brightness difference value D is used for detecting changes in shadow and lighting, so the pixels in the shadows and variable lighting regions are eliminated from appearing as a part of the foreground.

B. Removing the arm of the actor

After separating the actor from the original scene, the target arm of the actor is removed. For this purpose, the actor wears a green cloth that covers his arm. We remove the green arm from the body of the actor using a chorma-keying technique. Chroma keying techniques give the ability to remove a predefined color from the original footage, but green and blue are usually used because of their distinction from the skin color of most actors [10].

There are many chroma keying techniques that are used to generate a binary mask based on the predefined color [10]. The binary mask specifies the removed region in the original footage. Luma-key technique uses the luma version of the frame to generate the binary mask. Chroma-key technique uses the Euclidean distance to calculate the distance between any given pixel intensity and the predefined color intensity. Color difference technique calculates the difference between the green channel, in the case of green screens, and the maximum between the blue and red channel. A color difference technique based on the inverted channel generates the most suitable binary mask by generating an inverse of the channel of the predefined color, in our case it is the green channel. The green region of the inverted channel is dark. By scaling the inverted channel and improving it using the red channel value, the binary mask is generated. Now,
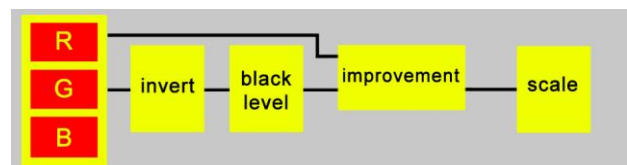

Fig. 4. Flow chart of the difference technique using the inverted channel of the predefined color.

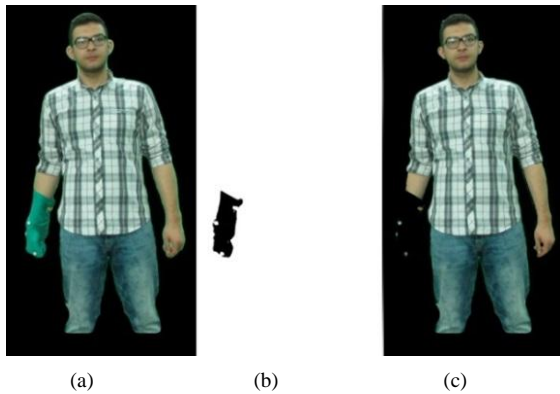(a)   (b)   (c)

Fig. 5. Removing the green arm. (a) The original frame. (b) The binary mask. (c) The result of removing the arm of the actor.

the green pixels are scaled down to zero, otherwise pixels are scaled up to one. The generated binary mask is the best result compared with the results of other techniques, see Fig. 3.

We use a color difference technique based on the inverted channel to remove the green arm of the actor. For each pixel in the current frame, we calculate a corresponding pixel in the binary mask as shown in Fig. 4. The generated binary mask is multiplied by the current original frame to generate a final frame without green pixels, see Fig. 5.

C. Body Completion

During the movement of the actor, his arm with the green cover may occlude a part of his body as shown in Fig. 6 (a). After removing the green arm of the actor, the occluded part of the body becomes empty as shown in Fig. 6 (b). Even though, a number of approaches have been developed in video completion, but most of those techniques either a time consuming or introducing a flickering in the video. So we use a simple and straightforward technique for solving this problem. First of all, we track the movement of the body, specifically the shirt, of the actor in the video using [13]. In the frame which has the arm overlapped with the shirt, we apply the tracking data to the shirt segment of another frame that contains the shirt without any overlapping. We use the technique which is proposed in [14] for blending the two frames as shown in Fig. 6 (c). The used technique is a simple one and saves more time compared with other
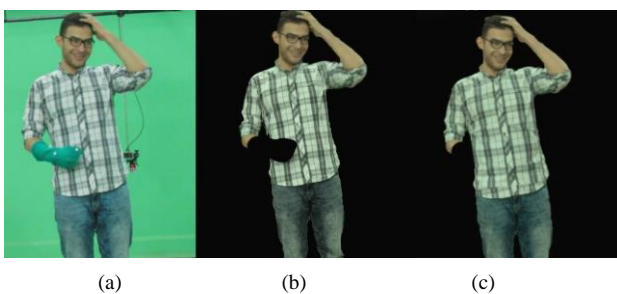

(a)   (b)   (c)

Fig. 6. The problem of the arm overlapping with the body. (a) The original footage. (b) After removing the green color. (c) After applying the completion technique.

(a)   (b)   (c)

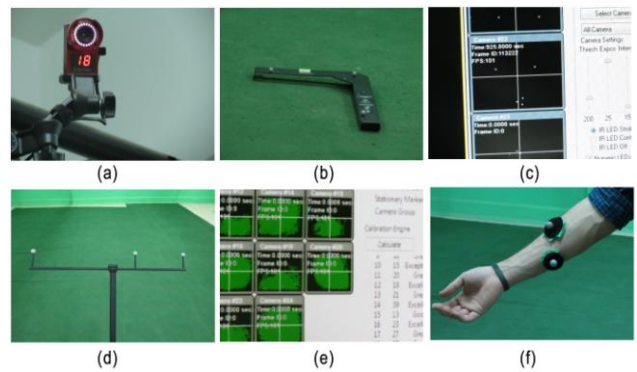(d)   (e)   (f)

Fig. 7. Configuration of the motion capture system.

techniques. In addition to the simplicity and time saving, the quality of the results is satisfactory for some extent.

D. Motion Capturing

In order to align the movement of the 3D model with the actor's movement, we need to know how the actor moves in the scene. The usage of a low-cost device such as Microsoft Kinect is at the expense of performance [15]. Accordingly, a MCS is used to track the actor's movement. MCSs fall into one of a three categories: (1) Magnetic, (2) Mechanical, and (3) Optical. The used MCS is an optical one that consists of 24 cameras. The cameras are aligned in two rows: top and bottom rows. Each row consists of 12 cameras, each with a frame rate of 100 FPS, see Fig. 7 (a). This construction gives the actor a $(12' \times 12' \times 8')$ as a capturing volume.

As the first step in using the MCS, we need to set up the system by means of the calibration square, see Fig. 7 (b), in which cameras view the same point of view as shown in Fig. 7 (c). Calibration of the system must be performed before proceeding into the capturing process. In this task, we use ARENA software package which is licensed by our system [16]. In combination with ARENA software, we used OptiWand Kit that is shown in Fig. 7 (d). In order to obtain more accurate calibration results, at least 1000 samples must be captured by each camera using OptiWand Kit, as shown in Fig. 7 (e).

After that, we start capturing the data in FBX format. The data is exported into 3Ds MAX software to align the movement of the 3D model with the actor's one, which is explained in the following subsection. The system is capable of tracking an overall human skeleton by wearing a special suit with markers in specific positions and joints. It is also capable of tracking a rigid body by using three sensors on the actor's arm such that the sensors do not lie in a straight line. We use three sensors to track the arm of
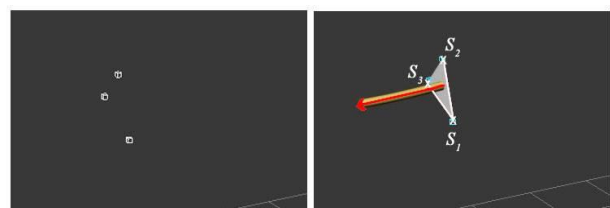

Fig. 8. The virtual sensors in the 3Ds max software. (a) The three virtual markers. (b) The normal of the generated polygon.
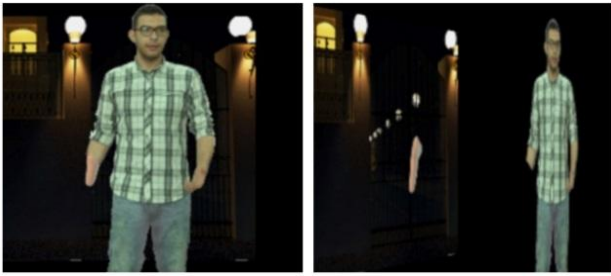
Fig. 9. Layer-based compositing technique.

the actor as shown in Fig. 7 (f).

E. Virtual object integration

After tracking the movement of the real arm using the MCS, we apply the tracking data to a 3D model to be integrated into the real scene. The MCS generates a tracking data that describes the XYZ position information of each sensor. We use three sensors to track the real arm, so we have three records of the position data of the sensors for each frame. The orientation of the real arm can be obtained using the cross product of the vectors of the three sensors [17] as shown in the following equation:

$$P = [S_2 - S_1] \times [S_3 - S_1] \qquad (2)$$

Where S_1, S_2, and S_3 are the 3D positions of the three sensors that do not lie in a straight line and P is perpendicular to the triangle that is generated using those three sensors, as shown in Fig. 8.

After finding the transformation matrix of the generated polygon, we set the transformation matrix of the 3D object that is integrated into the real scene to the transformation matrix of the generated polygon.

For creating a virtual camera that has the same parameters of the real camera, we use the registration method [18]. After the registration step, we render the 3D object that aligns the real arm of the actor. The rendering process generates frames of the 3D object that is used in the compositing step.

F. Compositing

After rendering the extracted actor and the integrated 3D object, the compositing step combines the extracted actor whose arm is separated with the integrated 3D object (e.g., a 3D Prosthetic, a 3D metal arm, or a 3D

Fig. 10. Final result after compositing

amputated arm). We use layer-based compositing technique to combine the real actor with the 3D object and background footage as shown in Fig. 9.

In some cases, we need to remove the rest part of the arm, so a skin color segmentation technique is needed. Modeling the human skin color is a very challenging field [19]. We use the 3D tracking data to generate an extended mask around the green arm to restrict the area of working and apply the Mahalanobis distance [20] to remove the skin color of the rest of the arm that gives high rates of skin detection process [21]. The Mahalanobis distance is considered a very efficient technique to determine the similarity between an unknown sample to a set of predefined samples [22]. The Mahalanobis distance can be obtained using the following equation

$$D_M(x) = \sqrt{(x-u)^T S^{-1}(x-u)} \qquad (3)$$

Where u and S are the mean and the covariance matrix of a sample of the skin colors, and x is the intensity of each pixel in YCbCr color space. After calculating the Mahalanobis distance, we remove any pixel under a predefined threshold.

Because of the possibility of the different lighting spaces between layers, color correction process handles this difference by using black and white point matching technique [10]. Black and white point matching technique adjusts the black pixels of the luminance version of the foreground layer to match the black pixels of the luminance version of the background layer. Similarly, in white pixels of the foreground layer and the background layer. Finally, we generate the final frame after integrating the 3D object (e.g. amputated arm) into the real footage of the actor with a background that has the same viewpoint as shown in Fig. 10.

III. RESULTS AND DISCUSSION

We have described a medium-cost system for integrating a virtual amputated arm with an actor. The proposed system removes a green arm of the actor to attach any 3D model that acts as the amputated arm. We
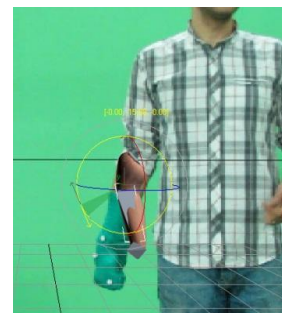

Fig. 11. Additional animation keys to give more realism to movement of the 3D object.

used the Naturalpoint MCS [23], which is considered a medium-cost MCS compared with other systems. Table 1

shows a comparison between the Naturalpoint MCS and other systems. The comparison shows that, the Naturalpoint MCS is the lowest cost MCS with some restrictions, such as the sensors must be placed above the clothes of the actor, and the calibration volume of the system is (12'×12'× 8'). We overcome the problem of the sensors by using only three sensors that are unnoticed.

The problem of the restrictions on the calibration volume is overcome by extracting the actor from the scene to be combined with another background that has the same viewpoint of the actor.

We applied the proposed system on a wide variety of records of the actor at 24 frames per second with 1920 x 1080 pixels for each frame. For modeling the 3D objects, the 3ds Max software was used.

We used a MCS to track the green arm as a rigid body, so there are some animation keys were set to give more realism to the movement of the 3D object, see Fig. 11.

We used several 3D objects (e.g. a virtual amputated arm, a virtual metal arm, a virtual skeleton arm, or a virtual bone arm) to be integrated with the actor as shown in Fig. 12.

We generated several examples of the integrated 3D object with the actor. The actor with a virtual amputated arm is shown in Fig. 13 (b). The actor with a virtual arm is shown in Fig. 13 (c-f). We replaced the real arm of the actor with a virtual smoke as shown in Fig. 13 (g).

The proposed system composites the actor with a background footage with restrictions on the viewport of the background. The final result of the proposed system is shown in Fig. 14.

Limitations

The limitations of the proposed system appear in the quality of the final results. Because of the foreground extraction process, the quality of the footage is not good enough. The reason for this limitation is that, we use a medium-cost MCS with a restricted tracking volume to keep the cost in a reasonable range. Another limitation of the proposed system is that, the proposed system is an off-line system. So any fault in the recording process requires repeating all the processes of the system.

IV. CONCLUSIONS AND FUTURE WORK

We have presented a medium-cost system that generates an integrated 3D object to a real actor with any
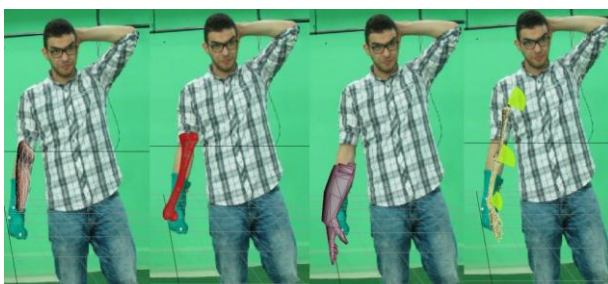
scene that has the same viewpoint of the actor. By extracting the real actor from the scene, the proposed system solves the problem of the clipping volume in the medium-cost MCS. The arm of the actor is removed using an efficient chroma-keying technique to generate the actor without his arm. The tracking data of the MCS is applied to any 3D object in the integration process. White point matching technique is used in the color correction process to match the lighting spaces between the layers of the final footage.

The proposed system is currently an off-line system because of the dependency on 3Ds max program that takes a great time to render the high quality frames. In the future work, we will develop a MAXScript to generate a real-time system for generating a draft version of the output. By using this draft version, we ensure that any fault in the recording process will be fixed simultaneously.

Table.1 A comparison between the Naturalpoint motion capture system and other systems

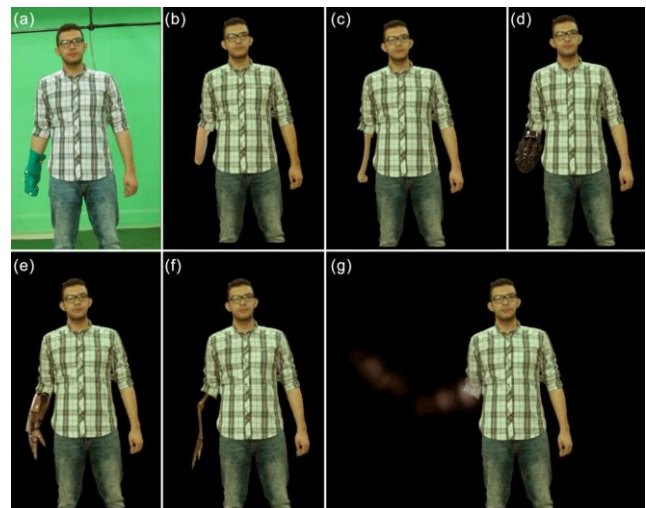|	Naturalpoint [23]	Xsens MVN [24]	FAB System [25]
Capture volume	12'×12'×8'	Unlimited	Unlimited
Cameras	24 cameras	Cameraless	Cameraless
Sensors	34 sensors that must be attached above clothing	17 small sensors that can be attached below or above clothing	17 MTx sensors
Cost	Medium	High	High

Fig. 13. The results of the integration of the 3D objects. (a) The original footage. (b) A virtual amputated arm. (c) A virtual bone arm. (d) A virtual weapon arm. (e) A virtual metal arm. (f) A virtual skeleton arm. (g) A virtual smoke that generated by a particle system.

Fig. 12. The integration of virtual objects with the actor

Fig. 14. The Result of the proposed system

See video in http://www.aun.edu.eg/faculty_computer_information/video/

REFERENCES

[1] Azuma, Ronald T. "A survey of augmented reality." Presence 6.4, 1997: 355-385.

[2] Carmigniani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., & Ivkovic, M. "Augmented reality technologies, systems and applications." Multimedia Tools and Applications, 51(1), 2011, 341-377.

[3] Khaled F. Hussain, Adel A. Sewisy, Islam T. El- Gendy. "Augmented Dressed Body System Controlled By Motion Capture Data", International Journal of Computing Academic Research (IJCAR), 2013, 2(1), 1-13.

[4] Frati, V., & Prattichizzo, D. (2011, June). "Using Kinect for hand tracking and rendering in wearable haptics." In World Haptics Conference (WHC), 2011 IEEE (pp. 317-321). IEEE.

[5] Alexander, O., Rogers, M., Lambeth, W., Chiang, J. Y., Ma, W. C., Wang, C. C., & Debevec, P. "The digital Emily project: Achieving a photorealistic digital actor." Computer Graphics and Applications, IEEE, 2010, 30 (4), 20-31.

[6] Jain, Arjun, Thorsten Thormahlen, Hans-Peter Seidel, and Christian Theobalt. "Moviereshape: Tracking and reshaping of humans in videos." In ACM Transactions on Graphics (TOG), ACM, vol. 29, no. 6, pp 148:1-148:10, 2010.

[7] Stargate studios, http://www.stargatestudios.de/shows/demos. Last accessed 26-4-2014.

[8] The Curious case of Benjamin Button, http://www.cgsociety.org/index.php/CGSFeatures/CGSFeatureSpecial/curious_case_of_benjamin_button. Last accessed 27-4-2014.

[9] Review: Stolen, http://www.riveronline.co.uk/art-ents/review-stolen. Last accessed 27-4-2014.

[10] Wright, S. "Digital compositing for film and video." CRC Press, 2013.

[11] Stauffer, Chris, and W. Eric L. Grimson. "Adaptive background mixture models for real-time tracking." Computer Vision and Pattern Recognition, IEEE Computer Society Conference on. Vol. 2. IEEE, 1999.

[12] Kelly, Philip and Ó Conaire, Ciarán and Monaghan, David and Kuklyte, Jogile and Connaghan, Damien and Pérez-Moneo Agapito, Juan Diego and Daras, Petros. "Performance analysis and visualisation in tennis using a low-cost camera network." Multimedia Grand Challenge Track at ACM Multimedia 2010.

[13] Chu, H., Ye, S., Guo, Q., & Liu, X. "Object tracking algorithm based on camshift algorithm combinating with difference in frame." In Automation and Logistics, IEEE International Conference on, 2007, (pp. 51-55). IEEE.

[14] Farbman, Zeev, et al. "Coordinates for instant image cloning." ACM Transactions on Graphics (TOG). Vol. 28. No. 3. ACM, 2009.

[15] Obdrzalek, S., et al. "Accuracy and robustness of Kinect pose estimation in the context of coaching of elderly population." Engineering in Medicine and Biology Society (EMBC), Annual International Conference of the IEEE. IEEE, 2012.

[16] Natural point: Products, http://www.naturalpoint.com/optitrack/products/arena/indepth.html. Last accessed 4-5-2014.

[17] Vince, John A. "Mathematics for computer graphics." London: Springer, 2010.

[18] Breen, D. E., Rose, E., & Whitaker, R. T. "Interactive occlusion and collision of real and virtual objects in augmented reality." Munich, Germany, European Computer Industry Research Center, 1995.

[19] Jagadesh, B. N., K. Srinivasa Rao, and Ch Satyanarayana. "A Robust Skin Colour Segmentation Using Bivariate Pearson Type II$\alpha\alpha$ (Bivariate Beta) Mixture Model." International Journal of Image, Graphics and Signal Processing (IJIGSP) 4(11), pp. 1-8, 2012.

[20] Hsu, R. L., Abdel-Mottaleb, M., & Jain, A. K. Face detection in color images. "Pattern Analysis and Machine Intelligence", IEEE Transactions on, 2002, 24(5), 696-706.

[21] Kakumanu, P., Makrogiannis, S., & Bourbakis, N. "A survey of skin-color modeling and detection methods. Pattern recognition", 2007, 40(3), 1106-1122.

[22] Pujari, Jagadeesh D., Rajesh Yakkundimath, and Abdulmunaf Syedhusain Byadgi. "Automatic Fungal Disease Detection based on Wavelet Feature Extraction and PCA Analysis in Commercial Crops." International Journal of Image, Graphics and Signal Processing (IJIGSP) 6(1), pp 24-31, 2013.

[23] Natural point, http://www.naturalpoint.com. Last accessed 4-5-2014

[24] Roetenberg, D., Luinge, H., & Slycke, P., "Xsens MVN: full 6 DOF human motion tracking using miniature inertial sensors." Xsens Motion Technologies BV, Tech. Rep, 2009.

[25] Biosyn systems, http://www.biosynsystems.net/f-a-b-system/. Last accessed 4-5-2014.
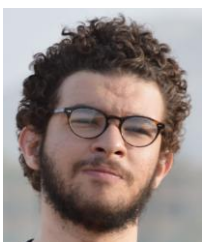
Mahmoud Afifi received the B.S. degree in Information Technology in 2009 from Faculty of Computers and Information, Assiut University, Egypt. He is currently a M.S. student and a teacher assistant at the department of Information Technology, Assiut University, Egypt. In addition, he is a 3D modeler and animator. His research interests include multimedia systems, digital actors, image processing, and computer vision.

Mostafa Korashy received the B.S. degree in Information Technology in 2011 from Faculty of Computers and Information, Assiut University, Egypt. He is currently a M.S. student and a teacher assistant at the department of Information Technology, Faculty of Computers and Information, Assiut University, Egypt.
His research interests include computer vision, computer networks, image processing, and robotics.

Ebram K. William received the B.S. degree in Information Technology in 2010 from Faculty of Computers and Information, Assiut University, Egypt. He is currently a M.S. student and a teacher assistant at the department of Information Technology, Faculty of Computers and Information, Assiut University, Egypt.
His research interests include computer networks, computer vision, and network security.

Ali H. Ahmed received the B.sc in Information Technology in 2008 from Faculty of Computers and Information, Assiut University, Egypt. He completed master degree in 2013 in wireless sensor networks. Besides the research work in his PhD, he is a teaching assistant in information technology department and he had more that 6 years in the academic work. His research interests include WSN, image processing, and computer vision.

Khaled F. Hussain his BSc and MSc Degrees in Electrical Engineering, both from Assiut University, Assiut, Egypt, in 1994 and 1996, respectively, and the PhD degree in Computer Science from the University of Central Florida, Orlando, in 2001. From 2002 to 2006, he worked as a Visiting Assistant Professor with the University of Central Florida. From 2006 to 2007, he worked as an Assistant Professor with the Department of Electrical Engineering, Faculty of Engineering, Assiut University. Since 2007, he has been with the Department of Computer Science, Faculty of Computers and Information, Assiut University, and is currently an Associate Professor, Executive Director with the Multimedia Laboratory, and Head of the Department of Multimedia, Faculty of Computers and Information, Assiut University. His major research interests include computer vision, computer graphics, augmented reality, and computer animation.