Motion Capture and Animation Using Digital Image Processing

Eliza Mae A. Saret and Jaime Samaniego

***Abstract*— We present a practical method for human posture recognition and animation through the use of digital image processing as an alternative for expensive means of motion capture. Given an input video of a moving subject wearing colored markers, the application identifies the position of the body parts and maps its movements into a cartoon character.**

***Index Terms*— motion capture, animation, digital image processing**

I. Introduction

A. Significance of the Study

According to an article by Steven Dent, motion capture or mocap is defined as the process of transferring complex actions of a model subject into an animated character. It can be traced back in 1914 where Animator Max Fleischer invented rotoscoping which uses the method of tracing the movements and facial gestures frame-by-frame from live-action footages. Creating this form of digital animation by hand takes a lot of time and human effort to complete. Wanting to automate this process, bio-kinetic researcher Tom Calvert from Simon Fraser University invented mechanical capture suites which used electronic sensors to output computer animated figures. This paved way to the use of more advanced methods in motion capture like the ones that are commonly used in animated films nowadays. On-set performace capture requires actors to wear tight suites with markers placed on different parts of the body and be surrounded by several special cameras while they act out their scenes. Various software applications are then used to map these markers on 3D characters such as Autodesks Maya, MotionBuilder and 3D Max. Recently, Xbox had also released KINECT 2 sensor, a hardware which allows bone and facial gesture tracking through infrared vision and could be used in navigating and interacting with the console. [1] [2]
Alexandre Szykman and João P. Gois mentioned on their study last 2014 that the cost of motion capture can be a huge barrier for people who want to create animations for the sake of art and science. They further stated that animations that are the same level as Hollywood films are overpriced and that independent professionals who want to create such animations often do not make enough profit to pursue mocap-based animation. It can also be seen from the previously given examples that using motion capture can be impractical for independent creatives especially for those who are only starting in the field of animation, those who aim to map human motions for scientific studies, or those who just want to explore motion capture for entertainment purposes. Due to these reasons, it is reasonable to look for a more convenient and less expensive means of motion tracking that will allow more people to use motion capture in animation. [3]
A possible way of tracking human gesture can be through the use of digital image processing. Instead of purchasing special cameras or electronic motion sensors to track the movements of the human body, the computer itself could be the one detecting the required anatomical features of a moving person by applying human body detection algorithms into a given input video. Color detection and the application of human body geometric constraints will be used in the identification of these anatomical features. Once the needed features are successfully detected, a skeletal figure of the moving human could be outputted and mapped into a character making the whole process more convenient.

The result of this study created a way to extend motion capture to people who are unable to afford the special equipments needed for this technology. By testing the use of digital image processing in determining human body movements, the feasibility of using these techniques in animating human motions was confirmed. These results can be used in the development of a more practical means of motion capture and the reduction of the production cost in film and game animation.

B. Objectives

This study aims to create a computer program that can be used to facilitate motion capture and animation by fulfilling the following objectives:

1) To use digital image processing and knowledge in human body geometric constraints in detecting and identifying human body parts
2) To use the detected human body parts in drawing a moving skeletal figure representing the motion executed
3) To map the motions of the moving skeletal figure in animating a character

C. Review of Related Literature

Few studies have been conducted on extracting data from human body images through the use of digital image processing. These informations were applied in various fields such as in surveillance systems, motion analysis and medical

Presented to the Faculty of the Institute of Computer Science, University of the Philippines Los Baños in partial fulfillment of the requirements for the Degree of Bachelor of Science in Computer Science

analysis. Researchers were able to separate detected human figures from photos and identify their body parts using various algorithms.

Park et al. (1999) reconstructed an image of a human body into a 2D model with labeling on ten main body parts (the head, two upper arms, two forearms, the torso, two thighs, and two shins), twelve joints (two shoulders, two elbows, two wrists, two hips, two knees, two ankles) and two pseudo joints (the neck and pelvis). Watershed segmentation algorithm and region merging was used in determining the regions in the image. Skin region extraction is then applied by marking prominent skin color regions. Curve segments are then used to enclose the detected skin color regions in order to create 2D ribbons that can fit a human body. To determine the posture of the body, the 2D ribbons were labeled into what parts they represent. After these steps, a stick figure representing the model could be generated. [4]

A study conducted by Chowdary et al. (2014) explored and discussed different gesture recognition methods in detecting the number of fingers held out by a hand on a plain background. These methods were the use of pixel count algorithm, detection of circles, morphological operations, and scanning method. All the given methods involved converting the input into a binary image representing the skin region as white pixels and the background as black pixels. In pixel count algorithm, the number of white pixels are counted and classified according to predefined ranges representing the number of fingers. Marking the fingers with circles before taking the image and counting these markers in the program can also be used in getting the needed information. The use of morphological operations involve using hit or miss operation which will result to an image with only one side of every finger. Dilation is used to be able to further identify and count the objects that remained in the image. The scanning method could be classified into two ways: linear horizontal scanning and linear vertical scanning which divides the image into two. The number of detected transitions from 1 to 0 in the binarized image is equal to the number of fingers held out. At the end of the study, the researchers were able to conclude that the best way to detect the number of fingers was through the scanning method which gave accurate results for 82.47% of the input images. [5]

On the same year, Toshev and Szegedy used deep neural networks (DNN) in human pose estimation. Researchers were able to detect four most challenging limbs: lower arms, upper arms, lower leg, and upper leg given an input image of a human body. Prediction of poses were used to identify the position of body parts that are not visible on the image. At the end of the research, it was concluded that the use of a DNN-based approach gives an advantage of capturing pose in a holistic manner and outputting a more realistic models for input images. [6]

These studies showed that the use of digital image processing techniques in determining human body parts in images and identifying their positions is possible through various algorithms. Since a video can be defined as only a series of images, some of these concepts were also applied in capturing motion in videos. The study aimed to identify the different poses done by a human in an input video and represent it in animating a model.

D. Date and Place of Study

This study was conducted from January to April, 2018 at the University of the Philippines Los Baños, Institute of Computer Science.

II. Materials and Methods

OpenCV 3.3.0 C++ and Qt Creator 4.5.1 were used in creating the application. OpenCV C++ was used in preprocessing the input videos, identifying body parts and in creating the character animation. Meanwhile, Qt Creator was used to create the user interface for playing and configuring the animations.

Since the study aims to create an application that will minimize the cost of motion capture, there was no need to use special cameras or electronic motion sensors in capturing the human subjects. Any device with a camera can be used to take the input video for the application. A total of 11 colored markers were placed on the human subject in order to track its movements. Seven colors respresenting different body parts are recognized by the application: red, blue, yellow, green, orange, purple and pink for the head, neck, hips, left hand and foot, left elbow and knee, right hand and foot, and right elbow and knee respectively. The captured videos were inputted in the application where its frames were subjected to preprocessing, body part identification, and animation using OpenCV C++.

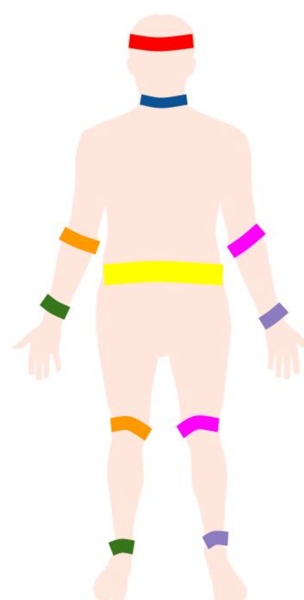


Fig. 1: Positioning of Colored Markers on Human Subject

A. Setup

A mini-studio was setup in order to produce an input video with colored markers that can easily be recognized by the program. This studio consists of two small led lights to provide balance amount of light on all parts of the frame, a plain white background to prevent any wrong color detections that do not belong to the subject, and a tripod to avoid shakiness in the video. The input video was taken using a cellphone camera.

B. Preprocessing

Due to possible differences in lighting where the input video is taken, options for adjusting some properties in the input videos was added in the application since it is an essential part in the process of body part identification. The properties included in the configuration part of the application were contrast, brightness, saturation, and the HSV (hue-saturation-value) values for the range of colors to be detected.

C. Body Part Identification

The location of the markers present inside a single frame was detected by finding the largest group of points with the given color representation. The midpoint of the chosen group is then assumed to be the coordinates of the given body part. In order to identify body parts when a single color is used to represent them, it was assumed that the one found at the upper part of the frame is either the hand or the elbow and the one found at the lower part is the leg or the foot. The endpoints representing the different detected parts are connected in order to create a simple stick figure.

D. Animation

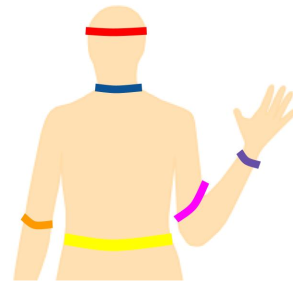
From the drawing of a stick figure, the animation was done by mapping the captured movements into a character. These characters were drawn using combinations of some drawing functions in opencv c++ which was mostly used for limbs and importing images on body parts that were more detailed and complex to draw manually. The angle between two endpoints were also taken in order to know how much an image is to be rotated before it is placed in the frame.

Three videos were displayed in the application. The first one displayed the adjusted input video. The second contained a stick model of the subject's movements. The last video showed the animated movement of the input video. The user can choose which character to be animated inside the video settings. At the same time, the user can also adjust settings such as contrast, brightness, saturation, and the threshold values for the colored markers used.

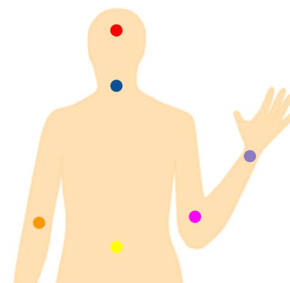
demo videos:

1) `https://drive.google.com/file/d/16WUcdiERpY_tmkyTIbcj1brSWtFplAHv/view?usp=sharing`
2) `https://drive.google.com/file/d/1w3JnSumqBOnleQeoQKxJUXnkWWxrVJZx/view?usp=sharing`

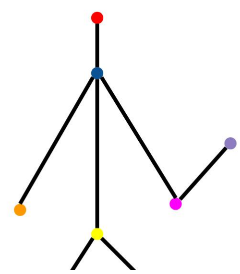
repository:
`https://github.com/lizsaret/Motion-Capture-Animator.git`



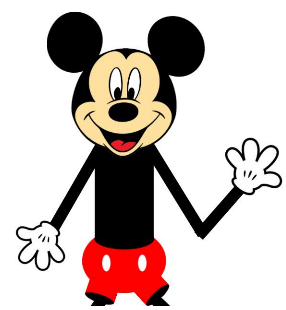
(a) Human Subject with Colored Markers



(b) Color Detection



(c) Body Part Identification and Model Construction



(d) Mapped Animation

Fig. 2: Body Part Identification Stages

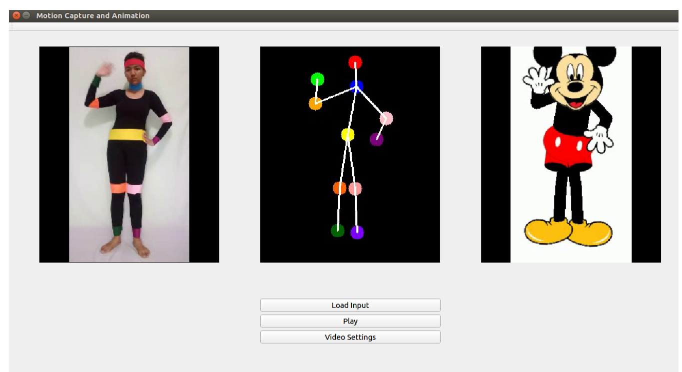


Fig. 3: Final output page of the application

III. Results and Discussion

Three sets of input videos were used to test the ability of the application to detect body movements and animate them. The first one was taken with the black background and inconsistent lighting which resulted in mistakes in body part identification and inaccurate animation of the subject's movements. The second set was taken using the mini-studio setup with a balanced amount of light throughout the frame. Detection of body parts and animation were more correct with less misdetections as compared to the first. For the third set, the mini-studio setup was still used but the subject's movements were made so that it would purposely break the body part assumptions of the program such as when not all the body markers are present in the frame and when the knee or foot is above the elbow or hand. There were more misdetection in the body part identification in the third set as compared to the first and second which lead to wrong postures in animating the movements of the human subject.

On running the input videos on a machine with Intel Celeron processor N3160, it takes the application a few seconds to a minute

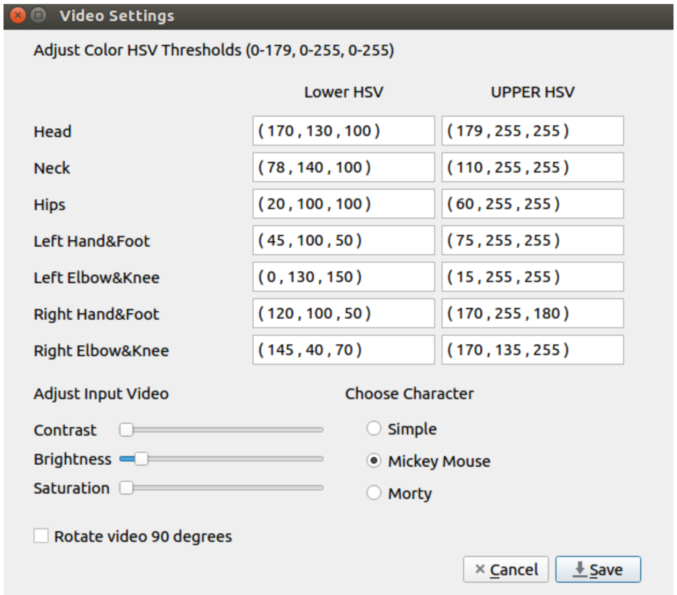


Fig. 4: Video configuration options

to produce the outputs depending on the length and the quality of the input video. For videos of lengths five seconds, ten seconds, fifteen seconds and twenty seconds all with 720p video quality, the application took an average of 8.28 seconds, 13.89 seconds, 22.41 seconds, and 30.53 seconds to process and write the outputs respectively. This is due to the amount of processes done on every frame of the input videos before finally displaying it to the user once all of the movement detection and animation is done.

IV. Conclusion and Future Work

The detection of colors using digital image processing worked well in the identification of body parts and in keeping track of the subject's movements. The output of the demo videos proved that digital image processing can be used as an alternative for expensive means of motion capture based animation. The application created can be used for studying the movements of a subject or for creating small animation projects.

Since the program uses threshold values for color detections, it is important that there is a consistent amount of light throughout the whole frame. Having colorful backgrounds for the input video is also discouraged in order to avoid wrong identification of body parts. Some issues in the animation, such as overlapping body parts and wrong posture detections, also arise when not all the markers are visible inside the frame.

Other improvements to the application can be done such as fixing the animation problems when the markers can't be seen on the frame or when the subject's body is not facing the camera. Using better human body constraint definitions can also help in increasing the amount of movements a subject can do.

References

[1] S. Dent. What you need to know about 3d motion capture. [Online]. Available: https://www.engadget.com/2014/07/14/motion-capture-explainer/

[2] ——. How i turned my xbox's kinect into a wondrous motion-capture device. [Online]. Available: https://www.engadget.com/2015/03/08/using-the-kinect-for-motion-capture\/?ncid=rss_truncated/

[3] A. G. Szykman and J. P. Gois, "Evaluating the feasibility of low cost motion capture systems," 2014.

[4] J. Park, H. Oh, D. Chang, and E. Lee, "Human posture recognition using curve segments for image," 1999.

[5] R. Chowdary, N. Babu, T. Subbareddy, B. Reddy, and V. Elmaran, "Image processing algorithms for gesture recognition using matlab," 2014.

[6] A. Toshev and C. Szegedy, "Deeppose: Human pose estimation via deep neural networks," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

Eliza Mae A. Saret is an undergraduate BS Computer Science student at University of the Philippines Los Baños. She enjoys reading graphic novels and sketching during her spare time.