Home Work 4: Assignment Report

Titash Mandal (tm2761@nyu.edu, N13592626)

Optical Flow

Question 1.1

1. SCOPE OF THE EXPERIMENT AND CONCEPT OF OPTICAL FLOW.

The objective of the Optical Flow method is to reconstruct the displacement vector field of objects captured with a sequence of images. Sequences of ordered images allow the estimation of motion as either instantaneous image velocities or discrete image displacements.

2. Computer Vision Technique followed to implement Optical Flow.

The optical flow method tries to calculate the motion between two image frames which are taken at times t and at $t + \Delta t$ each pixel position. These methods are called differential since they are based on local Taylor series approximations of the image signal; that is, they use partial derivatives with respect to the spatial and temporal coordinates. For a 2D dimensional case, a voxel at location (x, y, t) with intensity $I(x, y, t)$ have moved by Δx, Δy and Δt between the two image frames, and the following brightness constancy constraint can be given by:

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t)$$

Assuming the movement to be small, the image constraint $I(x, y, t)$ with Taylor series can be developed to get:

$$I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t) + \frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t + \text{H.O.T.}$$

(H.O.T= Higher order terms)

From these equations, it follows that:

$$\frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t = 0$$

or

$$\frac{\partial I}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial I}{\partial y}\frac{\Delta y}{\Delta t} + \frac{\partial I}{\partial t}\frac{\Delta t}{\Delta t} = 0$$

which results in

$$\frac{\partial I}{\partial x}V_x + \frac{\partial I}{\partial y}V_y + \frac{\partial I}{\partial t} = 0$$

where are the $\mathbf{V_x}$, $\mathbf{V_y}$ are the X and Y components of the velocity or optical flow of $I(x, y, t)$ and $\frac{\partial I}{\partial x}$, $\frac{\partial I}{\partial y}$ and $\frac{\partial I}{\partial t}$ are the derivatives of the image in the corresponding directions. I_x, I_y and I_t can be written for the derivatives in the following way:

$$I_x V_x + I_y V_y = -I_t$$

or

$$\nabla I^T \cdot \vec{V} = -I_t$$

This is an equation in two unknowns and cannot be solved as such. This is known as the **aperture problem** of the optical flow algorithms. To find the optical flow another set of equations is needed, given by some additional constraint. All optical flow methods introduce additional conditions for estimating the actual flow.

3. Experimental Procedure

The steps below have been implemented using the two images: "**toy_formatted7.png**" and "**toy_formatted8.png**" and **Gaussian Smoothing value =2.0**)

3.1 Data Capture

Two consecutive images from the given video sequence are chosen. The consecutive image frames are considered for a short interval span only because of the brightness constancy assumption. We assume 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.

toy2.png toy3.png toy4.png
toy5.png toy6.png toy7.png

Fig. 3.1 Colored mages from Video sequence provided to us.

3.2 Applying smoothing to the images.

The smoothing used will be bi-dimensional and separable Gaussian. Smoothing is necessary to ensure mathematical definition of the spatial gradients. It is done with respect to the motion we try to capture between the set of images. For the assignment, the Gaussian Smoothening value of 2.0 is used for both the images. In this assignment, the Gaussian smoothing kernel is defined by:

$$\begin{cases} f : \mathbb{R} \to \mathbb{R} \\ x \mapsto f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{(x-\mu)}{\sigma}\right)^2} \end{cases}$$

where σ is the standard deviation of the Gaussian curve, so this is the parameter that will have to match the size of the displacement we try to characterize. The time gradient is easily defined by:

$$I_t = \frac{\partial I}{\partial t} = I^*(x, t + \delta t) - I^*(x, t)$$

where $I*$ is the smoothed image.

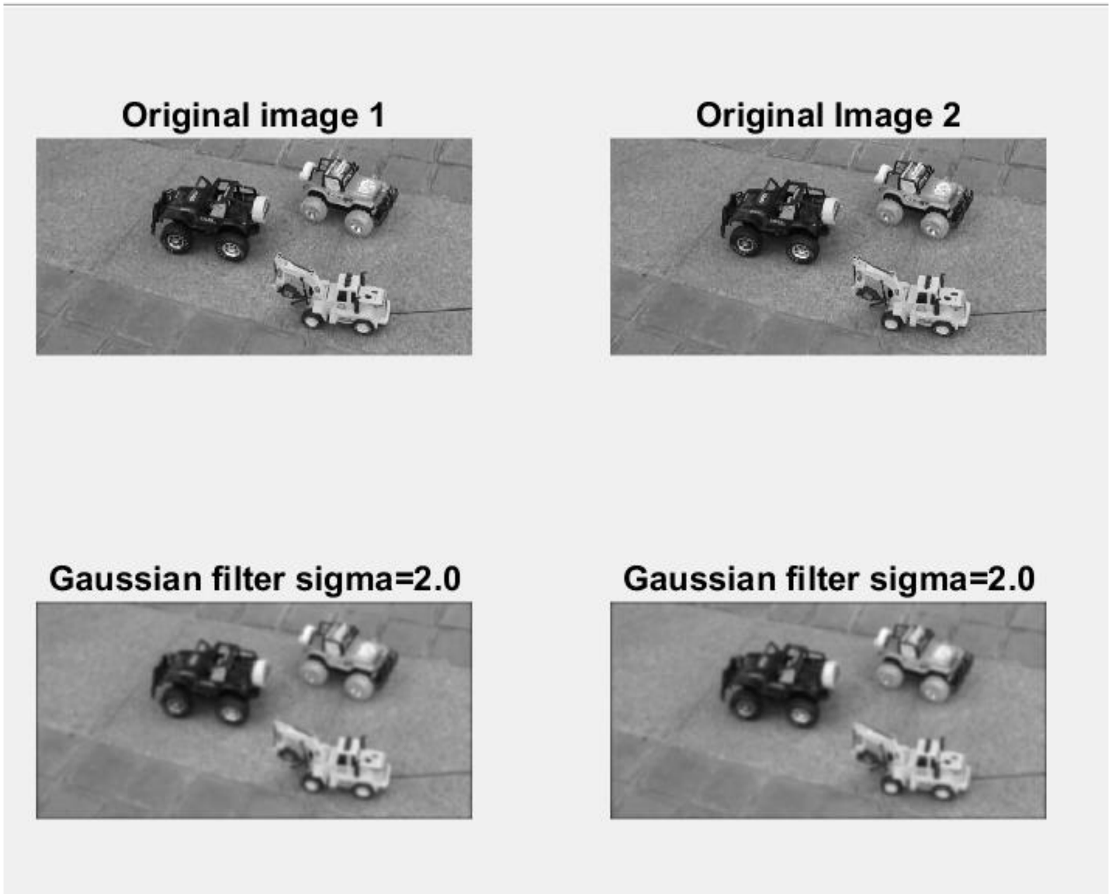


Fig. 3.2 The output images shown after the Gaussian Smoothing was applied to them.

3.3 Calculating the temporal gradient image $\partial E/\partial t$ via the difference of the blurred versions of the two consecutive frames.

The temporal gradient is simply calculated by subtracting the two image frames I(x, y,t+1) – I(x,y,t).

3.4 Estimating the spatial derivatives $Ex = \partial E/\partial x$ and $Ey = \partial E/ \partial y$

- We estimate the spatial derivatives by calculating pixel differences I(x+1,y ,t)-I(x ,y,t) for Ex and I(x,y+1,t)-I(x,y,t) for Ey.
- On the left image, Fig. 3.4a, we can clearly see that vertical edges are very well defined.
- This is the effect of the gradient on the horizontal direction (X direction).
- It enhances the difference of the values on this direction and the stronger are created by vertical lines
- The same thing is noticed for the image Fig. 3.4b where we have the gradient according to the vertical direction (Y direction).
- Here in Fig. 3.4b the strongest discontinuities enhanced are horizontal lines.
- The time gradient is the difference in the intensity values between the two images.

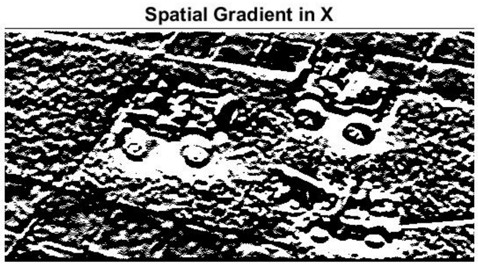


Fig. 3.4a Image showing the output of calculating the spatial gradient in the X direction.

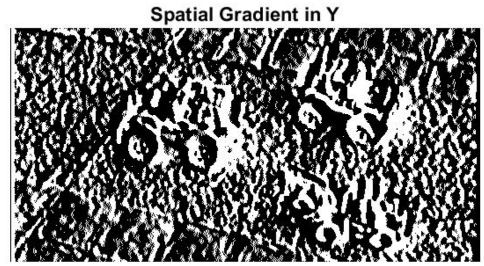


Fig. 3.4b Image showing the output of calculating the spatial gradient in the Y direction.

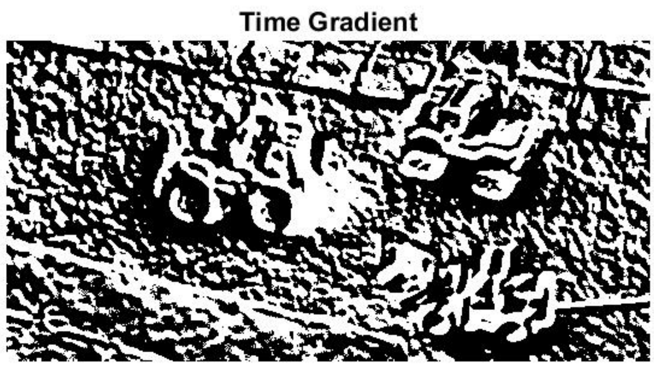
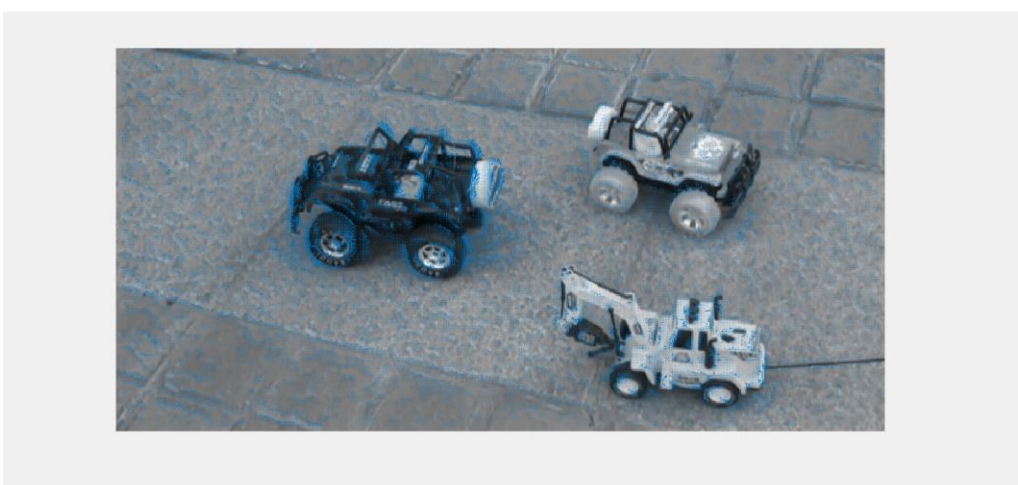


Fig. 3.4c Image showing the Time Gradient of the images.

3.5 Calculating the normal flow at each pixel.

In this method, we only deal with pixel information and do not take in account the neighborhood information. This could lead to an aperture problem where the overall motion of the object is not taken in account. We only compute for each pixel an estimation linked to the three gradients available. Here is the definition of the normal flow:

$$u_{\perp} = \frac{-I_t}{||\nabla I||}\frac{\nabla I}{||\nabla I||} \qquad \text{with} \qquad ||\nabla I|| = \sqrt{I_x^2 + I_y^2}$$



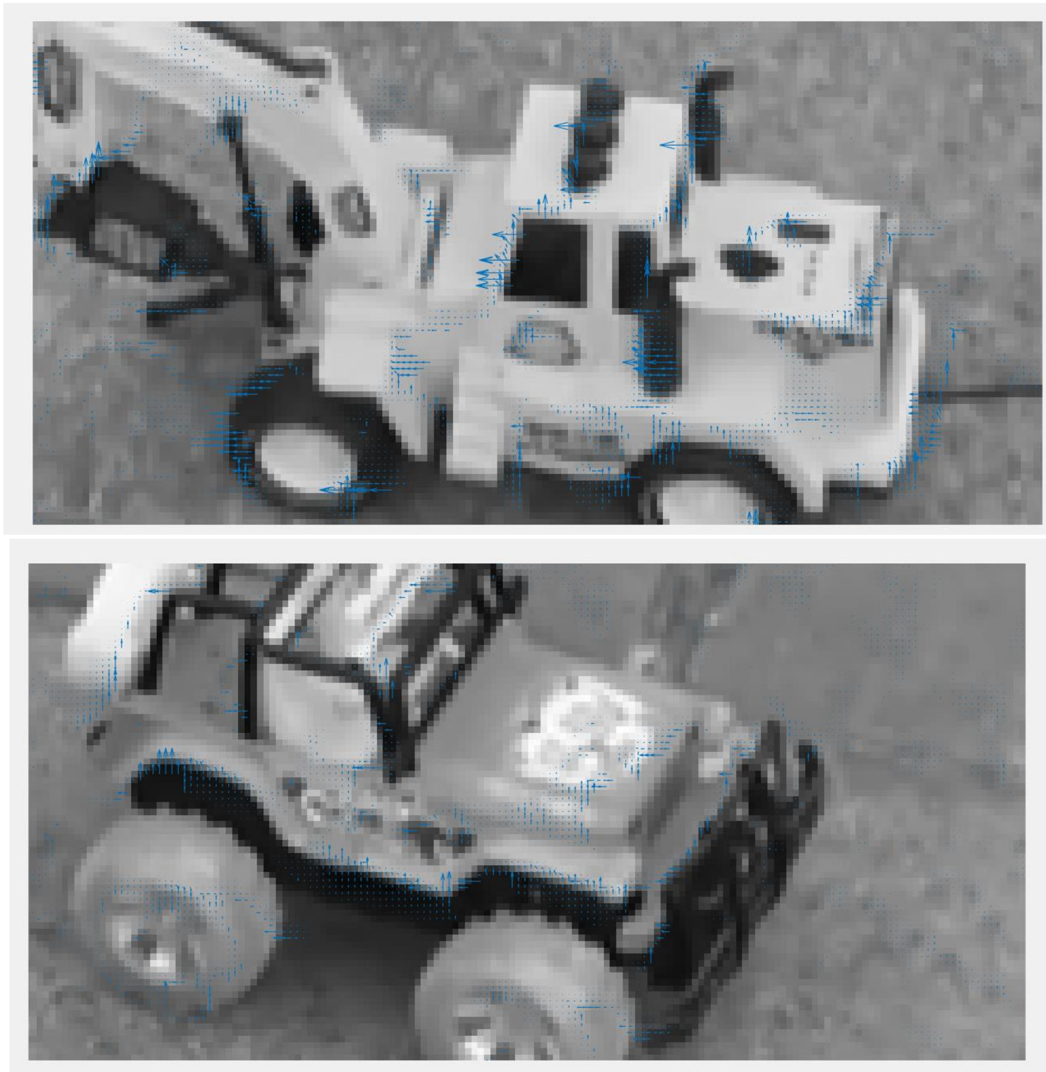


Fig. 3.5.1 Image showing the normal vectors overlaid with the gray level image.

Fig. 3.5.2 Image showing the maximized images of the individual cars to check the normal vectors.

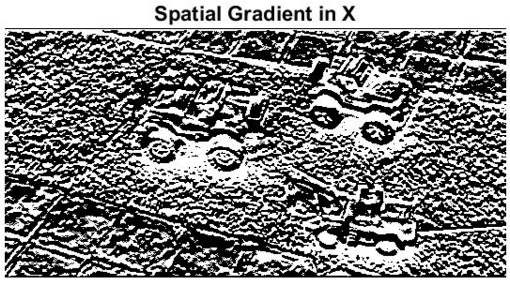
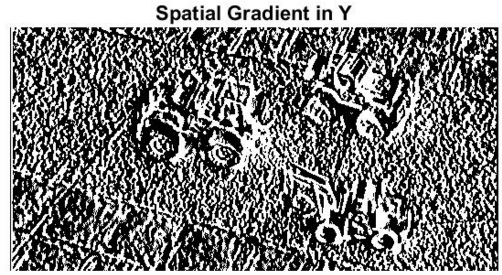
3.6 Discussion of the results of the surface normal vectors determining the direction of movement.

- It is very complicated to clearly see the associated displacement field vectors for each car because the results are not clearly visible as there are many directions associated with one surface.
- The original computed displacement field has a very low vector magnitude, so to have a more visible field to be analyzed, it is necessary to magnify the images.
- If we see closely, for the **first car**, we can see that most of the **displacement vectors are oriented to the left**, and for the **two others** most the arrows are **pointing to the back and some point to the right**. So the second pair of cars maybe moving in the forward direction and taking a turn towards the right.
- There may be another question coming into picture because the background of the image is also colored in white and seems to have a right to left motion.
- Due to the nature of the image acquisition there is a small motion of the camera between the frames which is contained in the temporal gradient and which is taken in account for the displacement computation. Hence, this may be an explanation of the motion associated with the background of the image and which could also be seen in the vector field.
- The size of the smoothing kernel used in the preprocessing step has an impact on the result of the velocity vectors we have obtained. The smoothing kernel must be big enough to capture the entire frame motion but it should not too big.
- If it is too big, a too wide smoothing result in smoothing the displacement field and a possible reduction of available accurate information which is not desired.

Fig. 3.6.1 Image of the first car zoomed to show that the arrows are directed towards the left.

Fig. 3.6.2 Image showing the direction for the second car pointing towards the right.

3.7 Implementing the similar procedure on two different images from the video sequence. The images chosen are "toy_formatted2.png" and "toy_formatted3.png" and Gaussian Smoothing value =2.0.

 Fig. 3.7a Image showing the output of calculating the spatial gradient in the X direction.	 Fig. 3.7b Image showing the output of calculating the spatial gradient in the Y direction.

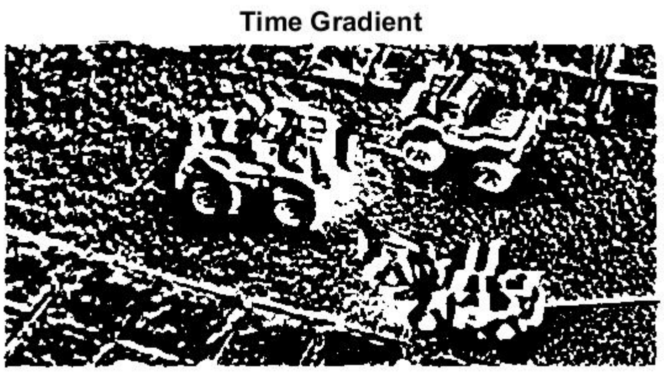


Fig. 3.7c Image showing the time gradient of the two images.

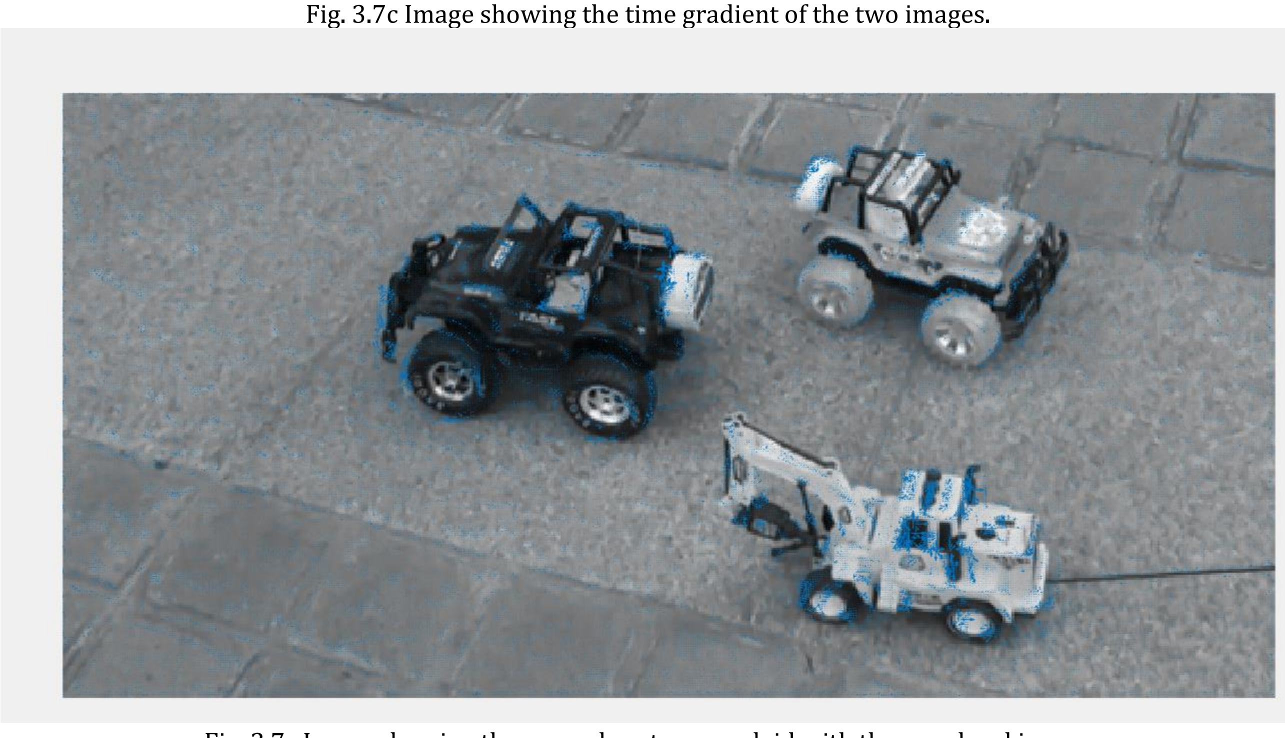


Fig. 3.7c Image showing the normal vectors overlaid with the gray level image.

3.8 Results obtained by changing the images chosen for the assignment.

- Using a new pair of images, we observe that the normal vectors are aligned in an even more complicated way with some surfaces having two arrows crossing each other. It becomes difficult to predict the direction of the movement of the cars.
- We have a horizontal motion of 3 pixels between the two frames as shown in the Fig. 3.8a
- Thus, the resulting field using Gaussian Smoothing value of 2 is very low and not very accurate but the reconstruction remains possible.
- This is an illustration that smoothing is a major contributing factor for the accuracy of our reconstruction of displacement field.

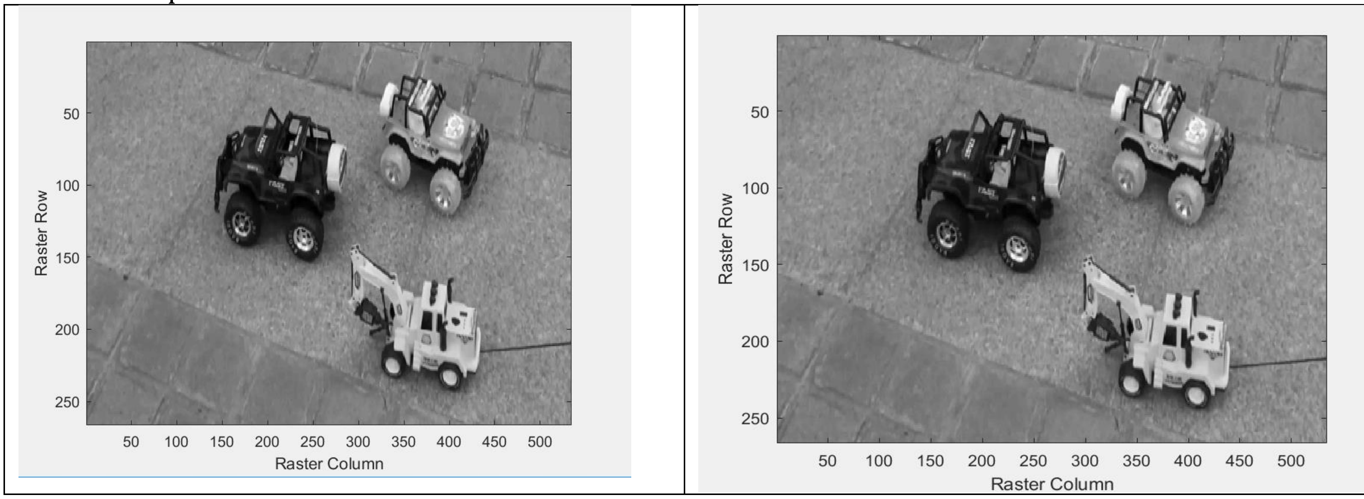


Fig. 3.8a Figure showing the two images chosen for the assignment are plotted on the co-ordinate system to determine their displacement with respect to each other.

Fig. 3.8b Image showing the output direction flow of the normal vectors overlaid on the grayscale image when the new set of images(toy_formatted2 and toy_formatted4) are taken.

Fig. 3.8c Image showing the output direction flow of the normal vectors overlaid on the grayscale image when the old set of images (toy_formatted7 and toy_formatted8) are taken.

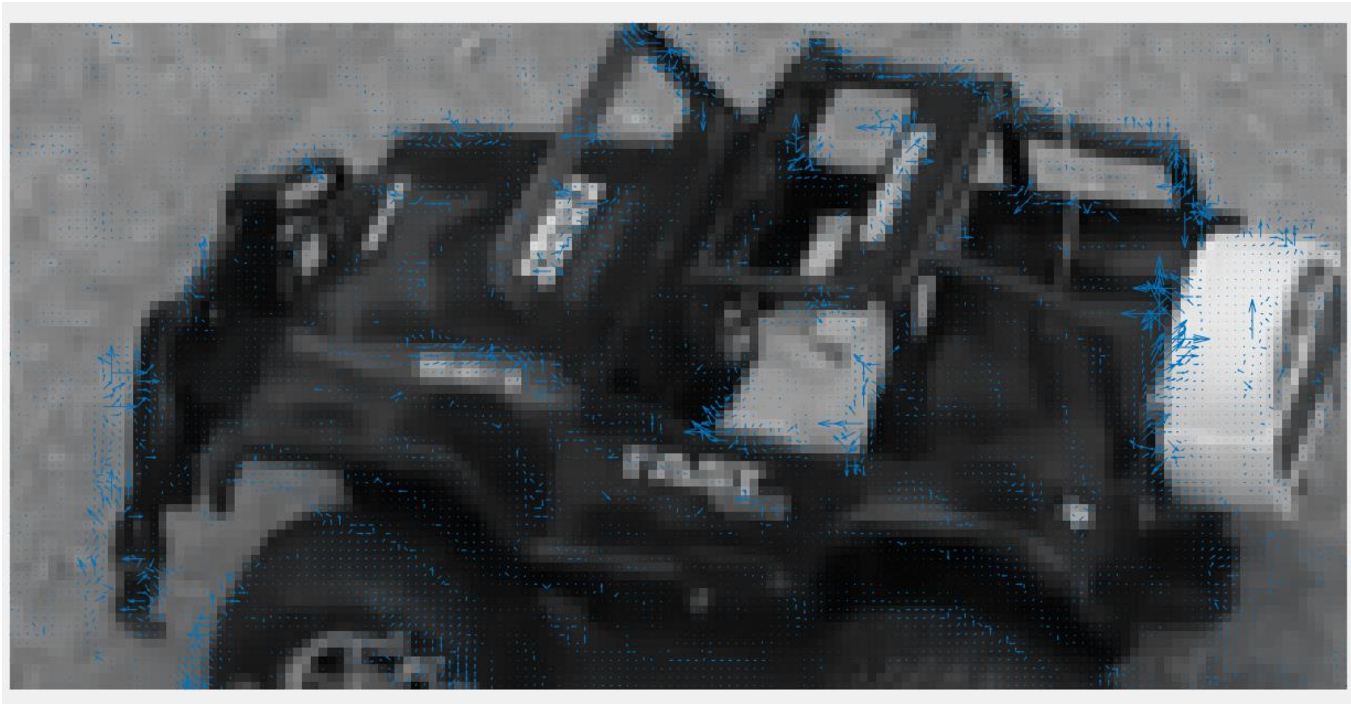
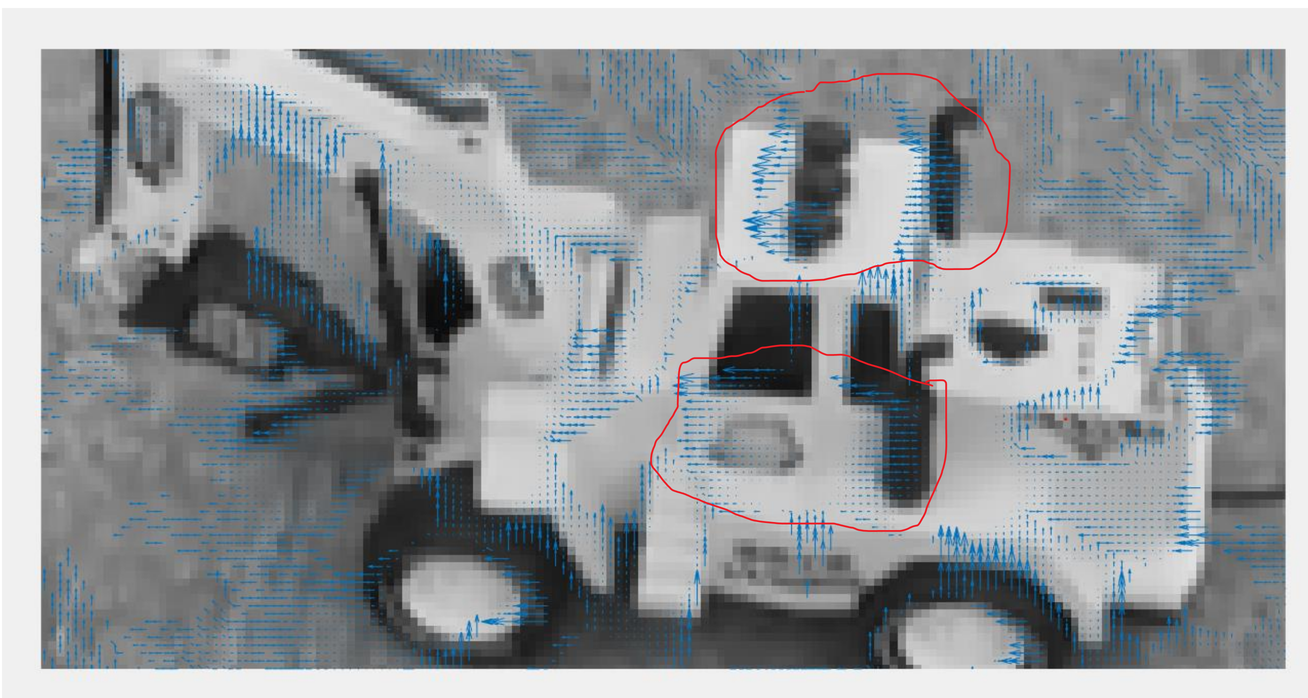


Fig. 3.8d Image of the first car zoomed to show that the arrows are directed towards the left.

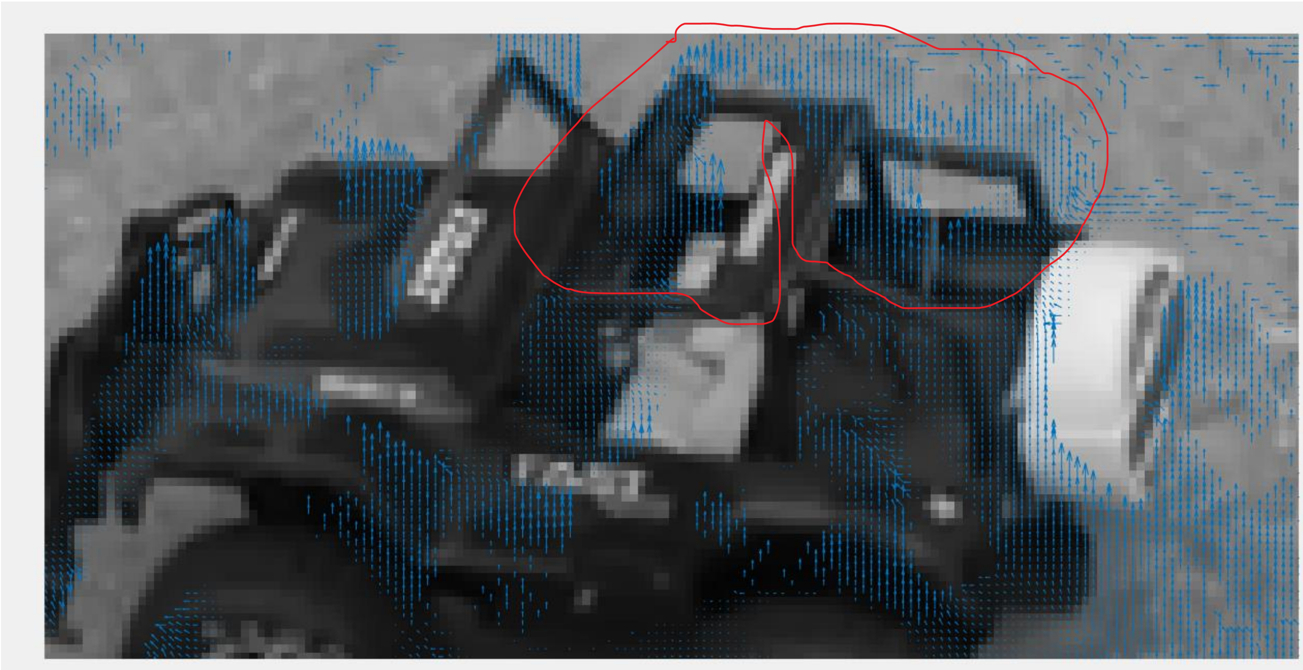
Fig. 3.8e Image of the second car zoomed to show that the arrows are directed towards the right.

Fig. 3.8f Image of the third car zoomed to show that the arrows are directed towards the right.

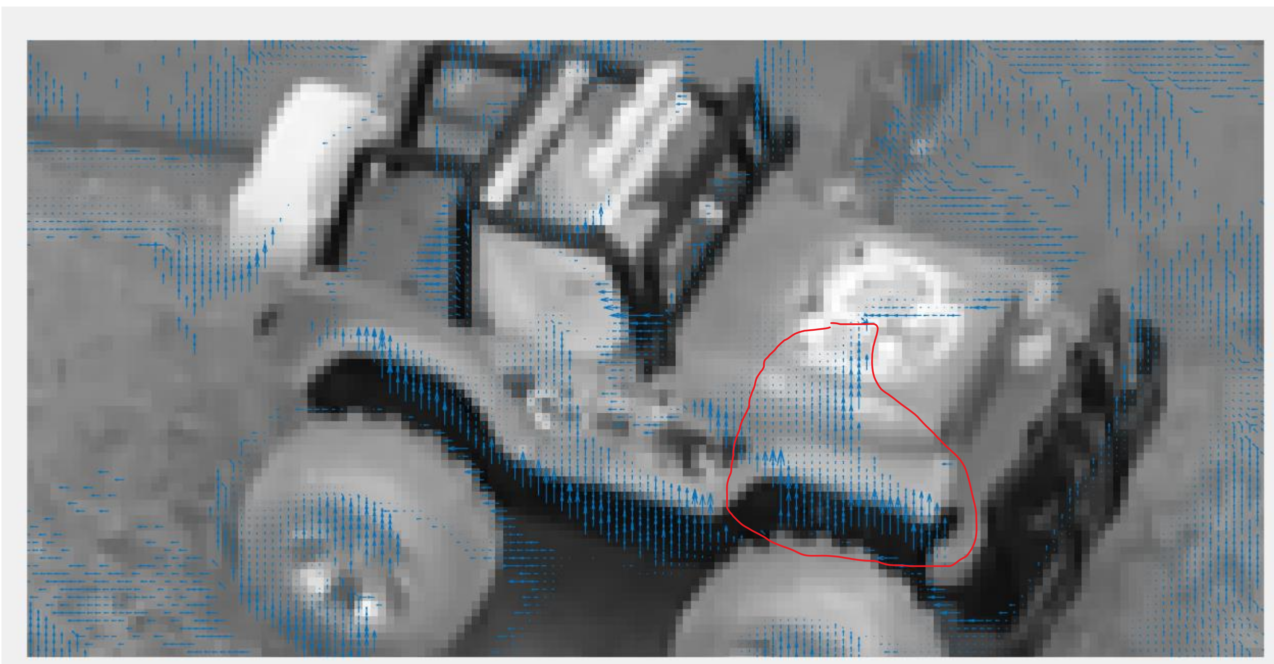
However, when applying a Gaussian Smoothing value of 3.0, the normal vectors are more correctly oriented and they show better results.



From the direction of the normal vectors, we can infer that this car maybe moving in the forward direction and maybe taking a turn towards the right direction.



The direction of normal vectors is upwards.



1.2 Flow calculated over pixel neighborhood

- Just like the previous method, smoothing was applied on the pair of images chosen from the video sequence.
- We then choose a 2x2 pixel neighborhood for the local estimate of velocities using the following strategy.

$$
\begin{aligned}
\nabla E.v + \frac{\partial E}{\partial t} &= 0
\end{aligned}
$$

now we have 4 measurements from 2x2 pixels to solve for v:

$$
\begin{aligned}
Av + b &= 0 \\
A^T A v &= -A^T b \\
v &= -(A^T A)^{-1} A^T b \\
C &= A^T A
\end{aligned}
$$

- The columns of A are the x and y components of the gradient ∇E.
- b is a column vector of the t gradient component of E, E_t.
- The inverse was calculated with Matlab's pinv.
- This calculation was performed at each pixel (x,y) in the image with the columns of A and b extracted within a neighborhood of size 2x2.
- For calculations that gave NaN at some pixel locations, an output equal to zero flow vector.

This method is applied to the following:

1.2.a) **Set of raw images.**

When the optical flow method over a neighborhood is applied directly on the set of raw images without applying the Gaussian Filter, we observe that the normal vectors isn't clearly visible on the image. Only one or two arrows indicate the direction of motion of the cars and this makes reconstruction of direction extremely difficult. The spatial gradients in the X and Y axis and the temporal gradient however show comparable results like the previous question. The images below show the result of implementing this method.

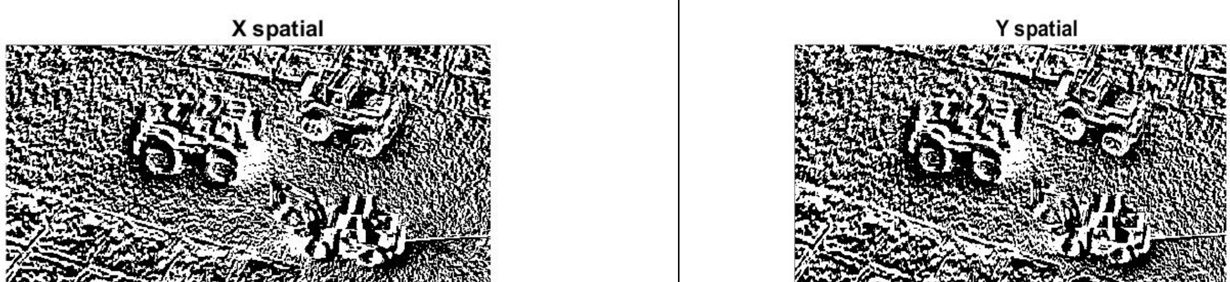


Fig. 3.9a Image showing the spatial gradient in the X and Y axis.

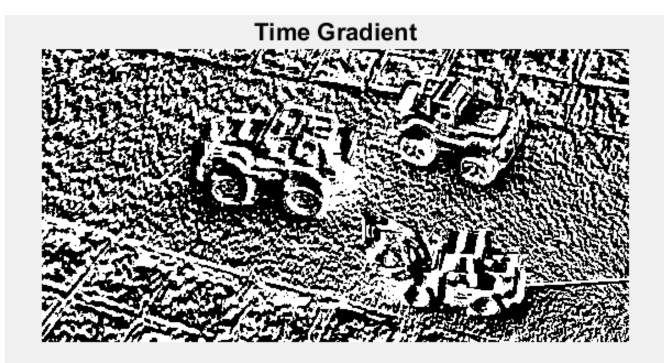


Fig. 3.9b Image showing the time gradient of the two images.

Fig. 3.9c Image of the first car zoomed to show that the direction of the arrows directed towards the left.


Fig. 3.9d Image of the second car zoomed to show that the direction of the arrows is directed towards the right.


Fig. 3.9c Image of the third car zoomed to show that the direction of the arrows directed towards the right.

b) **Set of two Images filtered by a Gaussian smoothing of sigma=1.**

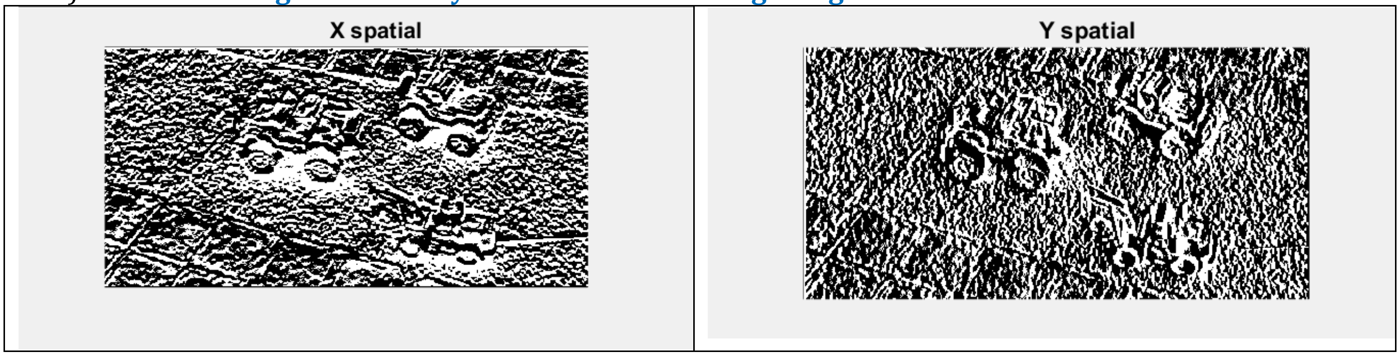


Fig. 3.10a Image showing the spatial gradient in the X and Y axis.

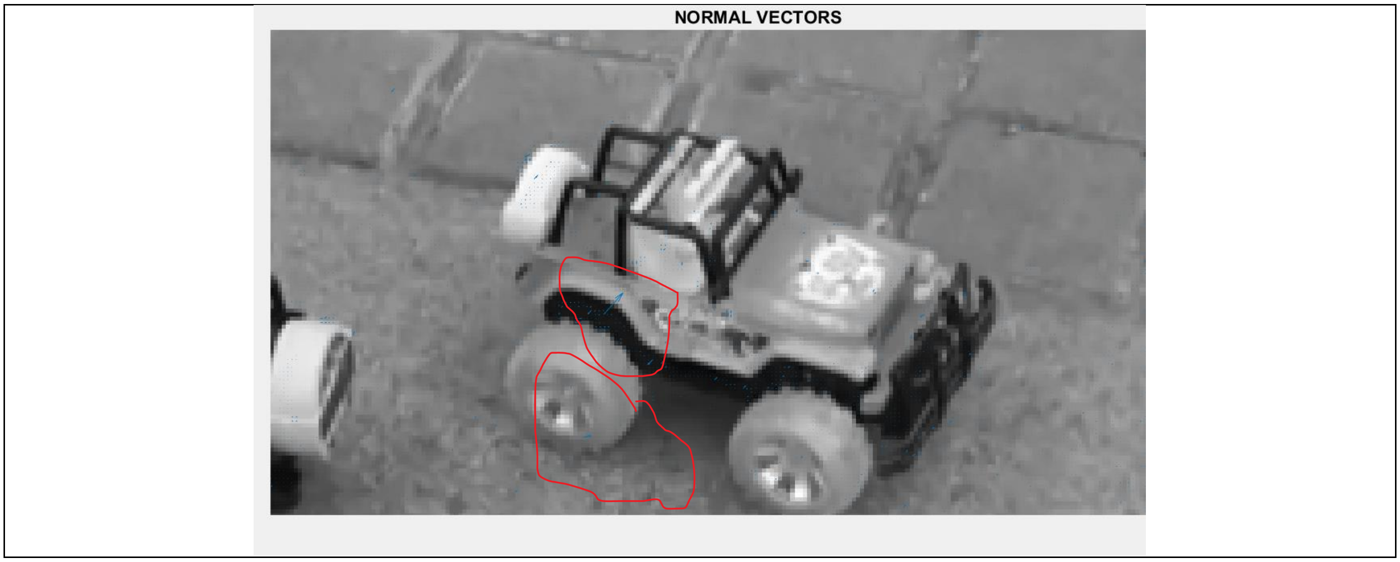


Fig. 3.10b Image showing the spatial gradient in the X and Y axis.

This gives a better result in terms of overlaying the normal vectors over the gray scale images. We can see the normal vectors aligned towards the appropriate direction. However, drawing conclusions still is difficult. As in the third image we can see some arrows point towards the right and some point to the left. So it is difficult for me to predict the direction and it gives a very ambiguous result.

c) **Set of two Images filtered by a Gaussian smoothing of sigma=2.**

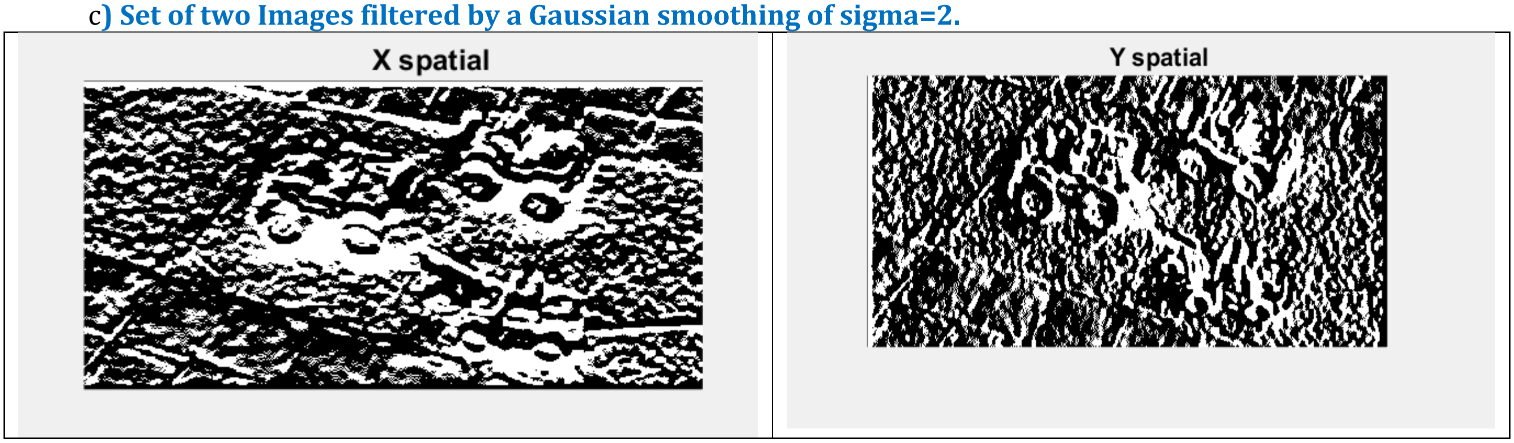


Fig. 3.11a Image showing the spatial gradient in the X and Y axis

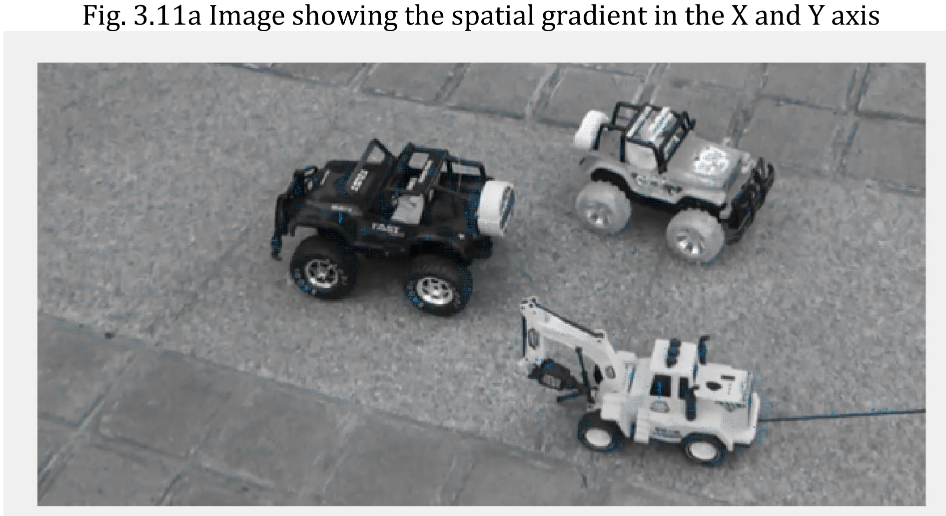


Fig. 3.11b Image showing the normal vectors overlaid with the gray level image.

The result for applying the Gaussian filter value 2.0 was almost the same as applying the value as 1.0.

Bonus 1: Horn and Schunck

The Horn–Schunck method of estimating optical flow is a global method which introduces a global constraint of smoothness to solve the aperture problem. The Horn-Schunck 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.The flow is formulated as a global energy functional which is then sought to be minimized. This function is given for two-dimensional image streams as:

$$E = \iint \left[(I_x u + I_y v + I_t)^2 + \alpha^2 (\|\nabla u\|^2 + \|\nabla v\|^2) \right] \mathrm{d}x \mathrm{d}y$$

where I_x, I_y and I_t are the derivatives of the image intensity values along the x, y and time dimensions respectively. $\vec{V} = [u(x,y), v(x,y)]^\top$ is the optical flow vector, and the parameter α is a regularization constant. Larger values of α lead to a smoother flow. This functional can be minimized by solving the associated multi-dimensional Euler-Lagrange equations. These are:

$$\frac{\partial L}{\partial u} - \frac{\partial}{\partial x}\frac{\partial L}{\partial u_x} - \frac{\partial}{\partial y}\frac{\partial L}{\partial u_y} = 0$$
$$\frac{\partial L}{\partial v} - \frac{\partial}{\partial x}\frac{\partial L}{\partial v_x} - \frac{\partial}{\partial y}\frac{\partial L}{\partial v_y} = 0$$

where L is the integrand of the energy expression, giving,

$$I_x (I_x u + I_y v + I_t) - \alpha^2 \Delta u = 0$$
$$I_y (I_x u + I_y v + I_t) - \alpha^2 \Delta v = 0$$

Discussion of the results of the surface normal vectors determining the direction of movement.

Advantages of the Horn–Schunck algorithm include that it yields a high density of flow vectors, i.e. the flow information missing in inner parts of homogeneous objects is filled in from the motion boundaries. On the negative side, it is more sensitive to noise than local methods. The Horn and Shunck method introduce also a good improvement of the computed displacement field with a completely different method.

We implemented the Horn and Schnuk Algorithm in the assignment and the result is as follows.

Discussion of results from Horn and Schnuk

The optical flow cannot be computed at a point in the image independently of neighboring points without introducing additional constraints, because the velocity field at each image point has two components while the change in image brightness at a point in the image plane due to motion yields only one constraint. In the Horn and Schnuk method, we have a choice of how many iterations to perform. We could iterate until the solution has stabilized before advancing to the next image frame. On the other hand, given a good initial guess one may need only one iteration per time-step. A good initial guess for the optical flow velocities is usually available from the previous time-step. The advantages of the latter approach include an ability to deal with more images per unit time and better estimates of optical flow velocities in certain regions. Areas in which the brightness gradient is small lead to uncertain, noisy estimates obtained partly by filling in from the surround. These estimates are improved by considering further images. The noise in measurements of the images will be independent and tend to cancel out. Perhaps more importantly, various parts of the pattern will drift by a given point in the image. The direction of the brightness gradient will vary with time, providing information about both components of the optical flow velocity.

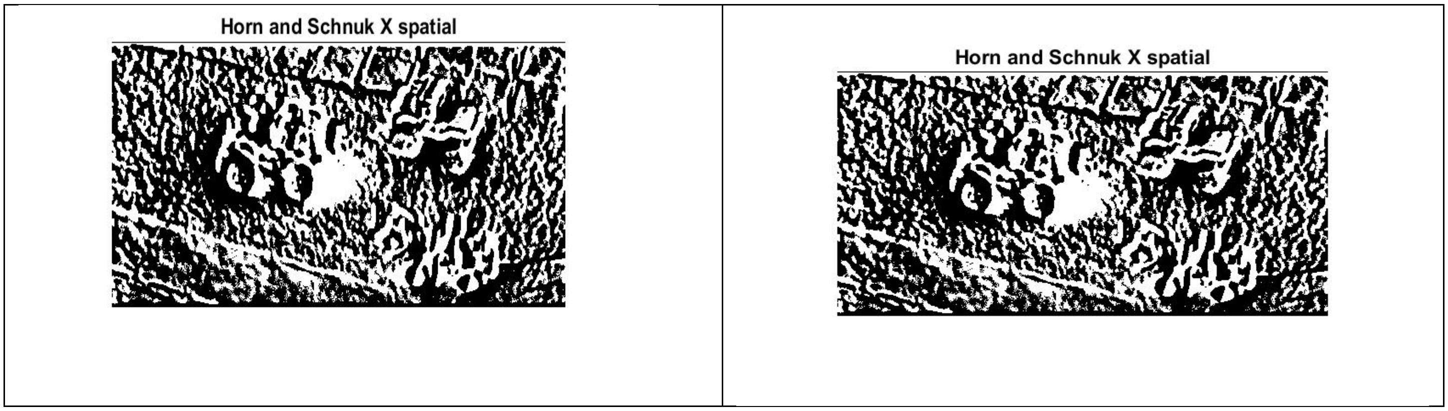


Fig. 3.12a Image showing the spatial gradient in the X and Y axis.

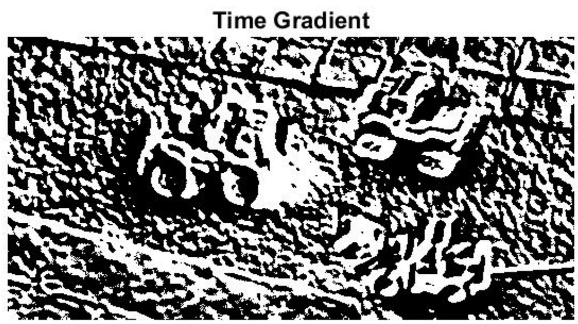


Fig. 3.12c Image showing the spatial gradient in the X and Y axis.

Fig. 3.12d Normal vectors overlayed over the original image.

Fig. 3.12e Image of the first car zoomed to show that the direction of the arrows directed towards the left.

Fig. 3.12f Image of the second pair of cars zoomed to show that the direction of the arrows directed towards the left.

Fig. 3.12g Image of the second pair of cars zoomed to show that the direction of the arrows directed towards the left.

RESULTS:

With Horn and Schnuk method, we can see that the arrows are placed in a very streamlined direction over the surface of the cars and are oriented towards the direction of the motion of the car. This gives a better understanding of velocity vectors and we can infer that the cars are moving forward and right and left as mentioned in the figures above.

BONUS TWO- RUNNING THE OPTICAL FLOW SIMPLE SMOOTHING ALGORITHM (as mentioned in question 1.1) BY USING OWN SET OF IMAGES FROM VIDEO SEQUENCE.

We tried to run the optical flow program using the video sequence of a moving ball and determine the direction in which the ball moves by looking at the direction of the velocity vectors. The images used for detecting motion are given in the figure below. The images were derived from a video sequence by taking every 3rd frame.

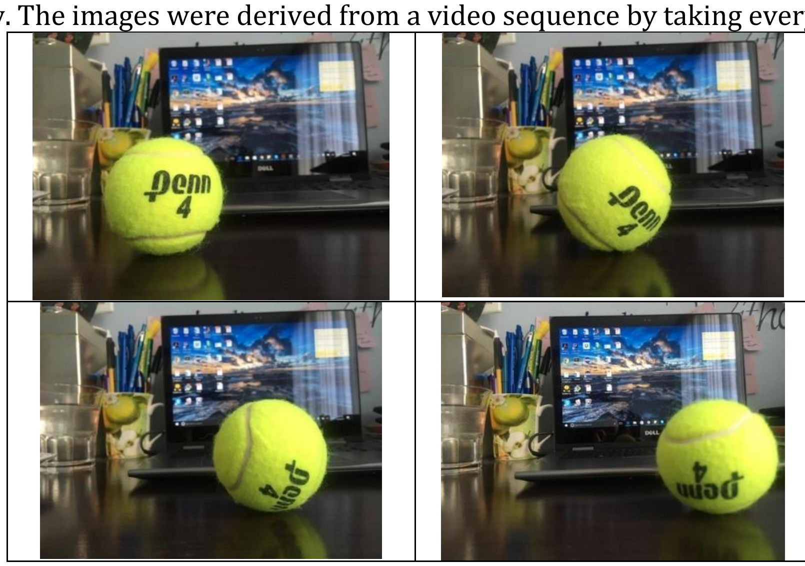


Fig. 3.13a Images extracted from the video sequence.

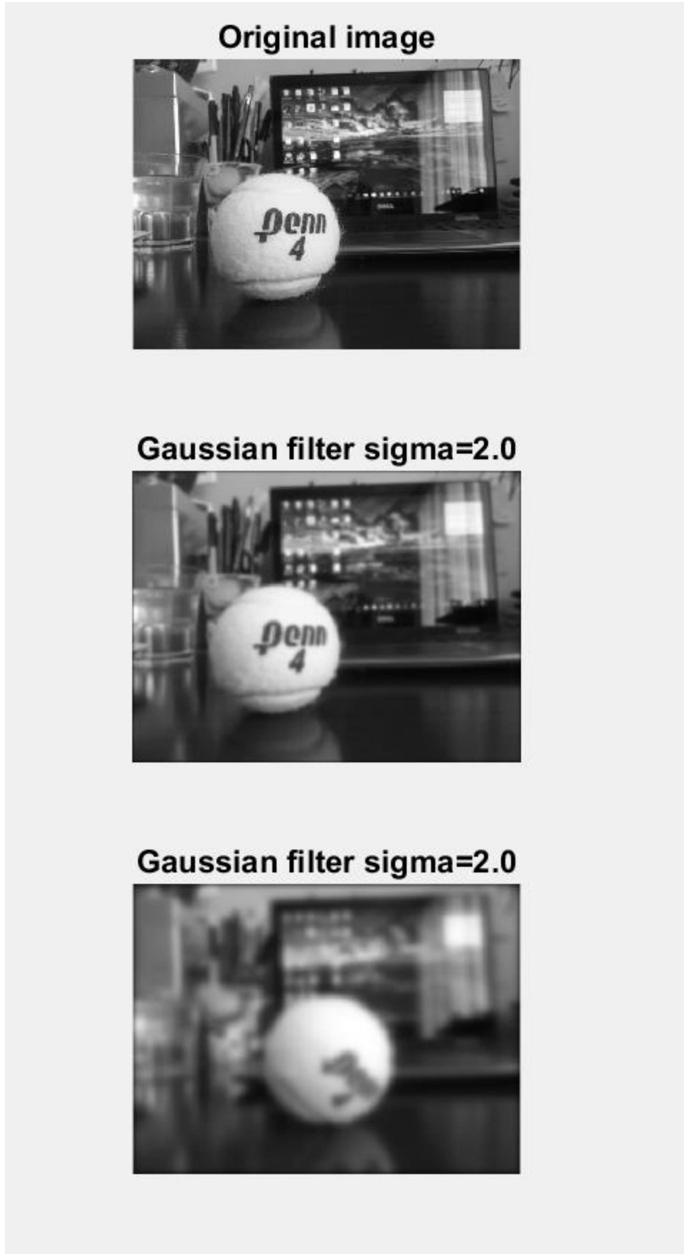


Fig. 3.13b Gaussian Smoothing applied to the images

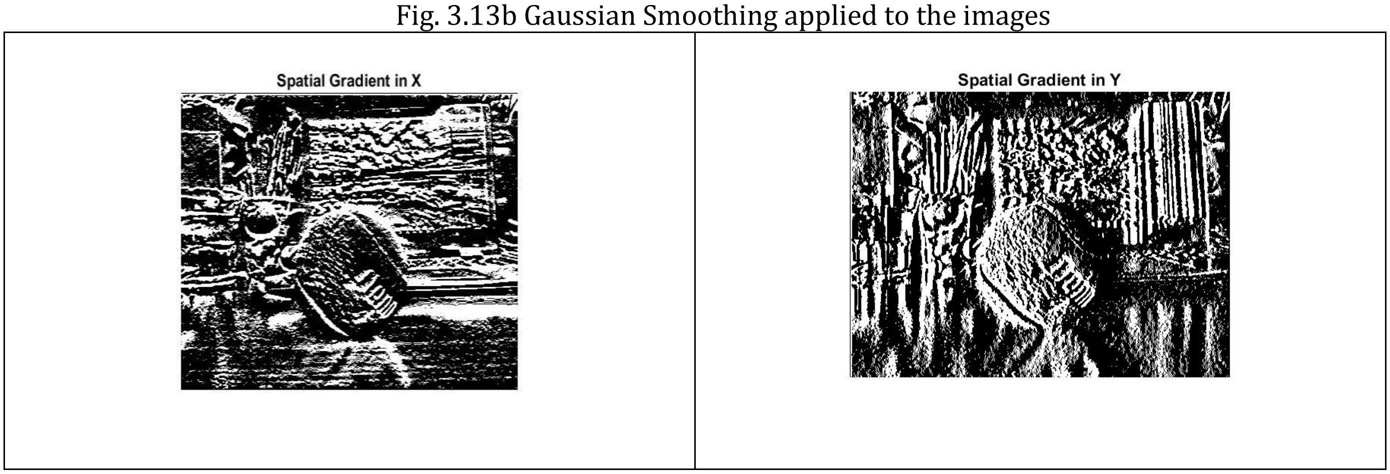


Fig. 3.13c Image showing the spatial gradient in the X and Y axis.

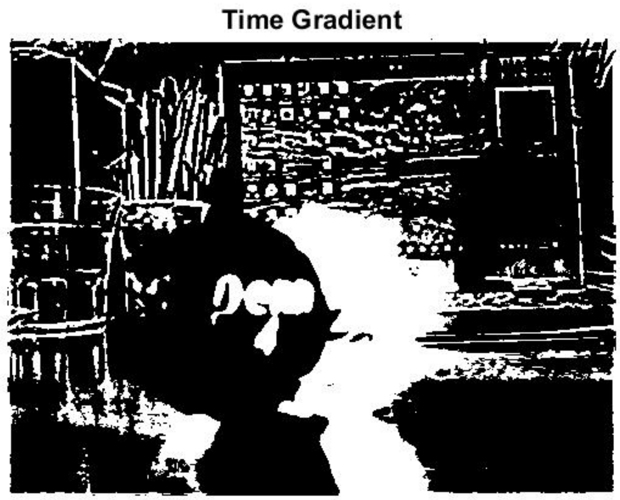


Fig. 3.13d Image showing the temporal gradient in the X and Y axis.

Fig. 3.12e Image showing the normal gradient in the X and Y axis.

Results:

The velocity vectors when zoomed on the image show that the direction of the movement of the ball is forwards and towards the right. The second mesh of the normal vectors drawn on the ball is slightly ahead of the first set of normal vectors drawn. Hence the direction can be easily predicted looking at the vectors. As we are using a gradient-based method to find the spatial and temporal partial derivatives (to estimate image flow at every position in the image), we have shown their outputs in the images presented above. As we assume that the image motion is not known in advance, we restrict it to a small range of possible values. We then apply a multi-scale analysis so that the scale of the smoothing prior to derivative estimation is appropriate to the scale of the motion. This can result in gradient-based methods being computationally expensive. This problem is the equivalent of the correspondence problem in feature-based methods. There are two major problems with the approach of Horn and Schunk. One is that only well-conditioned at parts of the image which have high gradient, as can be seen. Thus, results must be "spread" into those regions with low gradient if the aim of achieving flow estimates at every image position is to be achieved. The other problem is that the optic flow will be smoothed across flow discontinuities, resulting in inaccurate flow estimates.

APPLYING THE SAME SET OF IMAGES TO HORN AND SCHNUK ALGORITHM

The normal vectors lie on the surface of the ball and they give a good output.

CONCLUSION-

- Through this assignment, we studied and implemented an interesting computer vision technique to reconstruct the motion between the two images acquired through the video sequence.
- The reconstruction of the Optical Flow allows us to estimate the displacement of moving objects of a scene over time.
- In both the methods, we used a magnifying factor to produce more visible and understandable displacement field and to allow us to make meaningful analysis and comparisons.
- The velocity vectors were directed in different directions making it difficult to analyze and conclude the actual direction of movement of the objects.
- However, looking at the direction of the majority arrows we concluded the results.
- One of the fundamental property used by the previous method was to consider that, the inter frame motion is small.