IMA 208 - Motion estimation

Paul Fayard

April 2020

I, the undersigned Paul Fayard , born on 30/07/1999 in Toulon (France), certify on my honour that I am not using any form of fraud for the writing of this report.

I am well aware of the corresponding paragraphs of the school rules (paragraph 14.2, regularity of the written tests) and the penalties incurred.

Paul FAYARD

1 Introduction

2 Motion estimation techniques

2.1 Block matching

Question 3

Is motion estimation able to find the real scene motion ? Are there errors ? If yes, where and why ?

Overall, the MVF seems to respect the real scene motion (moving to the right). However, there are many errors in the upper right. Indeed, this is a homogeneous area so the block matching tends to make more errors because all the blocks in this region look similar, especially if there is a slight noise. That's why it is necessary to add a regularization.

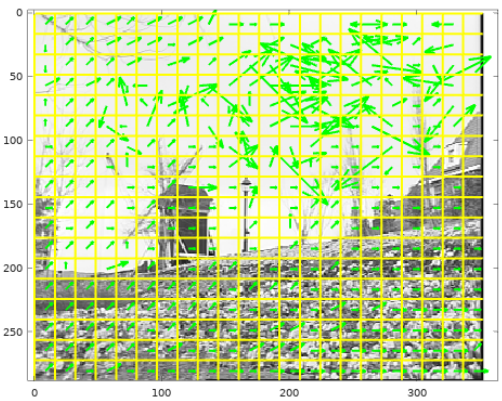

Figure 1: The MVF with SSD

Question 4

Change the block size and the search area radius. What is the impact on the previous question ?

When you reduce the size of the search area, we still get aberrant errors, especially in homogeneous areas. When you have a preconceived idea about a small displacement, it is better to reduce the search area. But when it is not the case, you mustn't take an area too small : you risk to miss the real displacement. A large search area increases the computation time and if the search area is too large, we risk also missing the real information, the real displacement in the homogeneous region.

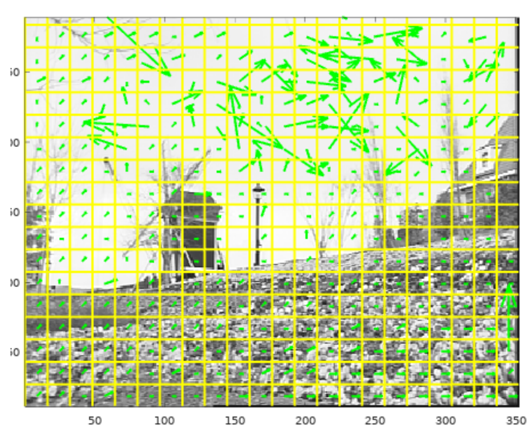

Figure 2: With a larger search area

For the size of the blocks, we get a more accurate result, with fewer errors, if the blocks are small. Indeed, the smaller the blocks are, the more obvious the correspondence between them is from one image to another. However, this does not solve our problem in the homogeneous area at the top right.

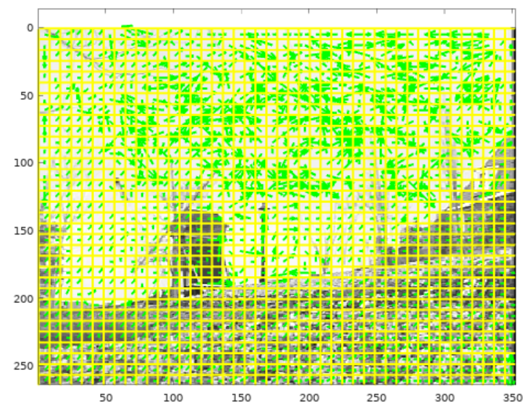

Figure 3: With smaller blocks

Question 6

Compute the PSNR of the MC-ed prediction with respect to the original image. Change the block size and comment the result.

Image	PSNR
Initial (8 and 16*16)	22,71
Larger search area (16)	22,71
Smaller search area (4)	22,71
Smaller blocks (8*8)	23,23
Larger blocks (32*32)	22,45

As we expected, working with smaller blocks gives a better result, a more precise result (higher PSNR) and conversely working with large blocks gives a poorer result. In this example, however, the size of the search area does not have a big impact. Indeed, the real displacement is small so it doesn't change much.

Question 8

Comment the difference observed between SSD and SAD in terms of regularity of MVF,PSNR of the prediction, impact of the block size.

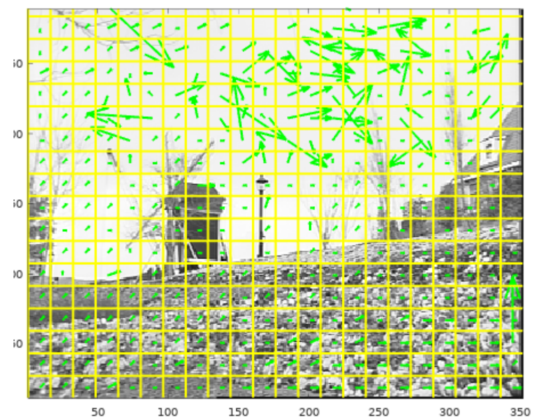

Figure 4: MVF with SAD

We see that the result with SAD is more regular than the one obtained with SSD. In fact, with the SSD, the square operator highlights errors, so they are bigger. However, the PSNR is equal to 22,45, so this is less precise than SSD, but it's normal because the prediction error has necessarily larger energy with the absolute value. We would prefer SAD because it does not emphasize the errors.

SAD	PSNR
Initial (8 and 16*16)	22,45
Larger search area (30)	22,62
Smaller search area (4)	22,45
Smaller blocks (8*8)	23,23
Larger blocks (32*32)	22,45

As we SSD, the result is better with small blocks. The search area as more impact here. We are a little more precise when it increases.

Question 11

Comment the difference observed between SSD and regularized SSD in terms of regularity of MVF, PSNR of the prediction, impact of the block size.

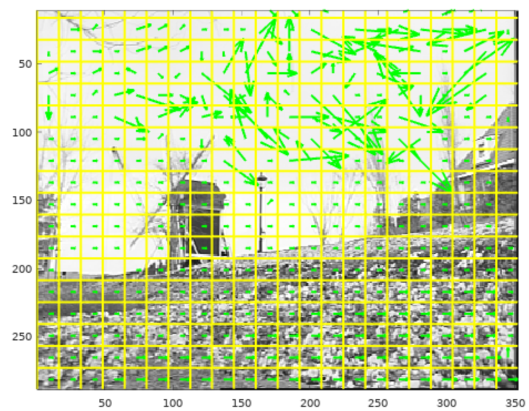

Figure 5: SSD

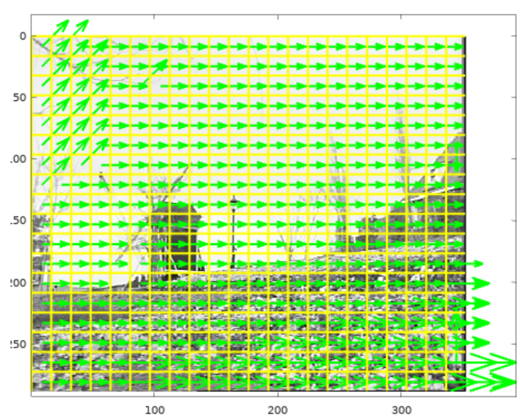

Figure 6: Regularized SSD

We modify the criterion by en penalizing vectors that are too much different from their neighbors. This avoids absurd displacements, especially in homogeneous areas. We can see that this way, all the blocks have almost the same displacement, to the right, which is satisfactory.

The PSNR score confirms that by regularizing, we obtain a result closer to the true motion.

Image	PSNR
SSD (16*16 and 13)	26,28
SSD Regularized (16*16 and 13)	26,23
SSD Regularized + bigger blocks (32*32)	22,73
SSD Regularized + smaller blocks (8*8)	26,95
SSD Regularized + smaller search area (5)	26,23
SSD Regularized + bigger search area	26,28

The impact of the parameters is always the same. It's better to widen the search area a bit by taking search area=30 and to take a small block size (8*8).This takes a lot of time, sometimes more than 40 seconds.

2.2 Optical flow

Question 4

Comment the result in terms of regularity of MVF, PSNR of the prediction.

The block matching algorithm uses a batch processing approach while the optical flow algorithm implements a pixel by pixel processing approach.

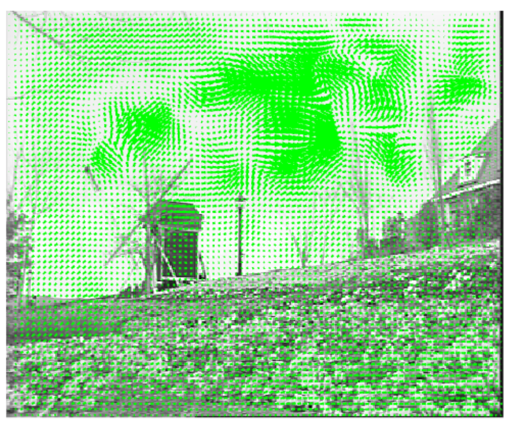

Figure 7: Optical flow with flower

We observe a much denser FVM because we now work pixel by pixel, and more regular than with SSD, and SAD. For flower, we always find a movement towards the top right. For akiyo, we can see that from the first video to the second one, she turns her head downwards to the left, which corresponds well to real motion.

For PSNRs, we observe that the precision is more important in the case of

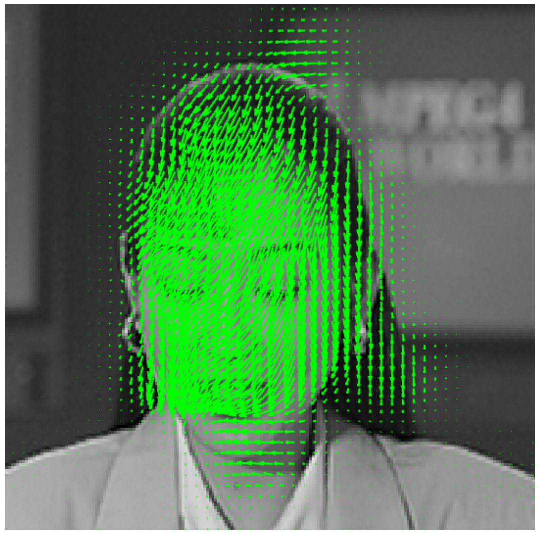

Figure 8: Optical flow with akiyo

the optical flow algorithm.

Image	Method	PSNR
Flower	SSD	22,71
Flower	H-S	24,68
Akiyo	SSD	40,46
Akiyo	H-S	42,25

3 Parametric estimation

3.1 Direct parametric estimation

Question 3

Compute the DTFT of b as a function of the one of a and of the displacement (c,d)

$$\hat{b}(v_x, v_y) = \sum_{n,m} a(n+c, m+d)e^{-i2\pi(v_x n + v_y m)}$$

$$\hat{b}(v_x, v_y) = \sum_{n,m} a(n, m)e^{-i2\pi(v_x(n-c) + v_y(m-d))}$$

(variable change)

$$\hat{b}(v_x, v_y) = \hat{a}(v_x, v_y)e^{i2\pi c v_x}e^{i2\pi d v_y}$$

Question 4

Compute the ratio between the DTFTs

With question 3 :

$$\frac{\hat{b}(v_x, v_y)}{\hat{a}(v_x, v_y)} = e^{i2\pi c v_x} e^{i2\pi d v y} = e^{i2\pi(c v_x + d v_y)}$$

Question 5

Final signals : In which hypotheses the ratio between a's and b's DFT is a linear phase signal ?

To have :

$$\frac{\hat{A}(k_x, k_y)}{\hat{B}(k_x, k_y)} = e^{\frac{i2\pi(c k_x + d k_y)}{N^2}}$$

We need to have :

$$\sum_{n=0}^{N-1} \sum_{m=0}^{N-1} a(n,m) e^{\frac{-i2\pi(n k_x + m k_y)}{N^2}}$$
$$= \sum_{n=c}^{N-1+c} \sum_{m=d}^{N-1+d} a(n,m) e^{\frac{-i2\pi(n k_x + m k_y)}{N^2}}$$

So

$$\sum_{n=0}^{c-1} \sum_{m=0}^{d-1} a(n,m) e^{\frac{-i2\pi((n-c)k_x + (m-d)k_y)}{N}} = \sum_{n=N}^{N-1+c} \sum_{m=N}^{N+d-1} a(n,m) e^{\frac{-i2\pi(n k_x + m k_y)}{N}}$$

$$\sum_{n=0}^{c-1} \sum_{m=0}^{d-1} (a(n,m) - a(n+N, m+N) e^{-i2\pi(k_x + k_y})) e^{\frac{-i2\pi(n k_x + m k_y)}{N^2}} = 0$$

If A is a periodical function of period N, then the ration of the DTFTs is a linear phase signal.

Question 7

Compute the phase of the ratio between the DFTs by using the function angle2D. Display the phase ϕ What is the behavior of ϕ? Is it coherent with the theoretical result ? Why ?

$$\phi = \frac{ck_x + dk_y}{N^2}$$

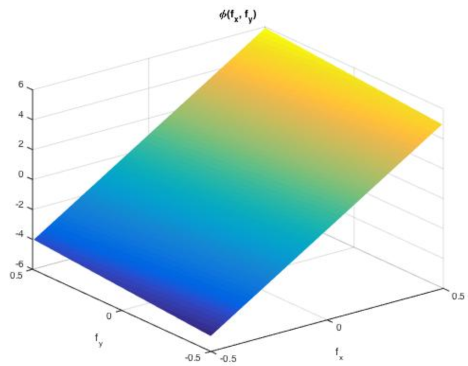

Figure 9: The phase ϕ

It increases with k_x and k_y so it's coherent.

Question 8

What is the theoretical value of the gradient ? Is it what you observe ? Comment

The gradient of ϕ is :

$$\nabla\phi = (c/N^2, d/N^2)$$

We obtain :
$\nabla\phi = (10, 1)$ and that corresponds to the displacement (c,d) with the spatial frequency factor.

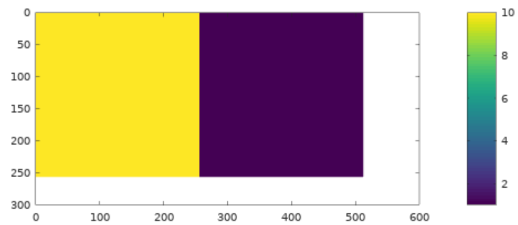

Figure 10: Median of the gradient

Question 10

Compare the results of the 2 methods and comment.

The second method that look for the best LS plane fitting ϕ also gives the good displacement (10,1), if there is no noise on the original image. Let's try with some noise.

Question 11

What happens if you add some noise to the image before translation ? You can do this by setting a non-nul value of sigma in line 6, for example, sigma=2.

When we add some noise, the results don't change, and both methods succeed to find the good displacement (c,d).

3.2 Indirect parametric estimation

Question 3

Use the median of MVF components to estimate the translation. Comment the result.

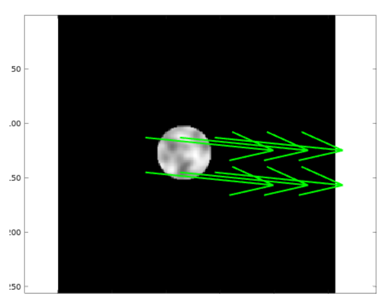

Figure 11: MVF Block Matching (SSD)

All the blocks that form the backgound are uniform and black. Therefore, the block-matching algorithm considers that they didn't move at all. Consequently, most of the motion vectors are (0,0) and the median is (0,0). So this method doesn't suceed to find the good displacement (10,10). We must add some noise on the original image (before the translation) to make sure that even the blocks of the background are able to find their pair in the translated image.

Question 4

Redo the previous questions when there is some noise, sigma=2. Comment and justify.

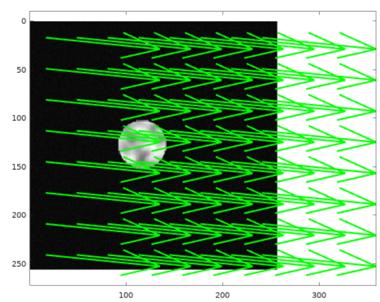

Figure 12: MVF Block Matching with noise on the original image (SSD)

As we said at question 3, adding some noise in the image before the translation will enable to better find the matching blocks. Here, we see that even the background blocks are considered to have moved. We get what we wanted : the median of the MVF is (10,1) ie the real displacement.

Question 5

Finally, compare all the parametric estimation methods when the translation values are not integer.

We set (c,d)=(10.3,1.2)

Method	bx	by
Grad+Med	10.155	1.059
LS plane	9.881	0.928
BM SSD	10.000	1.000
HS SSD	9.987	0.999
Actual	10.300	1.200

Figure 13: Results of the different methods

We see that the best algorithm is the one that use the median of the gradient of ϕ. The problem with the block matching method is that it returns only integers for c and d, so our median is an integer. If we want to be more precise, we are forced to use direct parametric estimation methods.

Question 6

Redo the questions of this part with HS instead of block matching

Without noise

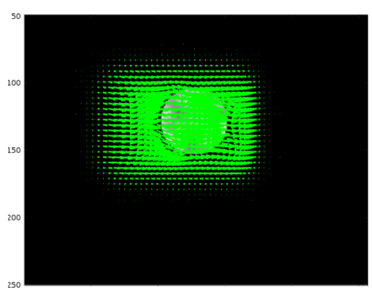

Figure 14: Results of the different methods

There is the same problem than with block-matching so (c,d)median=(0,0)

With sigma=2

The (c,d)found=(9.983,0.998) very close to (10,1). As we take the mean over.

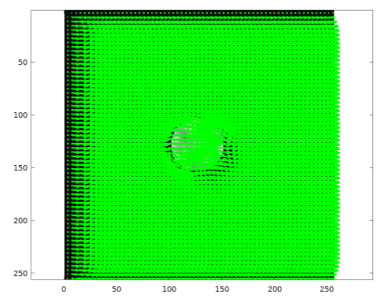

Figure 15: Results of the different methods

With non integer displacement (10.3,1.2)

The displacement found is exactly the same that with (10,1) because for every pixel, each MV is the same. We can't be more precise than 1 pixel in the calculus. So again we can say that parametric estimation methods are more profitable in this kind of cases.