EMITTER International Journal of Engineering Technology
Vol.3, No.1, June 2015
ISSN: 2443-1168

Remo Dance Motion Estimation with Markerless Motion Capture Using The Optical Flow Method

Neny Kurniati[1,2], Achmad Basuki[1], Dadet Pramadihanto[1]

[1]Electronic Engineering Polytechnic Institute of Surabaya
Jalan Raya ITS Sukolilo, Surabaya 60111
E-mail: neng_titi@yahoo.com, {basuki,dadet}@pens.ac.id

[2]Surabaya Polytechnic
Jl. Raya Kendangsari 109 Surabaya

Abstract

Motion capture has been developed and applied in various fields, one of them is dancing. Remo dance is a dance from East Java that tells the struggle of a prince who fought on the battlefield. Remo dancer does not use body-tight costume. He wears a few costume pieces and accessories, so required a motion detection method that can detect limb motion which does not damage the beauty of the costumes and does not interfere motion of the dancer. The method is Markerless Motion Capture. Limbs motions are partial behavior. This means that all limbs do not move simultaneously, but alternately. It required motion tracking to detect parts of the body moving and where the direction of motion. Optical flow is a method that is suitable for the above conditions. Moving body parts will be detected by the bounding box. A bounding box differential value between frames can determine the direction of the motion and how far the object is moving. The optical flow method is simple and does not require a monochrome background. This method does not use complex feature extraction process so it can be applied to real-time motion capture. Performance of motion detection with optical flow method is determined by the value of the ratio between the area of the blob and the area of the bounding box. Estimate coordinates are not necessarily like original coordinates, but if the chart of estimate motion similar to the chart of the original motion, it means motion estimation it can be said to have the same motion with the original.

Keywords: Motion Capture, Markerless, Remo Dance, Optical Flow

1. INTRODUCTION

Motion capture has been developed and applied in various fields, such as in the movies special effects, games, animation, security systems, and the arts, particularly the art of dance. Dance has been characterized as an exciting and vibrant art that can be used in the educational setting to assist the growth of the student and to unify the physical, mental, and emotional aspects of the human being [4]. Remo dance is a traditional dance that tells the struggle of a prince who fought on the battlefield. At first, Remo dance is used for the show *Ludruk* (East Java folk art form of drama that was shown with dancing and singing). However, the development of this dance is often danced separately as a speech on state guests, danced in state ceremonies, or in the local arts festival. The main characteristics Remo dance are the bells located at the ankle (called *Gongseng*), shawl (called *Sampur*), nod and shake head motion, and horses' leg motion. There are varying styles of Remo Dance costume. One of them is *Surabayan* style. It consist of a red headband, clothes without buttons are black with royal style in the 18th century, the extent of the mid-calf pants are crocheted with gold needles, batik sarongs *Pesisiran* that hung down to the knees, *Setagen* tied at the waist, and a dagger slip rear, and bell at the ankle. Dancers wearing two shawls. One tied at the waist and one pinned at the shoulder [1].

Remo dance motions are unique. Limbs motions are partial behavior. This means that all limbs do not move simultaneously, but alternately. It required motion tracking to detect parts of the body moving and where the direction of motion. Remo dancer does not use body-tight costume. He wears a few costume pieces and accessories. If the mark is embedded in the dancer's body, it can damage the beauty of the costumes. In addition, these markers can interfere with the motion of the dancer. Therefore, we need a markerless method to detect the motion of Remo Dancer. That method is Optical Flow.

Optical Flow is a simple method to motion detection and can also be used for motion estimation in a series of images or video frames by calculating the differential value of the frame. In this method do not have to use a monochrome background because the background subtracted automatically as long as using static background.

In this paper the data input is Remo dance video that danced by one person. Each frame does the optical flow with Horn and Schunck method [6, 13] to determine blob area. Each blob is marked by a bounding box. To estimate motion direction, the bounding box coordinates of the same object in each frame are calculated differential value for x coordinates and y coordinates. If the absolute differential value of x is greater than the absolute differential value of y, then motion estimation is horizontal. The object moves to the left if the differential value of x less than 0 and moves to the right if the differential value of x greater than 0. Whereas if the absolute differential value of x less than the absolute differential value of y, then the object moving vertically. The object moves up if the differential value of y less than 0 and moves down if the differential value of y greater than 0. The differential value of x equal 0 and the differential value of y equal 0 indicates the object is not moving.

2. RELATED WORKS

Previous work [2] presented A Virtual Reality (VR) Dance Training System Using Motion Capture Technology. Their system is inspired by the traditional way to learn new movements—imitating the teacher's movements and listening to the teacher's feedback. A prototype of their proposed system is implemented, in which a student can imitate the motion demonstrated by a virtual teacher projected on the wall screen. Meanwhile, the student's motions will be captured and analyzed by the system based on which feedback is given back to them.

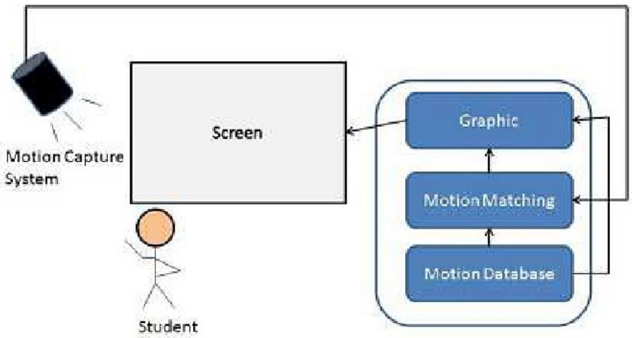

Figure 1. VR dance training system architecture [2]

The architecture of the system includes four components: 3D graphics, motion matching, motion database, and a motion capture system. Figure 1 shows the relationships between each component. The user's movements are obtained by the motion capture system are compared with the motions in the motion database through the motion-matching component. The 3D graphics component visualizes the movements by the user and the virtual teacher (template motion).

A virtual teacher demonstrating a dance motion, and the actual movement performed by a real dancer with marks on his body. The virtual teacher also appears when the student is practicing the moves. The student can thus imitate the moves of the teacher. Virtual teacher expressed in the avatar and student motion capture expressed in the skeleton articulated model [5].

Motion capture of dance movement in [2] paper became the basic idea of this paper. The same things between [2] and this paper is equally do capture of the dancers, then the dance movements simulated by the model. The differences are [2] captured in real time and the actual movement performed by a real dancer that wore a suit with markers attached. In this paper Remo dance capture from video file and Remo dancer wear costume without markers on it.

3. ORIGINALITY

In the paper [2,4,9,10,12] dancer wore body-tight costume with markers attached to it. In addition, they danced in front of a monochrome background. Different with [2,4,9,10,12], the originality of this paper is a markerless method and the dancer do not have to dance in monochrome background. Markerless is a suitable method because Remo dance has unique costume and accessories. Figure 2 shows the costumes and accessories consist of a red headband, clothes without buttons are black with royal style in the 18th century, the extent of the mid-calf pants are crocheted with gold needles, batik sarongs that hung down to the knees, *Setagen* tied at the waist, two shawls, a dagger slip rear, and bell at the ankle [1]. If using markers can interfere with the motion and the beauty of the costumes

(a) (b)

Figure 2. (a) Front View
(b) Back View

The optical Flow method used in this paper causes the background will be reduced automatically so it does not have a monochrome background. Another originality is the motion estimation method that using the development of differential methods. This method is simple, does not require complex mathematical calculations, so that it can be applied in real time.

4. SYSTEM DESIGN

Remo Dance Motion Estimation with Markerless Motion Capture Using the Optical Flow Method includes three components, i.e. Capture Video, Motion Tracking, and Motion Estimation is shown by Figure 3.

Capture Video is recording the movement of dancers that will generate AVI video file format. Each frame in the video is then processed motion tracking and motion estimation. Motion tracking is generally defined as identification of amounts to finding the object of interest from one frame in a subsequent frame of the video stream [6]. We may often want to assess motion between two frames (or a sequence of frames) without any other prior knowledge about the content of those frames. Typically, the motion itself is what indicates that something interesting is going on. Some kind of velocity can be associated with each pixel in the frame or, equivalently, some displacement that represents the distance a pixel has moved between the previous frame and the current frame. Motion estimation is the process of determining motion vectors that describe the transformation from one 2D image to another, usually from adjacent frames in a video sequence [6]. Detail of the System Architecture discussed below.

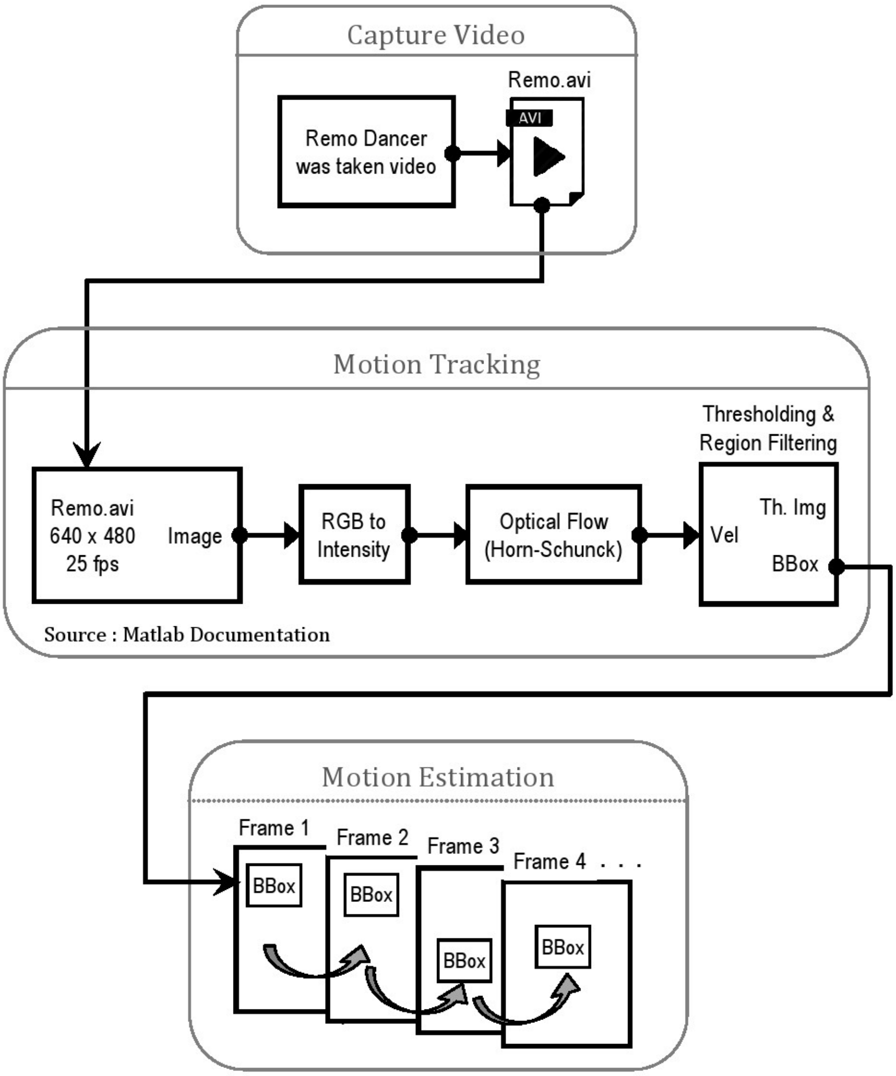

Figure 3. System architecture

4.1. Capture Video

Remo dancer wore costumes and accessories with no marks on his body were dancing on stage. In front of her was placed a static camera that recorded all the dance motions. Results captured in the form of video files to AVI format, as shown in Figure 4. AVI format chosen with consideration of this kind of format applies to standard Windows. AVI formats usually used as a basic format of video editing.

The Remo dance video has a duration of 7 minutes 33 seconds with frame rate 25 frames per second. Overall frame number 11,349 frames, which is partly shown in Figure 5. Each frame has a width of 640 pixels and height 480 pixels. This video will be processed motion tracking with optical flow method using the Horn-Schunck algorithm and motion detection process with Differential method.

Figure 4. Capture video

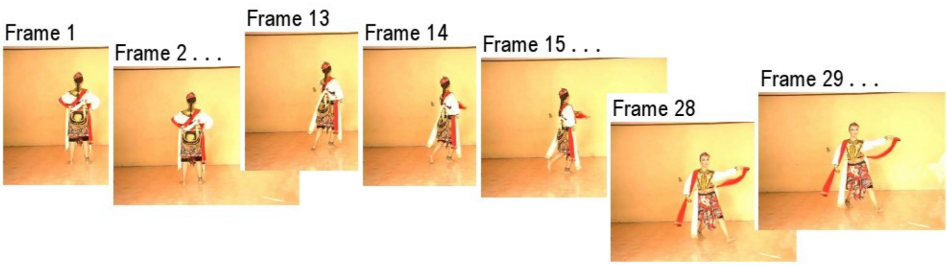

Figure 5. Video frames

4.2. RGB to Intensity

The RGB (red, green, and blue) to Intensity is converting RGB images to grayscale intensity image by eliminating the hue and saturation information while retaining the luminance [18]. Standard color TV and video systems generally are in the YUV color space [18]. The Y'UV model defines a color space in terms of one luminance (Y) and two chrominance (UV) components [18]. The YUV color example shown in Figure 6.

In video, luminance represents the brightness in an image (the "black-and-white" or achromatic portion of the image). Luminance is typically paired with chrominance. Luminance represents the achromatic image, while the chrome components represent the color information.

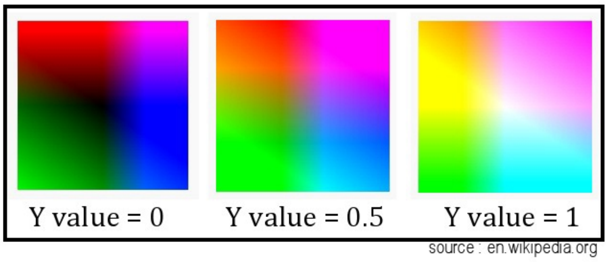

Figure 6. Example of UV color plane

YUV color space typically created from RGB source. Weighted values of R, G, and B are summed to produce Y, a measure of overall brightness or luminance. U and V are computed as scaled differences between Y and the B and R values [19]. *Recommendation BT.601*, more commonly known by the abbreviations *Rec. 601* or *BT.601* (or its former name, *CCIR 601*) is a standard originally issued in 1982 by the *CCIR* (an organization which has since been renamed as the International Telecommunication Union–Radiocommunication sector) for encoding interlaced analog video signals in digital video form [20]. Equation 1 defined conversion from RGB color space to intensity [21]:

$$intensity = \begin{bmatrix} 0.299 & 0.587 & 0.114 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

4.3. Optical Flow

Optical flow is the distribution of apparent velocities of movement of brightness patterns in an image [13]. Such a construction is usually referred to as a dense optical flow, which associates a velocity with every pixel in an image. The Horn-Schunck method attempts to compute just such a velocity field. One seemingly straightforward method is simply attempting to match windows around each pixel from one frame to the next.

To compute the optical flow between two images, must solve the optical flow constraint Equation 2 [16]:

$$I_x u + I_y v + I_t = 0 \quad (2)$$

In Equation 2, the following values are represented [16]:

- I_x , I_y and I_t are the spatiotemporal image brightness derivatives
- u is the horizontal optical flow
- v is the vertical optical flow

By assuming that the optical flow is smooth over the entire image, the Horn-Schunck method computes an estimate of the velocity field, $[u \quad v]^T$ which minimizes Equation 3 [13]:

$$E = \iint \left(I_x u + I_y v + I_t\right)^2 dxdy + \alpha \iint \left\{\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2\right\} dxdy \qquad (3)$$

In Equation 3, $\frac{\partial u}{\partial x}$ and $\frac{\partial u}{\partial y}$ are the spatial derivatives of the optical velocity component u, and α scales the global smoothness term [13]. The Horn-Schunck method minimizes the previous equation to obtain the velocity field, $[u\ v]$, for each pixel in the image, which is given by the Equations 4 and 5 [13]:

$$u_{x,y}^{k+1} = \bar{u}_{x,y}^{k} - \frac{I_x\left[I_x\,\bar{u}_{x,y}^{k} + I_y\,\bar{v}_{x,y}^{k} + I_t\right]}{\alpha^2 + {I_x}^2 + {I_y}^2} \qquad (4)$$

$$v_{x,y}^{k+1} = \bar{v}_{x,y}^{k} - \frac{I_y\left[I_x\,\bar{u}_{x,y}^{k} + I_y\,\bar{v}_{x,y}^{k} + I_t\right]}{\alpha^2 + {I_x}^2 + {I_y}^2} \qquad (5)$$

In Equation 4 and 5, $\left[u_{x,y}^{k} \quad v_{x,y}^{k}\right]$ is the velocity estimate for the pixel at (x, y), and $\left[\bar{u}_{x,y}^{k} \quad \bar{v}_{x,y}^{k}\right]$ is the neighborhood average of $\left[u_{x,y}^{k} \quad v_{x,y}^{k}\right]$. For k=0, the initial velocity is 0.

The Horn-Schunck method, u and v are solved as follows [16]:

1. Compute I_x and I_y using the Sobel convolution kernel: $\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$ and its transposed form for each pixel in the first image.
2. Compute I_t between images 1 and 2 using the $[-1 \quad 1]$ kernel.
3. Assume the previous velocity to be 0, and compute the average velocity for each pixel using $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ as a convolution kernel.
4. Iteratively solve for u and v.

4.4. Thresholding and Region Filtering

The simple process of conversion from intensity images to black and white images is thresholding. [17] described segmentation using thresholding. It said thresholding is the simplest segmentation method. The pixels are partitioned depending on their intensity value. There are several methods of thresholding as written [17], among others:

- Global thresholding, using an appropriate threshold T [17]:

$$g(x, y) = \begin{cases} 1, & \text{if } f(x, y) > T \\ 0, & \text{if } f(x, y) \le T \end{cases} \qquad (6)$$

- Multiple thresholding [17]:

$$g(x,y) = \begin{cases} a, & \text{if } f(x,y) > T_2 \\ b, & \text{if } T_1 < f(x,y) \le T_2 \\ c, & \text{if } f(x,y) \le T_1 \end{cases} \quad (7)$$

- Variable thresholding, if T can change over the image [17].
  - Local or regional thresholding, if T depends on a neighborhood of (x, y).
  - Adaptive thresholding, if T is a function of (x, y).

Peaks and valleys of the image histogram, as an example Figure 7, can help in choosing the appropriate value for the threshold. Some factors affect the suitability of the histogram for guiding the choice of the threshold [17]:
- The separation between peaks
- The noise content in the image
- The relative size of objects and background
- The uniformity of the illumination
- The uniformity of the reflectance

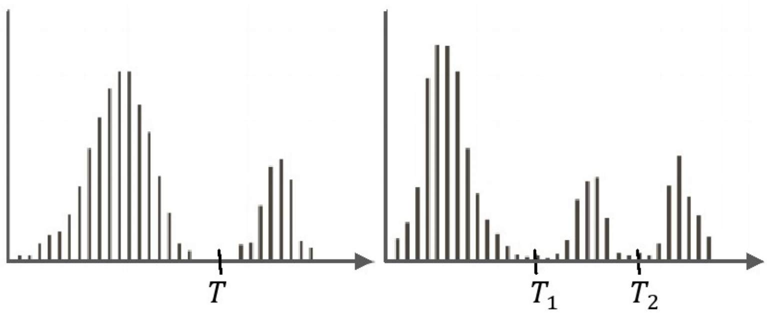

Figure 7. Image histogram [17]

Thresholding can cause noise in an image. Median filter is widely used and very effective at removing noise while preserving edges. The median filter works by moving through the image pixel by pixel, replacing each value with the median value of neighboring pixels. The pattern of neighbors is called the "window", which slides, pixel by pixel over the entire image pixel, over the entire image. Thresholding also cause imperfections shape of the object.

To improve the shape of the image used Morphological filters. Erosion and dilation is a morphological operation that will reduce and add pixels respectively on to the boundary between objects in a digital image. The erosion of the binary image A by the structuring element B is defined by Equation 8 [22]:

$$A \ominus B = \{z \in E \mid B_z \subseteq A\} \quad (8)$$

Where B_z is the translation of B by the vector z, i.e. Equation 9 [22]:

$$B_z = \{b + z \mid b \in B\}, \forall z \in E \quad (9)$$

The dilation can be obtained by the Equation 10 [22]:

$$A \oplus B = \{z \in E \mid (B^s)_z \cap A \neq \emptyset\} \quad (10)$$

Where B^s denotes the symmetric of B that is an Equation 11 [22]:

$$B^s = \{x \in E \mid -x \in B\} \quad (11)$$

Closing filter is one of morphological filter process for removing small holes. The closing of a binary image A by a structuring element B is the erosion of the dilation of that set such as Equation 12 [22]:

$$A \bullet B = (A \oplus B) \ominus B \quad (12)$$

Where $\oplus$ and $\ominus$ denote the dilation and erosion, respectively.

The next process is region filtering using blob analysis. Starting with estimate the area and bounding box, hereinafter referred to as bbox, of the blobs in the thresholded image. Select these boxes which are in the region of interest (ROI). When the ratio between the area of the blob and the area of the bounding box is above R (are expressed as a percentage) the bounding box for that object is used. Otherwise the bounding box is removed.

As described in the Matlab Documentation, the area represents the number of pixels in labeled regions. Bounding box (bbox) is an M-by-4 matrix of [*x y width height*] bounding boxes. The rows represent the coordinates of each bounding box, where M represents the number of blobs. Example, suppose there are two blobs, where x and y define the location of the upper-left corner of the bounding box, and w, h define the width and height of the bounding box. The block outputs at the bbox port $\begin{bmatrix} x_1 & y_1 & w_1 & h_1 \\ x_2 & y_2 & w_2 & h_2 \end{bmatrix}$

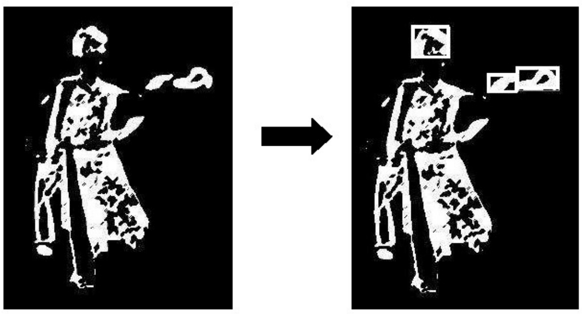

Figure 8. Region filtering

In Figure 8, the left image is the threshold image and the right shows the results of region filtering process using blob analysis method with the three bounding boxes in the head and scarf.

4.5. Motion Estimation

Not all frames have bbox. Each frame is not necessarily having the same number of bbox. Each bbox allocates dancers moving limbs. Bbox frame F_t relate to the bbox frame F_{t+1} if it has the shortest distance (less than or equal 20.0). Distance is calculated using the Euclidian method.

There are several kinds of bbox appearance in a sequence of frames as shown in Table 1. If limb's bbox appear (denoted 1) in the frame sequence that means a limb in the bbox perform a series of motions. If bbox of llimb does not appear (denoted 0) in the next frame or previous frame, it means there is no motion or change other limbs that moves.

Table 1. Appearance of bbox

F_{t-1}	F_t	F_{t+1}
0	0	0
1	1	1
1	0	0
1	0	1
1	1	0
0	1	0
0	1	1
0	0	1

To determine whether the limb in bbox moving or not, if it moves to where the direction of movement can be estimated using the differential method. There are eight kinds of motion estimation such as Figure 9.

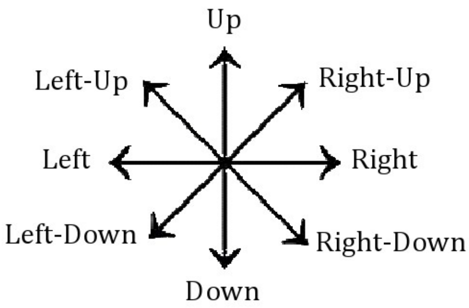

Figure 9. Motion direction

As explained previously that every bbox in a frame having information [x y width height]. The motion direction of bbox in F_t to F_{t+1} determined by calculating the value of dx_{t+1} and dy_{t+1} by means $x_{t+1} - x_t$ and $y_{t+1} -$

y_t respectively. Motion direction vertical-horizontal or diagonal is determined by the ratio r value of dx_{t+1} and dy_{t+1} as (13).

By using the values of x, y, dx_{t+1}, dy_{t+1}, and r can be calculated coordinates x'_{t+1} and y'_{t+1} which can be estimated the motion direction as shown Table 2.

$$r = \begin{cases} \dfrac{dx_{t+1}}{dy_{t+1}} & ;\ dy_t > dx_t \text{ and } dy_t \neq 0 \\ \dfrac{dy_{t+1}}{dx_{t+1}} & ;\ dy_t > dx_t \text{ and } dy_t \neq 0 \\ 0 & ;otherwise \end{cases} \tag{13}$$

Comparing the value to x'_{t+1} with x'_t and y'_{t+1} with y'_t as in Table 3 can be estimated direction of movement. The last line of Table 3 are expressed $m_{t+1} = m_t$ if all of the criteria in the previous line is not fulfilled, then the motion estimation is considered the same as the previous frame.

Table 2. Coordinate estimation of x'_{t+1} and y'_{t+1}

Ratio	**Condition**	x'_{t+1}	y'_{t+1}
$\geq r$	$dx_{t+1} \geq dy_{t+1}$	$x'_t + dx_{t+1}$	$y'_t + dx_{t+1}$
	$dx_{t+1} < dy_{t+1}$	$x'_t + dy_{t+1}$	$y'_t + dy_{t+1}$
$< r$	$dx_{t+1} \geq dy_{t+1}$	$x'_t + dx_{t+1}$	y'_t
	$dx_{t+1} < dy_{t+1}$	x'_t	$y'_t + dy_{t+1}$

Table 3. Motion estimation

Ratio	**Condition**	**Motion estimation (m_{t+1})**
$< r$	$y'_{t+1} < y'_t$	↑ UP
	$y'_{t+1} > y'_t$	↓ DOWN
	$x'_{t+1} < x'_t$	← LEFT
	$x'_{t+1} > x'_t$	→ RIGHT
$\geq r$	$x'_{t+1} > x'_t$ and $y'_{t+1} < y'_t$	↗ RIGHT-UP
	$x'_{t+1} > x'_t$ and $y'_{t+1} > y'_t$	↘ RIGHT-DOWN
	$x'_{t+1} < x'_t$ and $y'_{t+1} < y'_t$	↖ LEFT-UP
	$x'_{t+1} < x'_t$ and $y'_{t+1} > y'_t$	↙ LEFT-DOWN
Otherwise		m_t

5. EXPERIMENT AND ANALYSIS

From the entire data of the video capture is written in 4.1, the first 30 seconds of samples taken with the number of frame 752. Figure 10 from left to right and top to bottom respectively describes the results of the motion detection process from frame 25, i.e. original image, intensity image, the optical flow result, threshold region detection image, motion detection image, and motion vector image. The process begins with conversion from RGB images into an intensity that produces a grayscale image. Then make the process of optical flow using Horn-Schunck algorithm.

The Horn–Schunck method of estimating optical flow is a global method which introduces a global constraint of smoothness to solve the aperture problem. This method was chosen because it works in all areas of the image and generates the motion vectors with high density. This algorithm assumes smoothness in the flow over the whole image. Thus, it tries to minimize distortions in flow and prefers solutions which show more smoothness [13].

For comparison, Lucas-Kanade method assumes that the flow is essentially constant in a local neighborhood of the pixel under consideration, and solves the basic optical flow equations for all the pixels in that neighborhood, by the least squares criterion. Since it is a purely local method, it cannot provide flow information in the interior of uniform regions of the image [14].

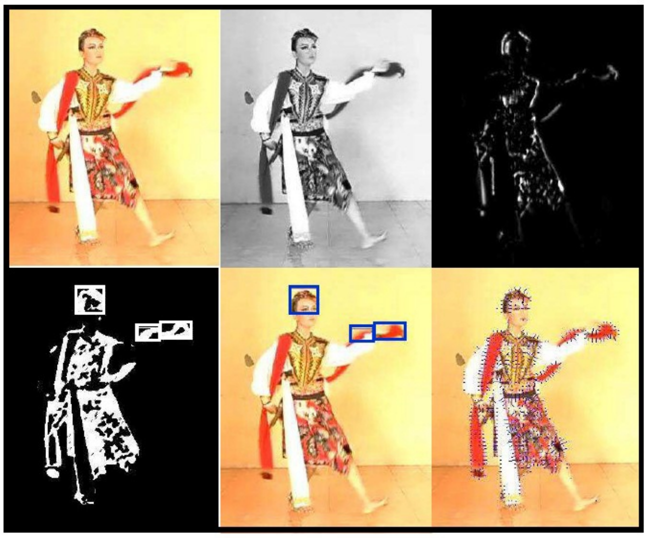

Figure 10. Motion detection process result

The leftmost image in the second row is the result of filtering region that begins with thresholding process. Thresholding cause noise and damage that repaired with Median filter and Morphological filter. Regions filter using blob analysis method. The following R value used in the experiment and its influence in the motion tracking.

As shown in Table 4, R value will determine the amount of bbox generated from a video file. For smaller R value, generate a greater number of bbox, and vice versa. But this does not mean that more the number of bbox be better, because it could be the same limb are at different bbox series. Conversely a smaller number of bbox resulting smaller number of limbs detected. The greater number of bbox causes slower motion estimation process. Great value ratio causes more frames that do not have bbox. Consequently bbox series of sequential frames becomes disconnected.

The next process is to estimate the bbox coordinates for F_{t+1} based on the series of bbox coordinate in F_t using the differential method. Distance or accuracy of the estimated coordinates of F_{t+1} with the F_t is determined by the ratio between the differential value dx and dy.

Table 4. Region filter

R	**Number of bbox**	**Frames having bbox**	**Frames without bbox**
0.35	155	591	161
0.4	120	559	193
0.45	80	492	260
0.5	55	379	373
0.55	36	243	509

Table 5. Error value frame 47 to 72

Frame	**Error value with different *r***									
	0.25		0.375		0.5		0.625		0.75	
	x'	y'	x'	y'	x'	y'	x'	y'	x'	y'
47	0	0	0	0	0	0	0	0	0	0
48	0	2	0	2	0	2	0	-2	0	-2
49	0	4	0	4	0	4	0	-5	0	-5
50	0	3	0	3	0	3	0	-6	0	-6
51	0	2	0	2	0	2	0	-7	0	-7
52	0	1	0	1	0	1	0	-8	0	-8
53	0	0	0	0	0	0	0	-9	0	-9
54	0	-1	0	-1	0	-1	0	-10	0	-10
55	0	2	0	-2	0	-2	0	-11	0	-11
56	0	5	0	-3	0	-3	0	-12	0	-12
57	0	7	0	-4	0	-4	0	-13	0	-13
58	0	9	0	-5	0	-5	0	-14	0	-14
59	0	9	0	-5	0	-5	0	-14	0	-14
60	0	10	0	-4	0	-4	0	-15	0	-15
61	-1	10	-1	-4	-1	-4	-1	-15	2	-15
62	-3	10	0	-4	0	-4	0	-15	3	-15
63	-3	10	0	-4	0	-4	0	-15	3	-15
64	-4	10	-1	-4	-1	-4	1	-15	4	-15
65	-4	10	-1	-4	-1	-4	1	-15	4	-15

Frame	Error value with different *r*									
	0.25		0.375		0.5		0.625		0.75	
	x'	y'	x'	y'	x'	y'	x'	y'	x'	y'
66	-4	10	-1	-4	-1	-4	1	-15	4	-15
67	-4	10	-1	-4	-1	-4	1	-15	4	-15
68	-3	10	0	-4	0	-4	0	-15	3	-15
69	-3	10	0	-4	0	-4	0	-15	3	-15
70	-3	10	0	-4	0	-4	0	-15	3	-15
71	-3	10	0	-4	0	-4	0	-15	3	-15
72	-3	10	0	-4	0	-4	0	-15	3	-15

Table 5 shows the value of the error between original coordinates and estimation coordinates by the value of *r* that used is 0.25, 0.375, 0.5, 0.625, 0.75. The *r* value is used to determine whether motion estimation for limb in the next frame to move vertically, horizontally, or diagonally.

The above table contains data error values of motion estimation with different *r* for bbox number 7 (head) on the optical flow *R* value 0.4. From 752 frames there were 56 series of head motion. One is the data of Table 5. Motion coordinate and motion charts of head motion between original motion and estimate motion from frame 47 to frame 72 is illustrated in Figure 11. In that figure, the green color represents the original motion chart and red color represents estimate motion chart. The shape of the chart represents motion of the head.

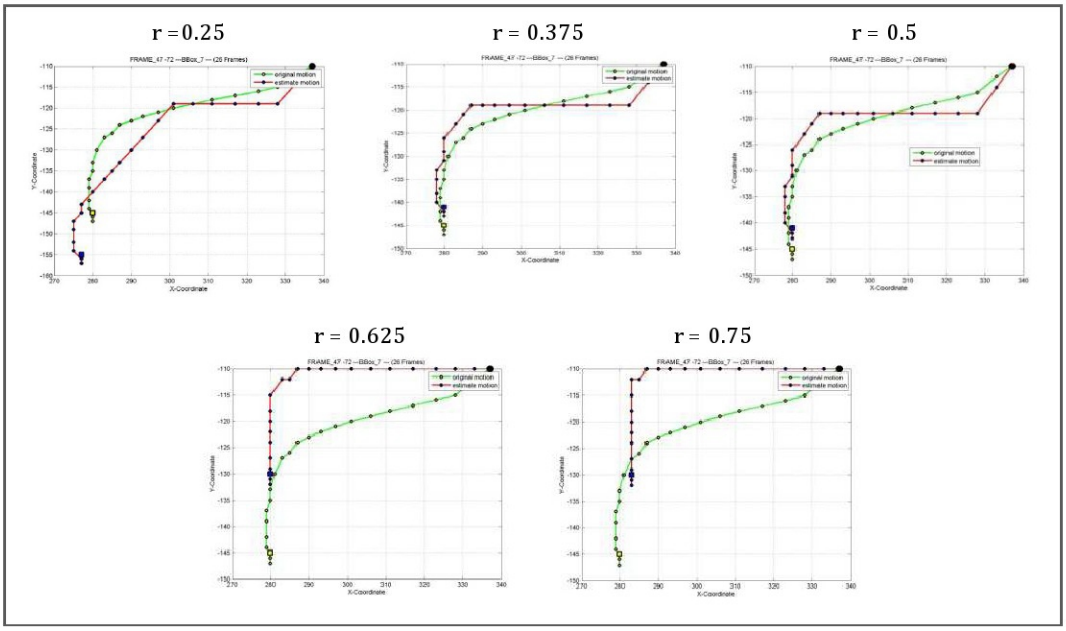

Figure 11. Motion estimation charts

Of the five charts with different *r* value it appears that *r* value 0.375 and r value 0.5 produces a similar chart between the original motion and estimate motion. Such a situation also occurs in 55 other series of head motion.

Table 6. Error value frame 149 to 153

Frame	Error value with different *r*									
	0.25		0.375		0.5		0.625		0.75	
	x'	y'	x'	y'	x'	y'	x'	y'	x'	y'
149	0	0	0	0	0	0	0	0	0	0
150	0	0	0	0	0	0	0	0	0	0
151	0	0	0	0	0	0	0	0	0	0
152	0	0	0	0	0	0	0	0	0	0
153	0	0	0	0	0	0	0	0	0	0

Estimate coordinates are not necessarily like original coordinates, but if the chart of estimate motion similar to the chart of the original motion, it means motion estimation it can be said to have the same motion with the original. The smaller number of frames allows the same motion and the same coordinate between the original motion and the estimation motion as shown in Table 6.

6. CONCLUSION

Remo dance is an Indonesian traditional dance originating from the province of East Java that have unique motions. Limbs motions are partial behavior. This means that all limbs move alternately. Remo Dancer wears a few costume pieces and accessories. Therefore, we need a simple markerless method to detect the motion of Remo Dancer. That method is Optical Flow with Horn-Schunck algorithm. The advantages of this method do not have to use the monochrome background and not require complex mathematical calculations, so that it can be applied in real time.

Performance of motion detection with optical flow method is determined by the value of the ratio between the area of the blob and the area of the bounding box. Here, using the value of 0.4. The smaller ratio value causes a greater number of ROI, vice versa. Motion estimation using the differential method. Distance or accuracy of the estimated coordinates and motion estimation is determined by the ratio between the differential value *dx* and *dy*. Based on the experiments that have been done obtained value ratio 0.375 to 0.5 produces motion estimation closer to the original motion.

Further research the motion estimation used in the motion capture process to mimic the motion of Remo Dance by human body model.

Acknowledgements

The author would like to thank the Electronic Engineering Polytechnic Institute of Surabaya especially to the supervisors for all the knowledge and guidance that has been given. Thanks also to the Surabaya Polytechnic, which has given more educational opportunities for author and all those who have helped up to the completion of this research.

REFERENCES

[1] Machdalati Rizky Cendani. **Buku Visual Tari Remo Surabayan Sebagai Media Pendukung Hak Paten Kesenian Khas Surabaya**. ITS-paper-22017-3407100048. 2012.

[2] Jacky C.P. Chan, H. Leung, Jeff K.T. Tang, and T. Komura. **A Virtual Reality Dance Training System Using Motion Capture Technology**. *IEEE Transactions on Learning Technologies*, Vol. 4, no. 2. 2011.

[3] LaViers and M. Egerstedt. **The Ballet Automaton: A Formal Model for Human Motion**. *IEEE/ACM International Conference on Cyber-Physical Systems - ICCPS*, 2011.

[4] N. Magnenat-Thalmann, D. Protopsaltou, and E. Kavakli. **Learning How to Dance Using a Web 3D Platform**. *Proc. Sixth Int'l Conf. Web-Based Learning*, pp. 1-12, 2007.

[5] Saboune, J., Charpillet, F. **Markerless Human Motion Capture for Gait Analysis.** *Clinical Orthopaedics and Related Research. 3rd European Medical and Biological Engineering Conference*. 2005.

[6] Gary Bradski, Adrian Kaehler. **Learning OpenCV - Computer Vision with the OpenCV Library**. O'Reilly Media, Inc. Ed. 1. 2008

[7] Richard Szeliski. **Computer Vision: Algorithms and Applications**. Springer. draft. 2010.

[8] Thomas Brox, Bodo Rosenhahn, Daniel Cremers. **Contours, Optic Flow, and Prior Knowledge: Cues for Capturing 3D Human Motion in Videos**. Springer. Volume 36. pp 265-293. 2008.

[9] Sophia Bakogianni, Evangelia Kavakli, Vicky Karkou, Maroussa Tsakogianni. **Teaching Traditional Dance Using E-learning Tools : Experience from The WebDANCE Project**. *Proceedings DVD of the 21st World Congress on Dance Research*, Athens. 2007.

[10] Magnenat-Thalmann, N., Joslin, C. **Learning How to Dance on the Internet**. *In: Interface Conference*. Hamburg. 2000.

[11] Thomas Brox, Bodo Rosenhahn, Daniel Cremers, Hans-Peter Seidel. **High Accuracy Optical Flow Serves 3-D Pose Tracking: Exploiting Contour and Flow Based Constraints**. Springer. Volume 3952. pp 98-111. 2006.

[12] Ballan, L., Cortelazzo, G.M. **Markerless Motion Capture of Skinned Models in a Four Camera Set-Up Using Optical Flow and Silhouettes.** In: 3DPVT. 2008.

[13] B.K.P. Horn and B.G. Schunck. **Determining Optical Flow**. *Artificial Intelligence. vol 17*, pp 185–203. 1981

[14] B. D. Lucas and T. Kanade. **An Iterative Image Registration Technique with an Application to Stereo Vision**. *Proceedings of Imaging Understanding Workshop*, pages 121—130. 1981.

[15] Bruce D. Lucas. **Generalized Image Matching by the Method of Differences.** Doctoral Dissertation. 1984.

[16] Barron, J.L., D.J. Fleet, S.S. Beauchemin, and T.A. Burkitt. **Performance of Optical Flow Techniques**. CVPR. 1992.

[17] Stefano Ferrari. **Image segmentation**. Universit`a degli Studi di Milano. 2012.
[18] Swamidoss Sathiakumar, Lalit Kumar Awasthi, Roberts Masillamani, S.S. Sridhar. **Advances in Intelligent Systems and Computing**. ICICIC Global. 2012.
[19] Rujuta R Mahambare. **Converting Grayscale Image to Color Image**. *International Journal of Computer, Information Technology & Bioinformatics*. Volume-1, Issue-4. 2013.
[20] Stanley Baron and David Wood. **Rec. 601 - The Origins of The 4:2:2 DTV Standard**. Ebu Technical Review. 2005.
[21] Ashok Banerji, Ananda Mohan Ghosh. **Multimedia Technologies**. Tata McGrawHill. 2010.
[22] Harley R. Myler, Atrhur R. Weeks. **The Pocket Handbook of Image Processing Algorithms in C.** Prentice Hall. 1993.