

# Motion Estimating Optical Flow for Action Recognition

(FARNEBACK, HORN SCHUNCK, LUCAS KANADE AND LUCAS-KANADE DERIVATIVE OF GAUSSIAN)

Rosepreet Kaur Bhogal, V Devendran<sup>#</sup>

*Lovely Professional University*

Phagwara, India

[rosepreetkaur12@gmail.com](mailto:rosepreetkaur12@gmail.com), svdevendran@gmail.com<sup>#</sup>

**Abstract**— Motion estimating is one of the methods which determines the movement from one frame to another in the videos. For an application of action recognition, choosing the optical flow can be an essential feature for recognizing actions. The optical flow consists of the information of the moving subject and objects in the video frames. This paper analyzes four motion estimating optical flow methods (Farneback, Horn Schunck, Lucas Kanade, and Lucas-Kanade Derivative of Gaussian explored based on visualization and PSNR. The NTURGB+D dataset uses for the analysis of experimental results.

**Keywords**—Optical Flow, Action Recognition, Videos, NTURGB+D, Motion Estimation

## I. INTRODUCTION

Human action recognition is one of the most challenging tasks nowadays. Various methods have been proposed in [1], which can use for pre-processing and recognizing activities. Motion estimation features are one of the main steps required to know activities in video sequences. The optical estimation technique gives in [2]. The optical flow gives motion information and helps to find an interesting point used to recognize actions. There are various non-parametric methods in which motion descriptor calculation helps estimate flow between one and the next frame. The flow descriptors aggregate the histograms on the temporal axis. The gradient-based methodology improves action recognition tasks [3]. There are various other handcrafted features like SIFT, HOG, GIST, and MHI [4]. These features have more focus on spatial information but not that much on temporal characteristics. The human motion analysis features the identification of human body shape, the relation of motion from one frame to the next, and one aspect of human activity recognition [5]. The optical flow based on warping explores which provided high-accuracy in [6]. This method can reduce the angle errors while computing optical flow. The optical flow can be used for video object segmentation, as mentioned in [7]. This model is attention based, which can use for object detection. Human activity recognition for video surveillance can design using an optical flow feature. Optical flow features include information related to the movement of subjects [8].

The mainly used optical flow estimation algorithm explore and analyse in the paper. Section 2 includes the research methodology. Section 3 includes experimental works, which show analysis based on a motion by considering random 15 RGB videos of NTURGB+D. Further, the conclusion of the

paper is which algorithm can prefer by the researcher for the estimation of activity.

## II. METHODOLOGY

To calculate optical flow, techniques analyses in this paper. The methods are Farneback [9], Horn Schunck [10], Lucas Kanade [11], and Lucas-Kanade Derivative of Gaussian [11].

The HOF features are widely used for optical flow estimation by differential equations. Despite calculated differences, the measurement can include three stages of processing. First, pre-filtering to extract interest points. Second, compute the spatial-temporal derivative (called velocity vectors). Third, the integration of measurements to produce flow fields [12]. The methods used to find the object's direction and object's speed of moving from one frame to another.

For computation of the optical flow features can use the equation as follows::

$$I_x u + I_y v + I_t = 0 \quad (1)$$

In equation,  $I_x$ ,  $I_y$  and  $I_t$  are the brightness derivatives.

### A. Algorithm 1: Horn Schunck [11]

The method estimates a velocity vector by considering the flow features are smooth across the whole image.

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

In the above equation 1, the spatial derivative of optical velocity components is  $\frac{\partial u}{\partial x}$  and  $\frac{\partial u}{\partial y}$ ,  $\alpha$  is the scaling factor that is related to smoothness. The method used to minimize equation 2 for each pixel in the frame with equations 3 & 4 is as follows:

$$u_{x,y}^{k+1} = u_{x,y}^{-k} - \frac{I_x [I_x u_{x,y}^{-k} + I_y v_{x,y}^{-k} + I_t]}{\alpha^2 + I_x^2 + I_y^2} \quad (3)$$

$$v_{x,y}^{k+1} = v_{x,y}^{-k} - \frac{I_y [I_x u_{x,y}^{-k} + I_y v_{x,y}^{-k} + I_t]}{\alpha^2 + I_x^2 + I_y^2} \quad (4)$$

The velocity estimator for pixel  $(x, y)$  is  $[u_{x,y}^k \text{ and } v_{x,y}^k]$  and neighborhood average of  $[u_{x,y}^k \text{ and } v_{x,y}^k]$  is  $[u_{x,y}^{-k} \text{ and } v_{x,y}^{-k}]$ . The initial velocity is zero. The steps to solve velocity vectors are as steps:

Step 1: Figuring of the brightness derivative i.e.  $I_x$  and  $I_y$  with Sobel kernels and its transposed form for each pixel in the first image.

Step 2: Computation of  $I_t$ , i.e. difference in first and second image using  $[-1 \ 1]$  kernel which basically used to determine the difference or change.

Step 3: Computation of the velocity average of each image pixel using  $[0 \ 1 \ 0; 1 \ 0 \ 1; 0 \ 1 \ 0]$  as convolution kernel.

Step 4: Iteratively solving of  $u$  and  $v$ .

#### B. Algorithm 2: Lucas Kanade [11]

To solve horizontal  $u$  and vertical  $v$  optical flow features, the LUCAS KANADE divides one frame into parts, assuming that each region has constant velocity. The method used weighted least square fit by minimizing the equation as follows:

$$\sum_{x \in \varphi} W^2 [I_x u + I_y v + I_t]^2 \quad (5)$$

In equation 5, “W” is the window function that focuses the constraints at each section's center. The final solution to minimizing the part is

$$\begin{bmatrix} \sum W^2 I_x^2 & \sum W^2 I_x I_y \\ \sum W^2 I_x I_y & \sum W^2 I_y^2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} W^2 I_x I_t \\ W^2 I_y I_t \end{bmatrix} \quad (6)$$

Step 1: Figuring the  $I_x$  and  $I_y$  by using the kernel  $[-1 \ 8 \ 0 \ -8 \ 1]/12$  with its transposed form.

Step 2: Computation of  $I_t$ , i.e, the difference of images 1 and 2 using the  $[-1 \ 1]$  kernel.

Step 3: After, step 2 required to smooth the gradient components  $I_x$ ,  $I_y$  and  $I_t$  by using 1D kernel.

Step 4: By solving of the 2-by-2 linear equations for each pixel and compare the eigenvalues with the threshold to calculate the  $u$  and  $v$ .

#### C. Algorithm 3: Lucas-Kanade Derivative of Gaussian [11]

The Lucas Kanade method computes  $I_t$  using the gaussian filter. For solving the Optical flow constraint equation for  $u$  and  $v$ .

Step 1: For computing  $I_x$  and  $I_y$  step follows as below:

- a) Perform temporal filtering by using the Gaussian filter.
- b) Perform spatial filtering by using the gaussian filter's derivative for smoothing each frame.

Step 2: Computation of the difference,  $I_t$  between images 1 and 2 using the  $[-1 \ 1]$  kernel

- a) Perform temporal filtering by using the derivative of the gaussian filter.
- b) Perform spatial filtering by using filter Step 1 (b) on the output of temporal filtering.

Step 3: Using a gradient smoothing filter, smooth gradient components  $I_x$ ,  $I_y$  and  $I_t$ .

Step 4: By solving of the 2-by-2 linear equations for each pixel and compare the eigenvalues with the threshold to calculate the  $u$  and  $v$ .

#### D. Algorithm 4: Farneback (based on polynomial expansion)

The Farneback algorithm [9] finds the displacement from low resolution to high resolution. It can compare with a pyramid structure, where optical flow estimation starts from one-point convergence and goes to many motion points convergences. The Farneback algorithm finds displacement with the polynomial expansion method. The polynomial uses to approximate the neighborhood of each pixel. The expression of the local coordinate system is as follows:

$$f(x) = x^T A x + b^T x + c \quad (7)$$

Where “A” is a symmetric matrix, “b” is a vector, and “c” is a scalar quantity. The algorithm-related step to estimate displacement given in [9].

### III. EXPERIMENTAL WORK

#### A. Dataset

The dataset used in work is NTURGB+D. The dataset divides into three categories, i.e., medical condition-based actions, dailya actions, and joint actions (where more than two persons involve). There are 60 classes in the dataset from which results show videos from each category. The dataset has 56880 videos recorded using the Microsoft Kinect v2 sensor. RGB videos recorded in the provided resolution of 1920x1080. The classification of the dataset under three types is as show in Fig. 1.

The experiment shows 15 random videos of NTURGB+D. The category of actions with the name of videos in the dataset is show in table 1. The results show in two ways: on one video of class drinking optical flow visualization and second PSNR-based results in table 2. In Fig. 2, the video-6 frames used show that (a) is the first frame of video-1, (b) is the third frame of video-6, and (c) shows the difference between the two frames, which depicts which changes are there between frame 1 and frame 3. If there is no difference means the whole difference frame is black. Every optical flow algorithm gives a velocity vector in the X direction, a velocity vector in the Y direction, magnitude and orientation. Optical flow is the movement of brightness patterns in X and Y directions. It gives information about the movement of the object in one frame to another, which uses for object tracking and estimation of actions in videos. The discontinuities in the objects help to segment objects in a frame. The magnitude depicts the velocities of pixels, and orientation indicates the direction of objects.

The Optical Flow cannot determine at one point. There is always involvement of neighborhood elements. At one point, the brightness varies, but the patch has no motion [10]. The magnitude of frame 1, shows in Fig. 3 for Farneback, Horn Schunck, Lucas Kanade, and Lucas-Kanade Derivative of Gaussian. More edges identify using Farneback, and no boundaries were detected using the Lucas Kanade derivative

of Gaussian. In another two methods, Lucas Kanade is better as compared to Horn Schunck.

In Fig. 4, the orientation shows using all-optical flow ways. Where the finding orientations using Lucas Kanade

outperforming over the other methods. In Fig. 4 (a), there is no visualization of the subject, and the same is in Fig. 4 (b) and (c) also.

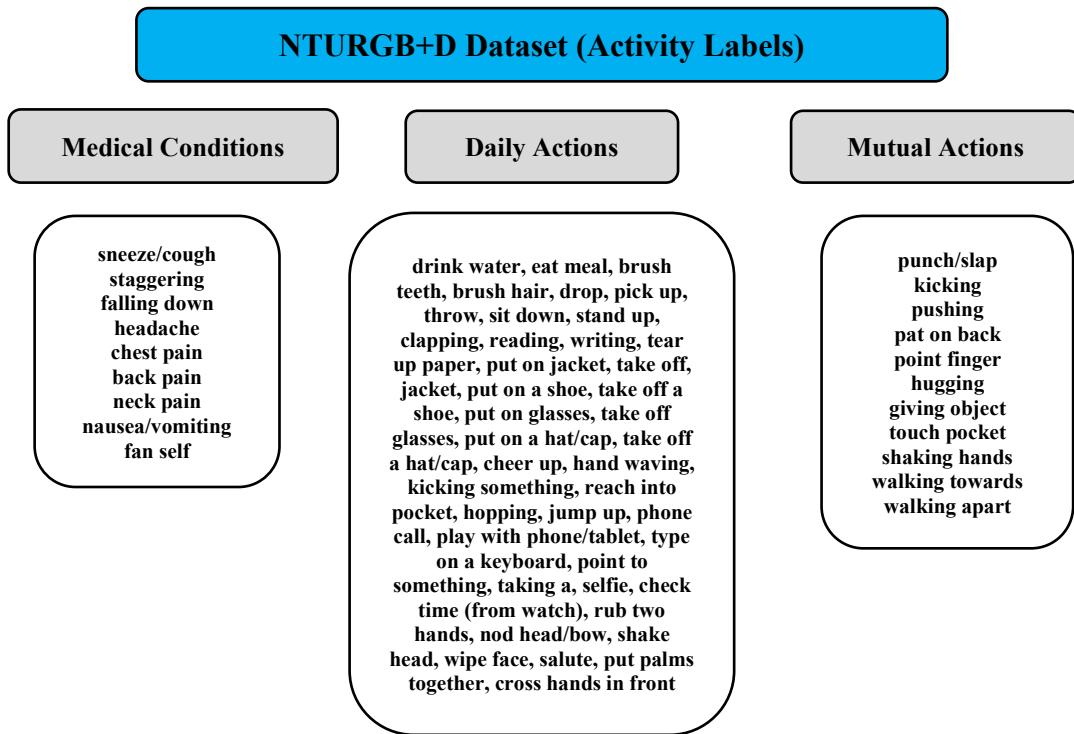


Fig. 1. The activity labels under three categories are medical conditions, Daily actions, and Mutual actions of NTURGB+D

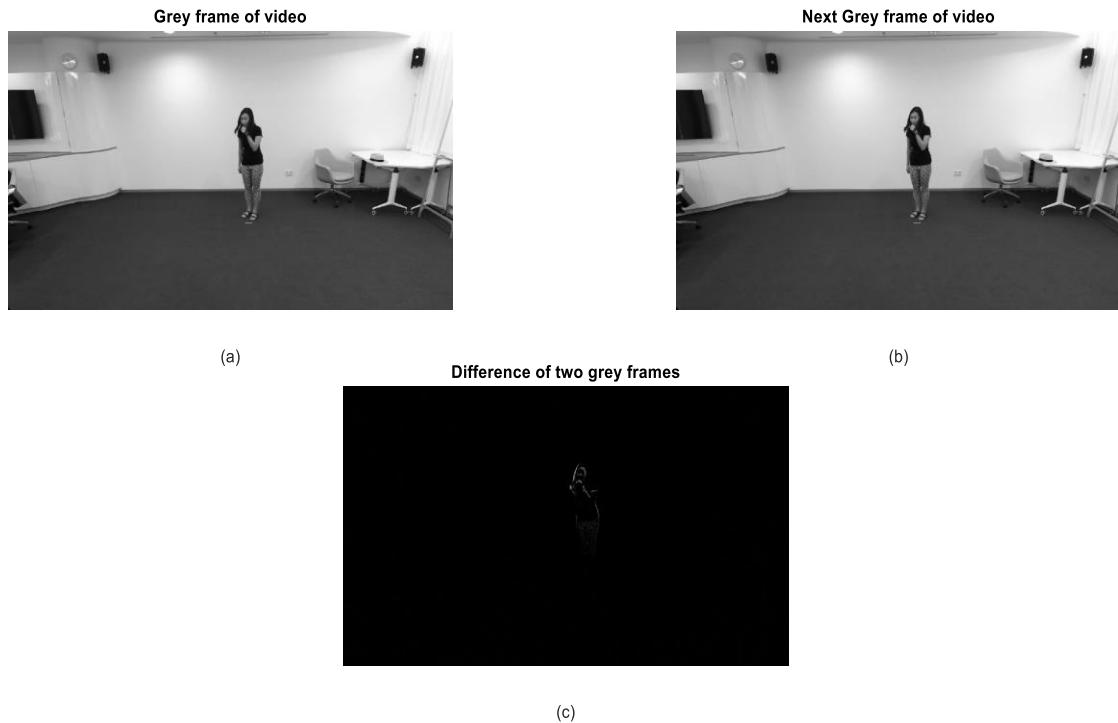


Fig. 2. (a) Video-1 (First frame), (b) Video-1 (Third frame), (c) Difference between two frames

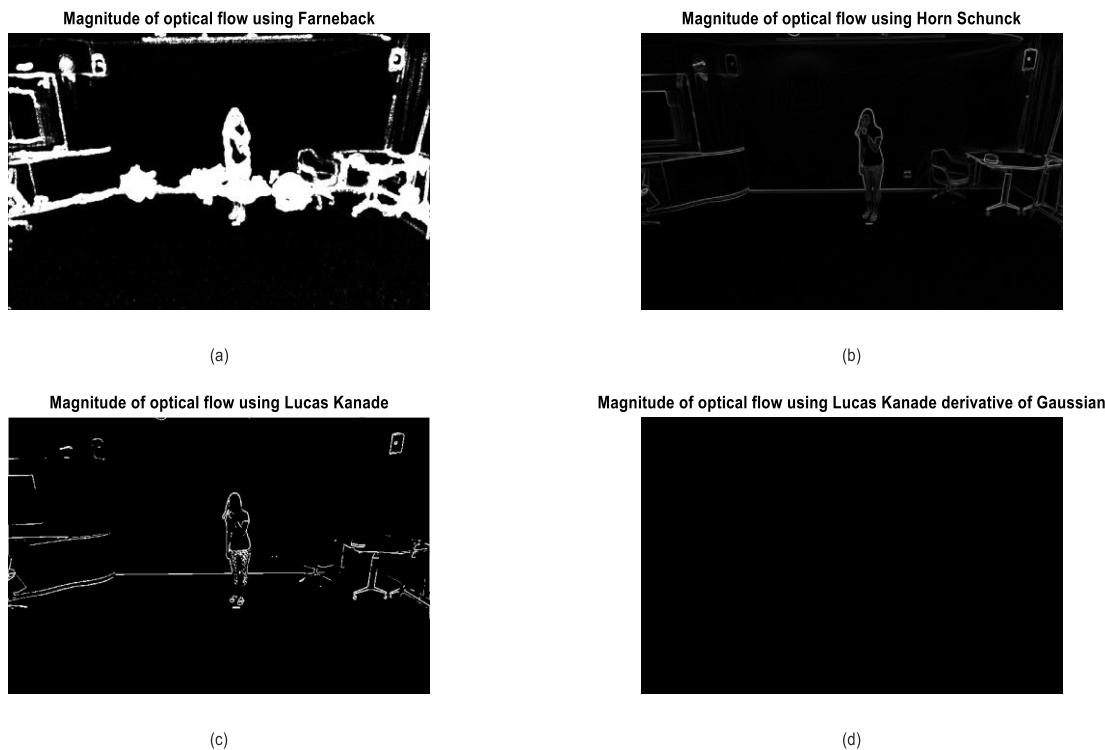


Fig. 3. Video-6 (Frame-1) magnitude of optical flow a) FARNEBACK, b) HORN SCHUNCK, c) LUCAS KANADE, and d) LUCAS-KANADE DERIVATIVE OF GAUSSIAN

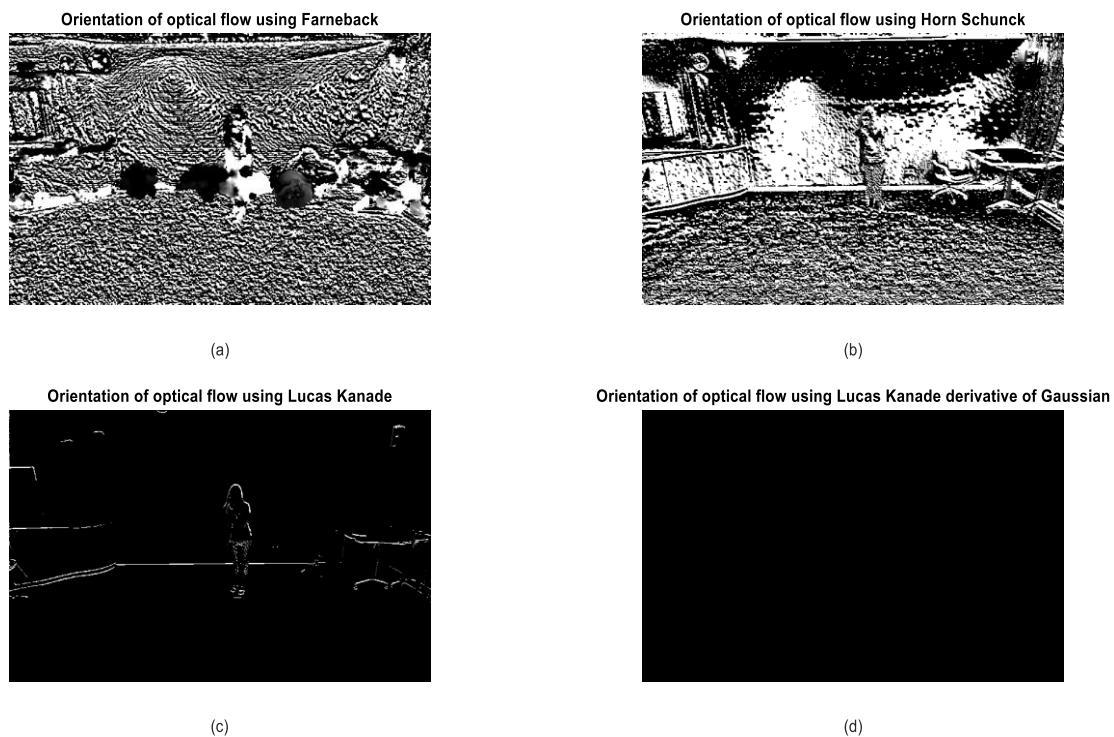


Fig. 4. Video-6 (Frame-1) orientation of optical flow a) FARNEBACK, b) HORN SCHUNCK, c) LUCAS KANADE, and d) LUCAS-KANADE DERIVATIVE OF GAUSSIAN

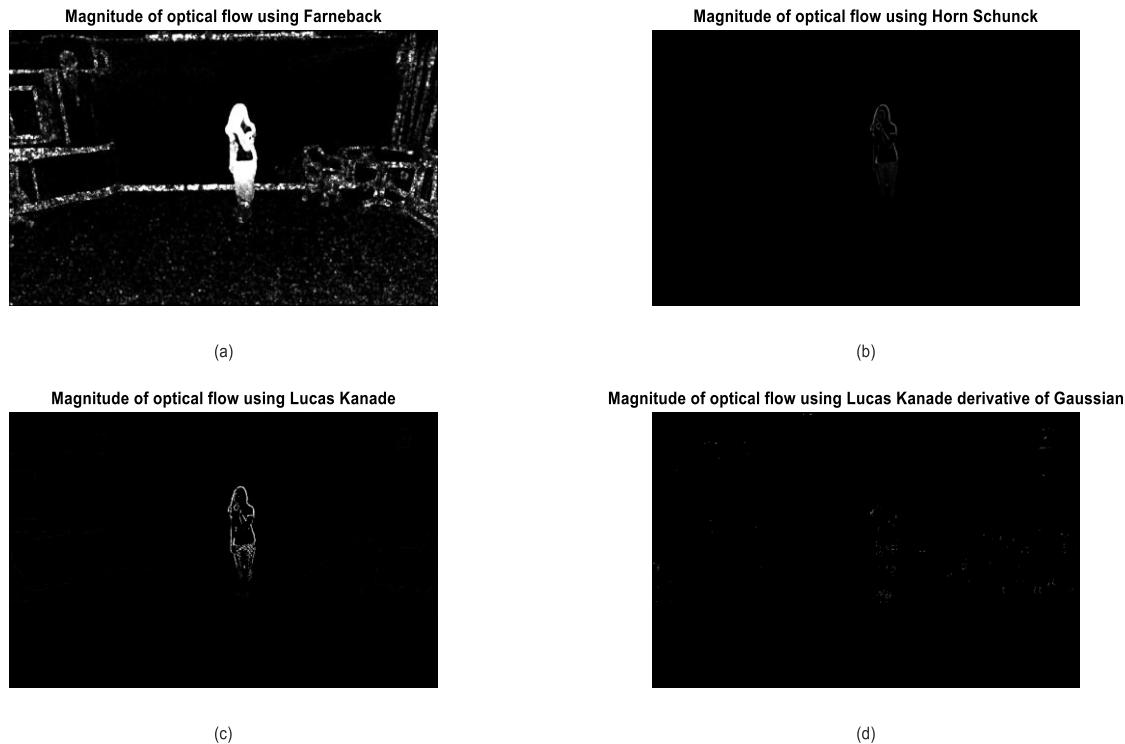


Fig. 5. Video-6 (Frame-3) magnitude of optical flow a) FARNEBACK, b) HORN SCHUNCK, c) LUCAS KANADE, and d) LUCAS-KANADE DERIVATIVE OF GAUSSIAN

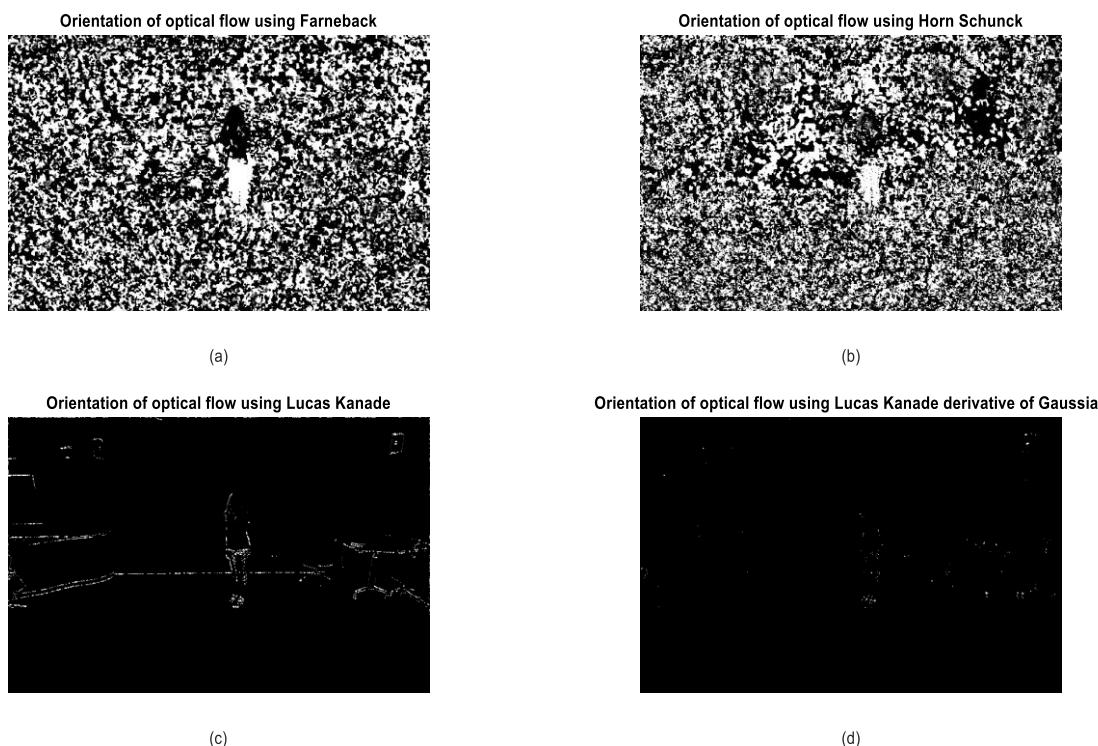


Fig. 6. Video-6 (Frame-3) orientation of optical flow a) FARNEBACK, b) HORN SCHUNCK, c) LUCAS KANADE, and d) LUCAS-KANADE DERIVATIVE OF GAUSSIAN

Fig. 5, with frame 3 of video-6 magnitude optical flow, shows that Lucas Kanade is better than other methods. Fig 5

(a) detected extra edges, which are not required, and in Fig. 5 (b) and (d), there is no visualization of subjects.

The orientation provides information on the directions of subjects. The orientation of video-6 (Frame 3) shows in Fig. 6, which depicts that there is no visualization in Fig. 6 (a), (b), and (c) instead of Fig. 6 (c).

Similarly, the magnitude and orientation of optical flow visualization shows in Fig. 7 and Fig. 8. Fig. 8 and 9 show the

difference between frame 1 and frame 3 of video 6. It shows a variation between frame 1 and frame 3 of video 6. And how the magnitude and orientation of optical flow affect frame-to-frame. From all four ways of optical flow, the Lucas Kanade is outperformed based on the frame and the difference of frames.

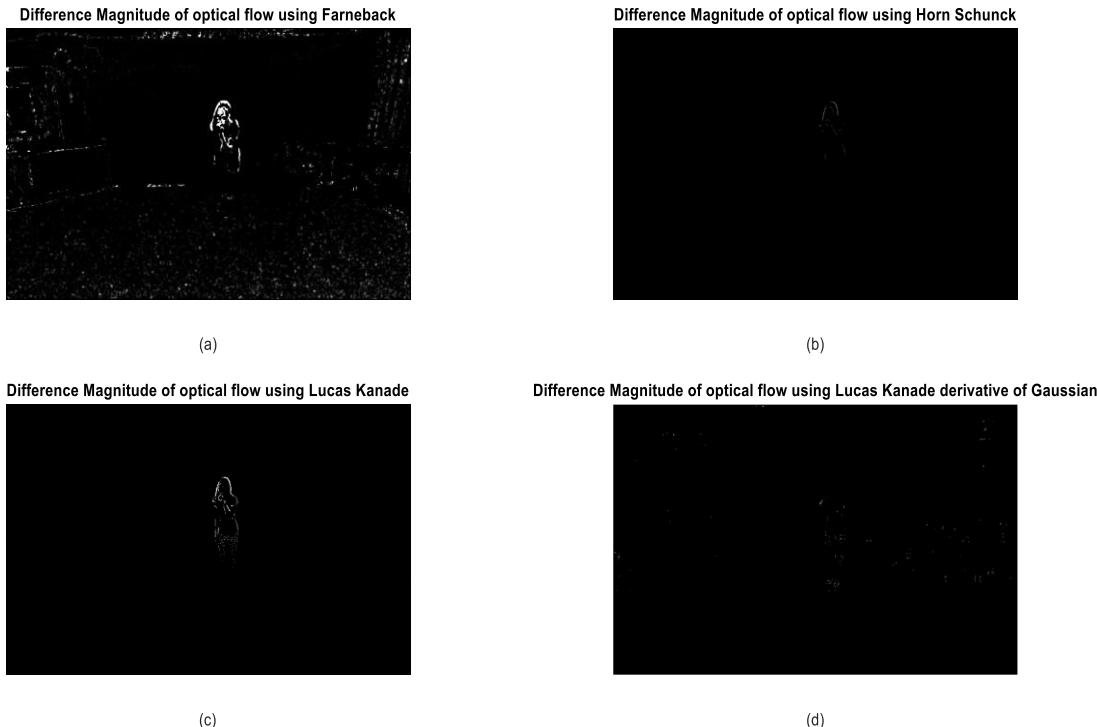


Fig. 7. Video-6 (difference of frame-3 and frame-1) magnitude of optical flow a) FARNEBACK, b) HORN SCHUNCK, c) LUCAS KANADE, and d) LUCAS-KANADE DERIVATIVE OF GAUSSIAN

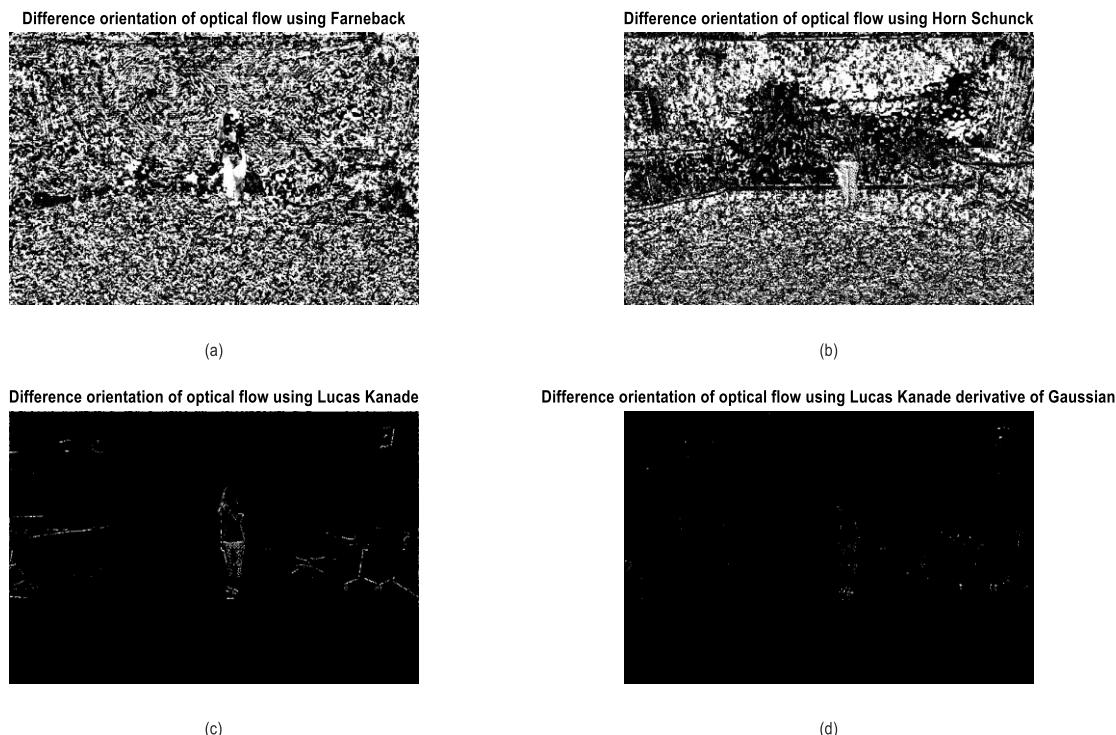


Fig. 8. Video-6 (difference of frame-3 and frame-1) orientation of optical flow a) FARNEBACK, b) HORN SCHUNCK, c) LUCAS KANADE, and d) LUCAS-KANADE DERIVATIVE OF GAUSSIAN

The experimental results show based on PSNR. The PSNR is the peak signal-to-noise ratio. It is one of the evaluation metrics. Its gives analysis between signal and noise. It is the ratio between signal power the noise power. The more positive the value of PSNR, the less the data is noisy. The PSNR evaluates 15 random videos of the dataset. The 5 videos used from each category as mentioned in table 1. The label of the video also gives in table 1. PSNR calculated on each video as mentioned in table 2 for all four optical flow methods, i.e., Farneback, Horn Schunck, Lucas Kanade, and Lucas-Kanade Derivative of Gaussian. With observation, the PSNR for OF using Farneback is negative. It shows that the variation in frames (Frame 3 and Frame 1 of videos) is the same as the variation in the magnitude optical flow ((Frame 3 and Frame 1 of videos) of all methods. Based on the observation, Lucas Kanade outperformed because the PSNR is the highest in many videos.

The PSNR calculate using expression as

$$PSNR = 10 \log_{10} \left( \frac{\text{peakval}^2}{MSE} \right) \quad (8)$$

The MSE is mean square error between difference of magnitude/Orientation of frame 1 and 3 with difference between grey frame 1 and frame 3.

#### IV. CONCLUSION

Motion estimation from videos for the understanding of motions can be one of the challenging tasks. In this paper, optical flow visualization analyses using four methods. Each technique helps calculate two main parameters: magnitude and orientation. The magnitude gives an understanding of brightness variation from pixel to pixel, and orientation gives the movement of directions. The optical flow visualization shows only one video, i.e., Video 6, mentioned in table 1 and table 2. It's also analyzed using PSNR that variation between two video frames is the same in the optical flow magnitude of each method. Based on the visualization and PSNR, the motion estimating optical flow using Lucas Kanade outperforms the other three motion estimators. The Lucas Kanade can be the preference over others for applications like action recognition, 3D reconstructions, or video coding. The work can extend to using these features for motion estimations.

TABLE 1: NAME OF ACTIVITIES CONSIDERED FOR ANALYSIS WITH VIDEO LABEL AND VIDEO LABEL IN NTURGB+D DATASET

Category of Actions	Name of activity	Video Label	Video label in NTURGB+D dataset
Daily actions	throw	Video-1	S001C001P001R001A007
	drink water	Video-2	S001C001P001R001A001
	stand up	Video-3	S001C001P001R001A009
	reading	Video-4	S001C001P001R001A011
	put on glasses	Video-5	S001C001P001R001A018
Medical Conditions	<i>nausea/vomiting</i>	<b>Video-6</b>	<b>S001C001P001R001A048</b>
	sneeze/cough	Video-7	S001C001P001R001A041
	staggering	Video-8	S001C001P001R001A042
	chest pain	Video-9	S001C001P001R001A045
	back pain	Video-10	S001C001P001R001A046
Mutual Actions	pat on back	Video-11	S001C001P001R001A053
	kicking	Video-12	S001C001P001R001A051
	pushing	Video-13	S001C001P001R001A052
	hugging	Video-14	S001C001P001R001A055
	shaking hands	Video-15	S001C001P001R001A058

TABLE 2: PSNR OF EACH VIDEO USING FARNEBACK, HORN SCHUNCK, LUCAS KANADE , AND LUCAS-KANADE DERIVATIVE OF GAUSSIAN

Method	FARNEBACK	HORN SCHUNCK	LUCAS KANADE	LUCAS-KANADE DERIVATIVE OF GAUSSIAN
Video-1	-33.1463	-11.5332	<b>-11.5215</b>	-11.5492
Video-2	-30.8491	<b>-3.3579</b>	-3.4085	-3.4121
Video-3	-30.5089	-3.7779	<b>-3.7546</b>	-3.8372
Video-4	-31.3553	<b>1.4849</b>	1.196	1.3705
Video-5	-31.2982	<b>1.0258</b>	0.7899	0.9159
<b>Video-6</b>	-32.0707	-13.2235	<b>-13.2216</b>	-13.2336
Video-7	-33.6824	-10.3174	<b>-10.3111</b>	-10.3342

Video-8	-31.6548	-12.7715	<b>-12.7623</b>	-12.7858
Video-9	-33.725	-11.8957	<b>-11.8845</b>	-11.9109
Video-10	-35.2208	-6.3572	<b>-6.3453</b>	-6.3951
Video-11	-31.8951	-6.3589	<b>-6.3536</b>	-6.41
Video-12	-29.2746	-16.3174	<b>-16.313</b>	-16.3265
Video-13	-29.7789	-14.2096	<b>-14.1955</b>	-14.2236
Video-14	-30.1065	-8.518	<b>-8.4931</b>	-8.5484
Video-15	-28.6731	-9.5152	<b>-9.5015</b>	-9.5395

#### ACKNOWLEDGEMENT

The authors are thankful to all members of the Rapid-Rich Object Search (ROSE) Lab from Nan-yang Technological University, Singapore, and Peking University, China.

#### REFERENCES

- [1] T. B. Moeslund, A. Hilton, and V. Krüger, “A survey of advances in vision-based human motion capture and analysis,” *Comput. Vis. Image Underst.*, vol. 104, no. 2-3 SPEC. ISS., pp. 90–126, 2006.
- [2] U. Mahbub, H. Imitiaz, and M. A. Rahman Ahad, “An optical flow based approach for action recognition,” in *14th International Conference on Computer and Information Technology (ICCIT 2011)*, 2011, pp. 646–651.
- [3] M. Lucena, N. P. De La Blanca, J. M. Fuertes, and M. J. Marín-Jiménez, “Human action recognition using optical flow accumulated local histograms,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5524 LNCS, pp. 32–39, 2009.
- [4] A. Gupta and M. S. Balan, *Action recognition from optical flow visualizations*, vol. 703. Springer Singapore, 2018.
- [5] J. K. Aggarwal and Q. Cai, “Q. Cai, Human Motion Analysis A Review-1,” vol. 73, no. 3, pp. 428–440, 1999.
- [6] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, “High Accuracy Optical Flow Estimation Based on a Theory for Warping,” 2004, pp. 25–36.
- [7] G.-P. Ji, D.-P. Fan, K. Fu, Z. Wu, J. Shen, and L. Shao, “Full-duplex strategy for video object segmentation,” *Comput. Vis. Media*, vol. 9, no. 1, pp. 155–175, Mar. 2023.
- [8] Y. Bouafia, L. Guezouli, and H. Lakhlef, “Human Detection in Surveillance Videos Based on Fine-Tuned MobileNetV2 for Effective Human Classification,” *Iran. J. Sci. Technol. Trans. Electr. Eng.*, vol. 46, no. 4, pp. 971–988, Dec. 2022.
- [9] G. Farneb, “Two-Frame Motion Estimation Based on,” *Lect. Notes Comput. Sci.*, vol. 2749, no. 1, pp. 363–370, 2003.
- [10] B. K. P. Horn and B. G. Schunck, “Determining optical flow,” *Artif. Intell.*, vol. 17, no. 1–3, pp. 185–203, 1981.
- [11] J. L. Barron, D. J. Fleet, S. S. Beauchemin, and T. A. Burkitt, “Performance of optical flow techniques,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1992-June, pp. 236–242, 1992.
- [12] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, “Performance of optical flow techniques,” *Int. J. Comput. Vis.*, vol. 12, no. 1, pp. 43–77, Feb. 1994.