

The Used of Video Tracking for Developing a Simple Virtual Boxing

David HabsaraHareva^{#1}, Martin^{#2}

^{#1,2}Informatics Department, Computer Sciences Faculty, PelitaHarapan University

Jl. M.H. Thamrin Boulevard Tangerang, 15811 Banten, Indonesia

¹david.hareva@uph.edu

Abstract— Now, video tracking is a feature that used a lot in computer gaming. This feature makes games more interesting because it offers direct interaction between the player and computer while in the same time player can exercise his/her body and eye coordination. For this reason, implementing video tracking in virtual boxing is a challenging task. Compared to the standard computer game controllers, controlling game using video tracking is more appealing and attractive. The aimed is to explore the used of video tracking at developing a simple virtual boxing. In this study the used method for detecting object such punch types was a Mean-shift method. This detection was done based on the moving area of the object's surfaces, thus this information could be used to differentiate punches type, which is classified as Jab, Hook, and Uppercut. The making of Virtual Boxing comprises of video tracking development and gaming design. This development of virtual boxing involves a number of configurations such as webcam configuration, color tolerance setting in the detector, and tolerance of the color detection. Every video tracking modification and gaming design is evaluated thoroughly. All of configuration was evaluated to ensure that this virtual boxing meet the initial expectation. The result shows that the punch detection reading in using library color tracking using Mean-shift in a boxing game is good enough, but wasn't perfect yet. The Mean-shift method needs to be combined with other method in order to detect punches perfectly.

Keywords—Gesture, Mean-shift, virtual game, video tracking, webcam

I. INTRODUCTION

Video tracking has been involved in research area of video processing and analysis. Video tracking is an activity to find position of a single motion or multiple objects over time on a video camera [1,2]. It is used for several applications including human computer interaction, security [3,4,5] and surveillance [6], tele-video conference, augmented reality [7], traffic control, medical imaging, smart room [8], compressed video [9], and virtual games.

The peripheral devices like a mouse, a keyboard, a joystick, light pen, track ball, keypad, are used for interaction between human and computer machine, even though that those devices have attributes of unnaturalness and limitable response. Now days, video tracking is the most challenging way to get in touch with electronic or machine devices. It has been showed by

[10], video tracking improves creativities of input technique and offering more features and capabilities than traditional devices. Obviously, this innovation has opportunities to support an important component of natural language processing, that is gesture approach. As a method of computer system interaction, it will reduce users' labor to learn how to use them; they do not need to learn how to use the peripheral input devices, such as a mouse holding or a keyboard typing, except knowing how the application operates on the device.

Among the former of object gesture detection in the field of video tracking, Mean-shift is a poplar method to segment image and video [11,12]. It is a vigorous approach toward feature space or color space for mapping inputs from a small subset data at a time [13]. The Mean-shift algorithm is a simple method that searching a similar color randomly around the pointed object color. The searching color is applying based on the RGB (Red, Green, and Blue) value. This algorithm is efficient method for processing the object tracing of the 2-dimensional images [14]. But there is still a weakness when it is used for the exposure change (contrast). The illumination change causes alteration of the RGB color value, and then influences to accuracy of the object tracing [15].

For a couple of years, the popularity of video tracking in video games increased significantly. Therefore this study was done to enrich the variation of the video game in the case of virtual boxing. The user is able to directly interact in the game environment, because he/she is marked as an acting object. The marked object is identified through video processing that capture by embedded or separated camera. This is the main favorable feature of the game when a player can control the object in the game virtually without any peripheral device or cables for connecting to a computer machine; so long the object is in under coverage area of the view camera. In this game, a player will be trained to do several kinds of punches: Jab, Hook, and Uppercut by the trainer which is inside the virtual boxing game. For the computer to able indentify the user's fist, the right and the left hand use the glove which hasa different color. Hopefully this study will foster the virtual boxing not only just a game but also basic knowledge to who is interested in boxing techniques in the future.

II. METHOD

A. About Boxing

Boxing is a combat sport invoking two people that fight each other using their fists in a competition. This sport is typically consisted of tree basic techniques: stance, punches, and defence. For simply in development of virtual boxing, this study was focused on the punches technique to hit a target as an opponent. There are three basic punches in a boxing competition: Jab, Hook, and Uppercut (Fig. 1). If a boxer is right-handed, his left hand is the lead hand, and the right hand is the rear hand, and for left-handed is vice versa. The Jab is a straight punch down by a rear hand rotates 90 degrees with the lead hand from the hold position. Hook is a semi-circular punch thrown with the lead hand to the side of the opponent's head with the rear hand tucked firmly to protect to protect the chin. Uppercut is a punch with a rising punch thrown with the rear hand to lift the opponent's body [16].

B. Video Tracking for simple virtual boxing

The hand gestures are mainly used in virtual boxing game, where the kind of gestures in this limit study was punches technique. The computer comprehends with machine language, but not human language such gestures, therefore an interface protocol is needed to help interactions between them. For that purpose, a webcam sensor is able bridging their communication by recording the boxer's fist movements and its surroundings into the form of video format. The computer's task is to track a unique gesture which exhibited by a fist gesture, whether it was Jab, Hook, or Uppercut through analysing the video content. The detectable punch is interpreted as input by a computer machine to give the boxernotification responses [16]

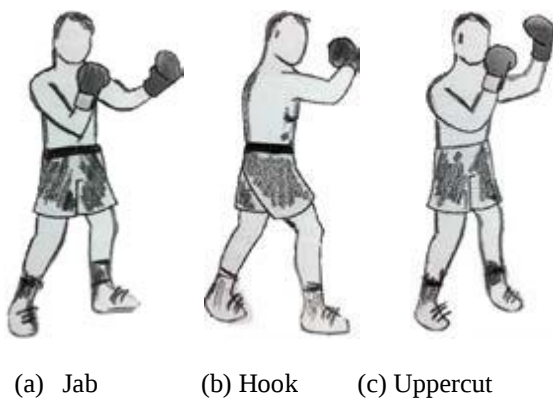


Fig. 1 The classification of the punch type

c. Video Tracking Process

The interactive communication that used gesture as natural language can be accomplished through implementing the video tracking process. Implementing video tracking in development of virtual boxing has several steps including webcam configuration, color detection, threshold and Mean-shift algorithm as depicted in Fig. 2.

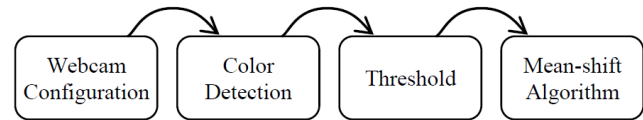


Fig. 2 Video tracking process

The first step is Webcam Configuration to detect the webcam device availability and to determine a display size of webcam's projection and frame value per second. In the case of virtual boxing, the windows size is considered to have a length (horizontal line) and a width (vertical line) 800 pixels and 600 pixels respectively. This size was compatible for all monitor in general. Several tests of frame rate were done to find the best performance, for example 24, 25 and 30 frame per seconds following the frame rate of film or TV.

The second step is color detection. The boxer is request to use hand gloves or anything that has a different color to distinguish the right or the left hand gestures. The hand glove color is marked by pointing it via mouse cursor to the pixel of glove color. So there are two pixels in the form of hexadecimal which consists of RGB value that will be detected as valid object gestures.

The third step is thresholding to regulate a number of grey image levels. The thresholding is the process to specify a value of tolerant color. Generally, it is done by changing the image quantization. In thresholding, the grey level value can be changed accordingly to remove unexpected pixel color. The background color can be omitted to easy process only a target color which has been defined as hand glove's color. The purpose of utilizing the tolerant color toward the defined glove's color is to observe the influence of its color toward its surroundings in the target punch area. The tolerant value is applying to put apart the glove's color and its background. For example if the value of tolerant color is 50, and the value of glove's color of Red, Green, Blue is 210, 50, 120 respectively, then the range glove's color of Red, Green, and Blue will be 160-260, 0-100, and 70-170, respectively. The color that those in specific ranges will be in process, others are disregarded. In this experimental study, the used density color for thresholding was 20, 30, 40, 45, 50, 55, 60, and 70. These color values are taking into account to obtain the appropriate tolerant color for punch gestures.

The last step of the process is Mean-shift method to specify a window around a dense region that are present in the feature

space. The Mean-shift algorithm based on ideas proposed by Fukunaga and Hostetler [17]. The dense region corresponds to local maxima or modes. Mean-shift considers feature space as density function of empirical probability. If there is a set of pixel color, the Mean-shift indicates them as sampling from the probability density function. It computes the mean of local estimated density and shifts the center of the window to the mean and repeat the algorithm until convergence.

The value of maximum window area is based on the value of a middle point where the clustered object is identified. Fig. 3 is an example of maximum window area with value 400 pixels, then window area will be traced from horizontal direction, left side to right side, and vertical direction upper side to bottom side. If the horizontal position of middle point is x , y then searching window area is traced along $x - 400$ to $x + 400$. Such in the case of vertical position is y then searching is performed for $y-400$ to $y+400$. For that, the window area with value 400 pixels will be 800 x 800 pixels area

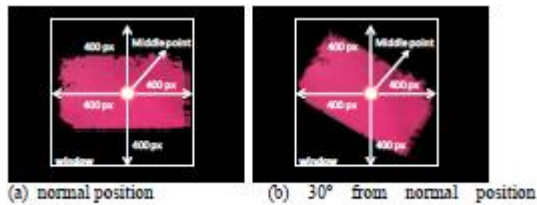


Fig. 3 The window area with tolerant value = 400 pixels

Deciding the middle point of window area is carried out by finding the length and width of detection area. The obtained length is summing the pixels from the left to the right side area, and also the width is summing the pixels from upper to bottom side area. The middle point is the crossing point between the length and the width. For each sequence pictures of a video, the middle point is searched by the Mean-shift algorithm.

III. RESULT

A. Webcam Configuration

The aspect ratio for webcam resolution used standard resolution, it was 800 x 600 pixels. This resolution was suitable enough for grasping the punch gestures. Several frame rates: 24, 25, and 30 frame/seconds were tested on webcam and found that 30 frame/seconds obtained more detail image and better in reading the fist gestures; However the image process and file storage became heavier.

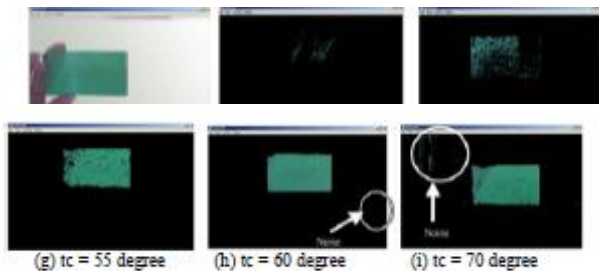


Fig. 4 Comparison of the tolerant color value in the thresholding process

B. Thresholding

The suitable tolerant color (TC) for this simple virtual boxing was examined. The Fig. 4 shows several objects' color which has different tolerant color. Those can be observed for TC = 50 degree, the object appearance shows clearly than lower TC value (TC= 20, 30, 40, and 45 degree). If the degree of TC is increased (TC = 55, 60, and 70 degree), the other color is came out of the object (noise) even though the object appearance is clearly seen. Therefore the used TC for the thresholding was 50 degree.

C. Mean-shift Algorithm

The maximum of window area for object's color detection was operated by reading the window area after stroked as depicted in Fig. 5. This trial was performed by striking punches 30 times, as a result the maximum width and height of detected window area was 397 pixels and 258 pixels respectively (Fig. 6). For easier in computation, this maximum width was rounded became 400

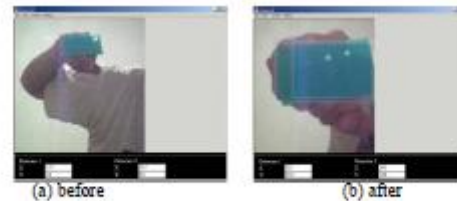


Fig. 5 The window area before and after stroked

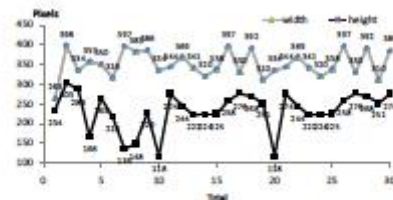


Fig. 6 Result of 30 punches for width and height of window area

D. The Punch Trials

The aim of this game was to detect the punch type (Jab, Hook, or Uppercut) where each punches consist of two hands represented by left glove and right glove (Fig. 7). For each trial, the successful strike was tested using Gaussian-Mean-shift. The expectation of local maxima is Gaussian, and when it is non-Gaussian, Mean-shift generalizes the local maxima. It means Mean-shift algorithm processes each point pixels become convergence in general [17].

The single user used gloves with different color between left and right hand to distinguish left or right strike. The glove's

color was identified by the computer automatically when glove color appeared on the screen. To classify the punch type, information of fist area before and after doing a strike was measured. The performing punch trial was limited toward punch type which provided in the virtual boxing game, which is left Jab, right Jab, left Hook, right Hook, left Uppercut, and right Uppercut. There are several parameters used to detect the glove area, which are height ($y_{max}-y_{min}$), width ($x_{max}-x_{min}$), and area of hand glove.

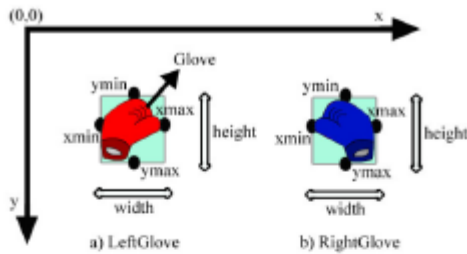


Fig. 7 The hand gloves and each window area around the gloves

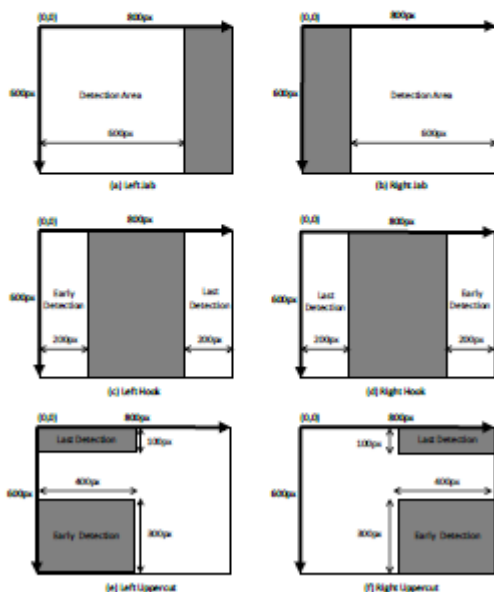


Fig. 8 The defined boundaries for targeted punches area

There are six boundary areas to distinguish the punch type as depicted in Fig. 8. If a punch is done in the target punch area of (a), 600 x 600 pixel from initial coordinate ($x=0$, and $y=0$); then it will be acquainted with a Left Jab. If a punch is begun in early detection area of (c) and ended in last detection area of (c), then it will be known as Left Hook. If it is happen to a punch which started in early detection area of (e) to the area of last detection of (e), it is recognized as left Uppercut, and so on. The target of the punch is mitts which hold by a boxing trainer.

In this game, the trainer will give several instructions to hit the mitt that is shown in Fig. 9.



Fig. 9 Instructions by boxing trainer

A player has to follow the instructions and hit the target as much as possible during 60 second. The punch which succeed hits the target will get a succeed notification and adding a credit score game. To give a more interesting game, there are three level difficulties, where the player is given a time limit for hitting the a mitt target: 3 second, 2 second, and 1 second for easy, medium, or hard level, respectively. Table I The alteration domain used for each type of punches. x-ord and y-ord are the middle point change

Punch type	window area		x-ord		y-ord	
	start	end	start	end	start	end
Left Jab	31534 - 39850	126911 - 150909	292 - 334	423 - 501		
Right Jab	30167 - 34926	118796 - 143860	495 - 565	526 - 618		
Left Hook	46243 - 54233	45524 - 53862	108 - 148	702 - 757		
Right Hook	44076 - 52554	43882 - 51669	694 - 756	116 - 162		
Left Uppercut	40110 - 48089	43485 - 51993	167 - 249	245 - 313	353 - 443	61 - 87
Right Uppercut	45023 - 53601	44782 - 53571	607 - 707	584 - 706	342 - 432	60 - 79

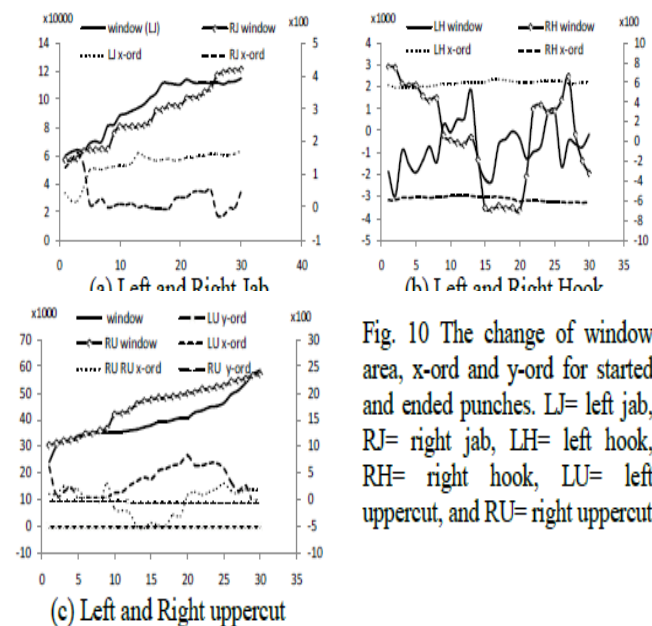


Fig. 10 The change of window area, x-ord and y-ord for started and ended punches. LJ= left jab, RJ= right jab, LH= left hook, RH= right hook, LU= left uppercut, and RU= right uppercut

Testing the window area and the movement of middle point of x-ordinate (horizontal direction) and y-ordinate (for Uppercut) are conducted in an experiment with 30 times punches. The boundary of all punches was followed the rules showed in Fig. 8. Table I shows results the alteration of window area and the middle point at initial position until accomplish of all punches. For example Left Jab punches have window area alteration 31.534 – 39.850 at starting point and 126.911-150.909 at ending point. The movement of middle point x-ord is 292-334 at initial point and 423-501 at ending point. The comparison of left and right punches for each punch types was almost similar, except the Hook punch, where the change of x-ord was reversed. The alteration of the window area of Jab at start to end point was significantly increasing from 60.000 to 120.000. It was happened with the Uppercut from 25.000 to 60.000. But not with the Hook, its movement change almost in horizontal direction. The alteration positions in middle point of Jab and Uppercut were fluctuated, but not with Hook, they appear remain the same. The movements of y-ord are not significantly different for all punches, and therefore this parameter was used only for the Uppercut punched.

IV.DISCUSSION

The implementation of simple virtual boxing using video tracking could be done well. The Mean-shift algorithm was useful to detect the gesture of the hand glove and successfully enough to provide natural inputs to this virtual boxing game. The speed response toward the game's instructions to strike the mitts has time-delay, it depends on the webcam and computer hardware specification. The boundary of punches using the hand glove color, the window areas, and the movement of middle points could be used to classify the punch type like Jab, Hook, and Uppercut.

REFERENCES

- [1] H. Kato, M. Billinghurst, *Marker Tracking and HMD Calibration for a video-based Augmented Reality Conferencing System*. In Proceedings of the 2nd ed., International Workshop on Augmented Reality (IWAR 99). October, San Francisco, USA. 1999.
- [2] S. Kang, J. K. Paik, , A. Koschan, B. R. Abidi, and M. A. Abidi, *Sixth International Conference on Quality Control by Artificial Vision*. 2003, vol 5132, pp. 103-111.
- [3] V. Kettner and R. Zabih, *Bayesian multi-camera surveillance*. Journal Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Fort Collins, CO, 1999.
- [4] M. Greiffenhagen, D. Comaniciu, H. Niemann, and V. Ramesh, *Design, analysis and engineering of video monitoring systems*. Proceedings of the IEEE, 2001, vol 89(10).
- [5] R. Collins, A. Lipton, H. Fujiyoshi, and T. Kanade, *Algorithms for cooperative multisensor surveillance*. IEEE. 2001, vol. 89(10), pp. 1456-1477.
- [6] R. Collins, A. Lipton, H. Fujiyoshi, and T. Kanade, *Algorithms for cooperative multisensor surveillance*. IEEE. 2001, vol. 89(10), pp. 1456-1477.

- [7] I. Ferrari, t. Tuytelaars, and I. V. Gool, *Real-time affine region tracking and coplanar grouping*. Proc. IEEE conf. On computer vision and pattern recognition, Kauai, Hawaii. 2001, vol. II.
- [8] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, *Pfinder: Real-time tracking of the human body*. IEEE Trans. Pattern Anal. Machine Intell. 1998, vol. 19, pp. 780-785.
- [9] A. Purushothaman, K. R. Shankarkumar, R. Rangarajan, and A. Kandasawam. *Compressed Novel Way of Tracking Moving Objects in Image and Video Scenes*. European Journal of Scientific Research. 2011, Vol.64, No.3, pp. 353-360.
- [10] J. Alon, V. Athitsos, and S.Sclaroff, *Accurate and Efficient Gesture Spotting via Pruning and Subgesture Reasoning*. Boston U. Computer Science Tech. Report No. 2005-020. 2005.
- [11] K. Newlander, J. Goldstein dan J. Means, *Video Change Detection and Tracking*. Design, 2007.
- [12] S. Paris and F. Durand, *A Topological Approach to Hierarchical Segmentation using Mean-shift*. IEEE Conference on Computer Vision and Pattern Recognition, 2007.
- [13] D. Comaniciu and P. Meer, *Mean-shift: A robust approach toward feature space analysis*. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002.
- [14] D. Comaniciu, V. Ramesh, and dan P. Meer, *Kernel-Based Object Tracking*. Real-Time Vision and Modeling Department Siemens Corporate Research and Electrical and Computer engineering Department Rutgers University, 2003.
- [15] Y. Hayashi dan H. Fujiyoshi, *Mean-Shift-Based Color Tracking in Illuminance Change*. Proc. of ROBOCUP 2007 International Symposium, 2007.
- [16] R. Gordon, *Boxing Basics: The Techniques and Knowledge Needed to Excel in the Sport of Boxing*. 2008.
- [17] M. A. Carreira-Perpinan, *Gaussian mean-shift is an EM algorithm*, IEEE Trans Pattern Anal Mach Intell. 2007, 29(5), pp.767-76.