ECE462 – Lecture 20

Video Compression Technologies

1. Video of data consists of multiple image frames per second (typically 15-30 frames/sec). This corresponds to enormous bit rate requirements → compression is of outmost importance.
2. While still image frames exhibit spatial redundancy video frames exhibit significant "temporal redundancy" which can be exploited for compression

→Code the difference between adjacent video frames (small variations in time)

→ More effective approach: Use motion estimation and compensation between frames

Motion Compensation (MC)

Three basic steps:

1. Motion estimation (motion vector)
2. Motion compensation (motion prediction)
3. Prediction Error

Process:

→Image is divided into macro blocks of size NxN pixels (usually N=16 for luminance images)

→Reference image: previous or future frame

→Target frame: current frame

→Compare each macroblock of the target frame to each macroblock of the reference frame using a “matching criterion”

→ Find best match between two macroblocks and then calculate the displacement between them (motion vector)

→The difference between two “matched macroblocks” is called the prediction error

→ Forward prediction: Reference frame in past error

→ Backward prediction: Reference frame error in future

EX

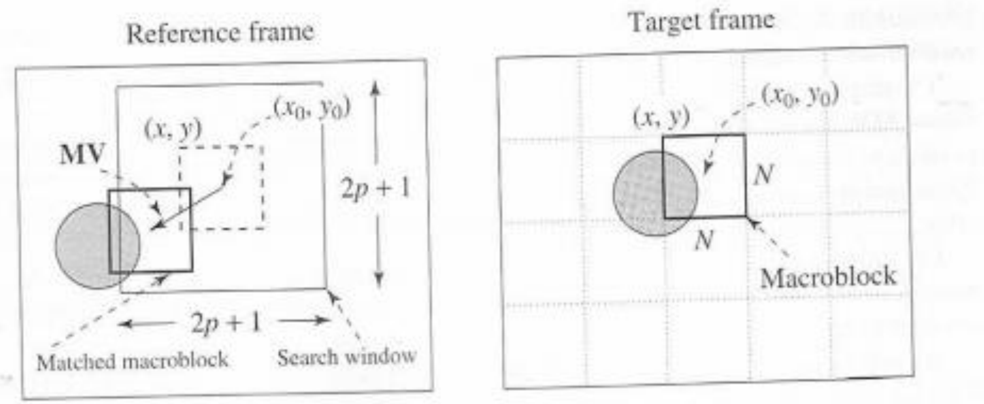


FIGURE 10.1: Macroblocks and motion vector in video compression: (a) reference frame; (b) target frame.

Compression Strategy: Given a number of vide frames

→Code first frame

→Code only motion vectors and difference macroblocks in all subsequent frames

Matching Criteria

→Mean absolute difference (MAD)

MAD(i,j) = $\frac{1}{N^2}\sum_{k=0}^{N-1}\sum_{l=0}^{N-1}|T(k,l)-R(k+i,l+j)|$

Where, T(k,l) : Macroblock in Target frame
R(k+i, l+j) : ” in Reference frame

→Mean Square Difference Cartesian (MSD)

MSD(i,j) = $\frac{1}{N^2}\sum_{k=0}^{N-1}\sum_{l=0}^{N-1}(T(k,l)-R(k+i,l+j))^2$

<u>Objective:</u> Find (i,j) so that MAD (i,j) or MSD(i,j) is minimum.

e.g. Let (u,v)={(i,j)/MAD(i,j) is minimum}

Then, the motion vector (k,l) $\longrightarrow$ (k+u,l+v)

Note:

To find the minimum each macroblock in the target frame must be compared to all possible macroblocks in Reference frame

<u>Thus:</u>

For frames of size LxL there are approximately L^2 possible macroblock positions.

MAD calculations $\sim N^2$/block

So the order of calculations $O(L^2 N^2)$

So for 1 sec of videos at 30 frames /sec we require

# of macroblocks in target frame X # of MAD operations per macroblock X # of frames

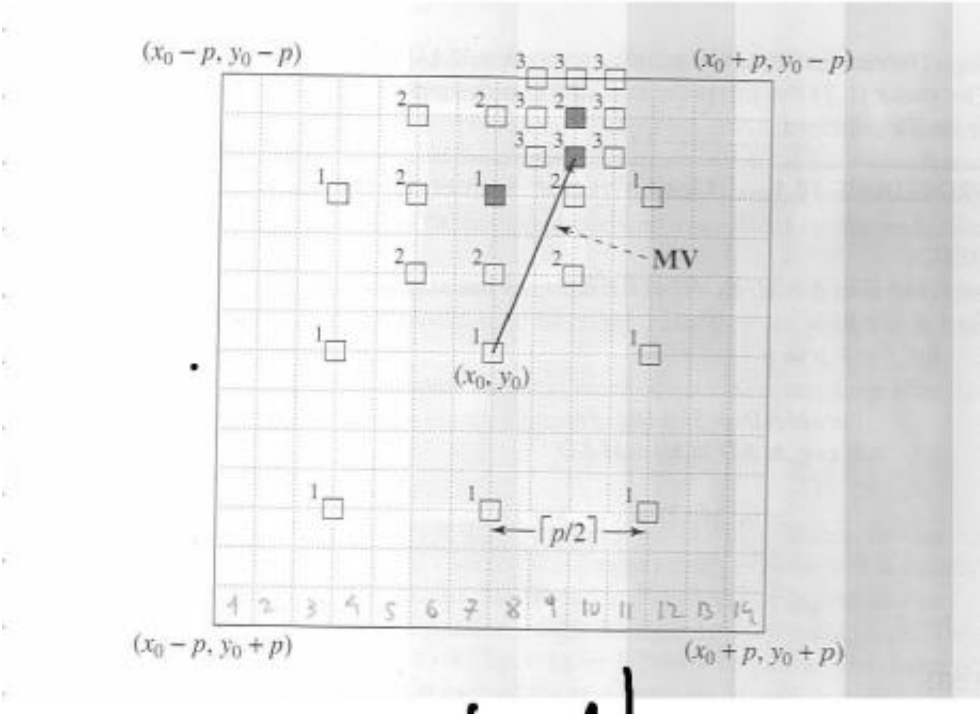
$$\sim \quad \frac{L^2}{N^2} \quad X \quad L^2 N^2 \quad X \ 30$$

$\sim 0(30.L^4)$ This makes real time video encoding very difficult

To reduce the high complexity of full search a number of computationally more efficient procedures are applied in practice such as

- Logarithmic search approach
- Hierarchical search approach

<u>Logarithmic Search</u>



Logarithmic Search

- Specify only 9(nine) macroblocks in interest area of size PxP in reference frame
- Find the best matching which now defines the center of a new search area of size $\frac{P}{2} \times \frac{P}{2}$
- Define nine macroblocks in new search area and continue the process ,…
- Stop after M steps where

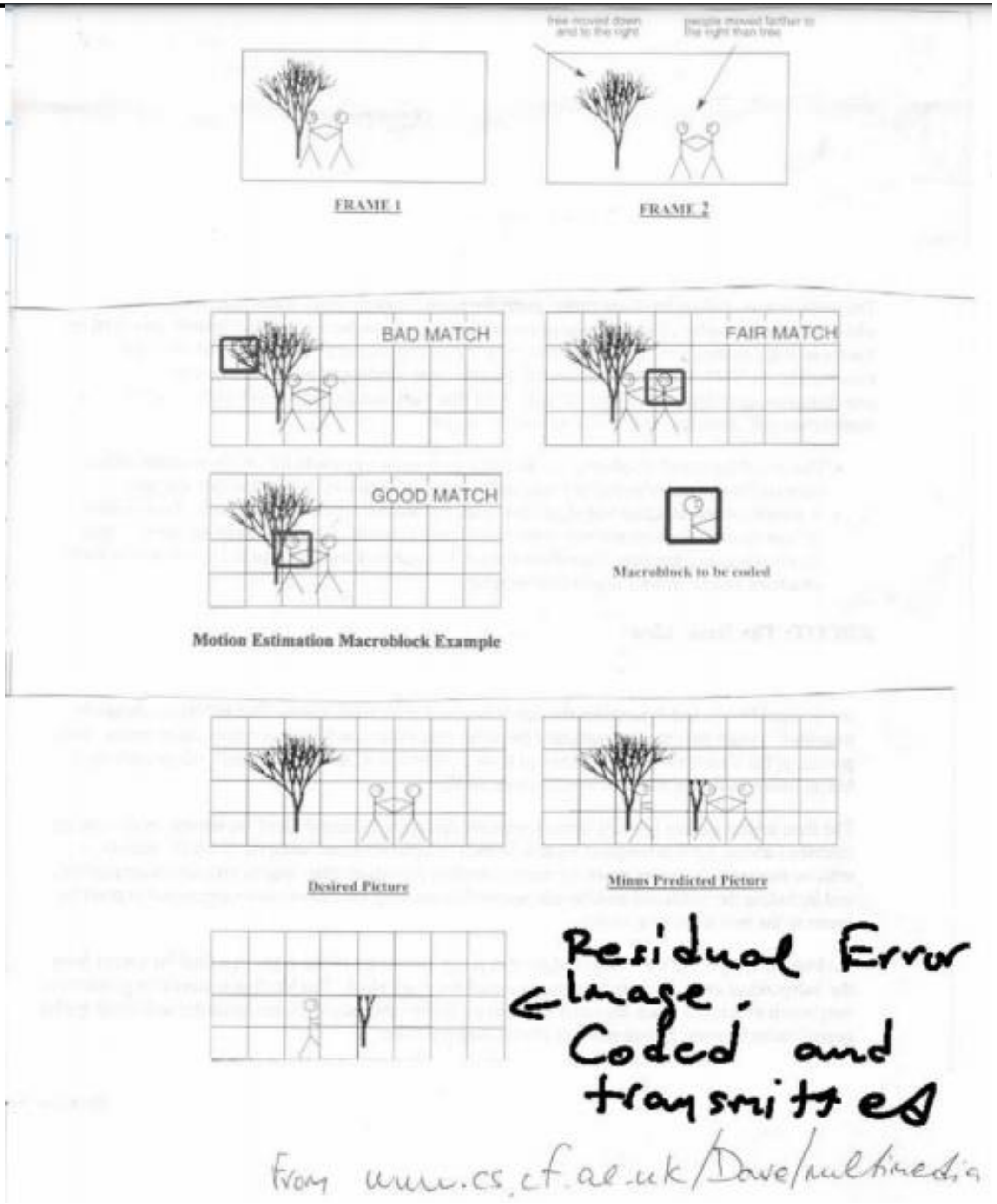
$$M \sim \log_2 P$$

and the complexity for a search area of PxP is reduced to $\sim O(\log_2 P \,.\, N^2)$ from $\sim O(P^2N^2)$ that is required by full search.

Hierarchical Search

- Decompose the image in various resolution levels
- Match macroblocks at a low resolution level
- Refine matching in progressively higher resolution levels.

Example of Motion Estimation Prediction

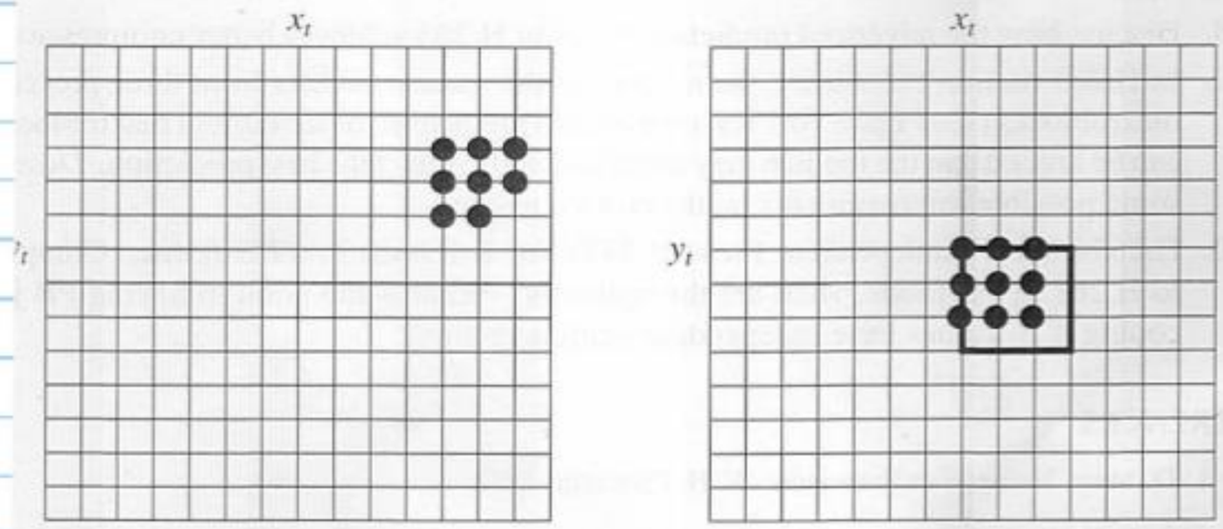


Working example on motion estimation

Work out the following problem of 2D Logarithmic Search for motion vectors in detail (see Figure 10.14).

The target (current) frame is a P-frame. The size of macroblocks is 4×4. The motion vector is $\mathbf{MV}(\Delta x, \Delta y)$, in which $\Delta x \in [-p, p]$, $\Delta y \in [-p, p]$. In this question, assume $p \equiv 5$.

The macroblock in question (darkened) in the frame has its upper left corner at (x_t, y_t). It contains 9 dark pixels, each with intensity value 10; the other 7 pixels are part of the background, which has a uniform intensity value of 100. The reference (previous) frame has 8 dark pixels.



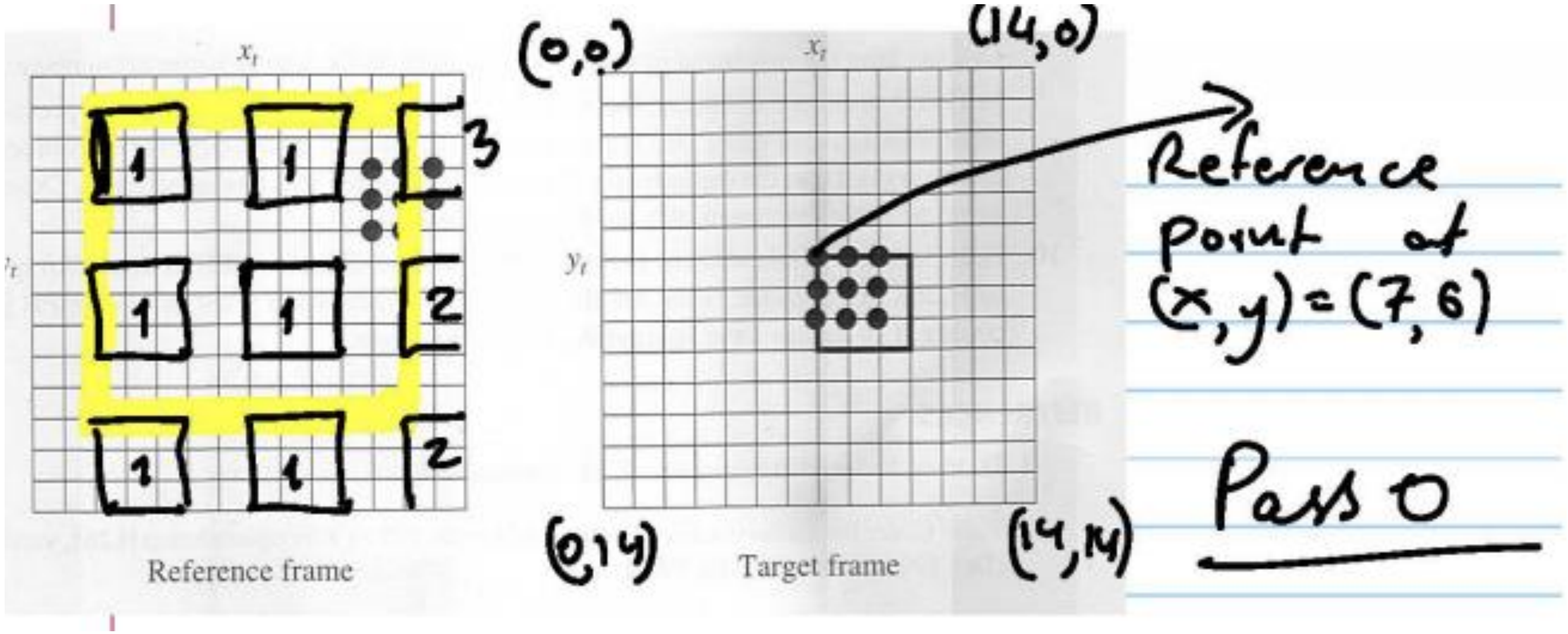
Reference frame

Target frame

Fig. 10.14

● Pixel with intensity value 10

Other background (unmarked) pixels all have intensity value 100



With p=5 the search are (-5,5)x(-5,5) is too large for this size image (incomplete blocks)

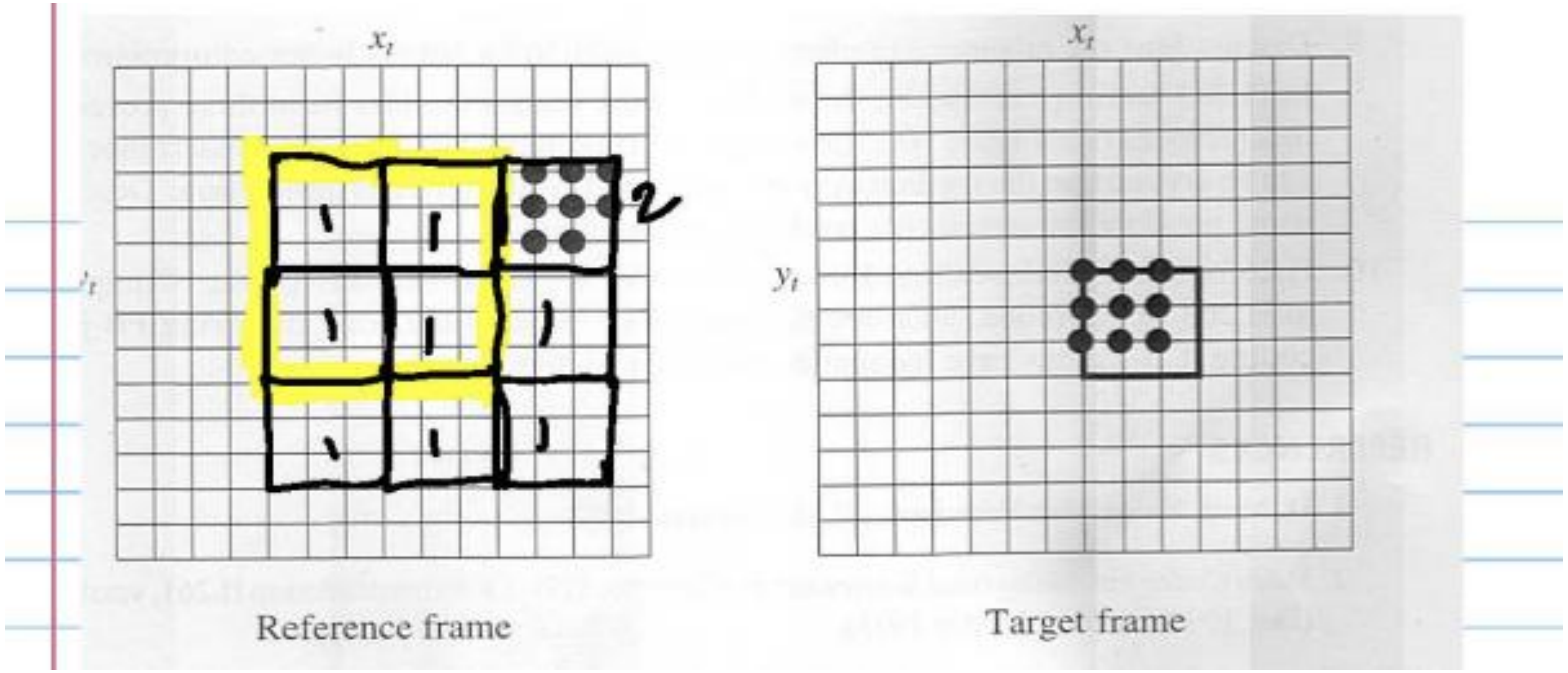
→For blocks marked 1 in the reference frame

MAD(u,v) = $\frac{1}{16}\sum_{k=0}^{3}\sum_{l=0}^{3}|T(i,j)-R(u+i,v+j)|$ = $\frac{1}{16}$. 9|100-10| = $\frac{9x90}{16}$ = $\frac{810}{16}$

→ For blocks marked 2, MAD(u,v) = $\frac{9x90}{16}$ = $\frac{810}{16}$

→For blocks marked 3, MAD(u,v) = $\frac{9x90}{16}$ = $\frac{810}{16}$

So this pass is inconclusive



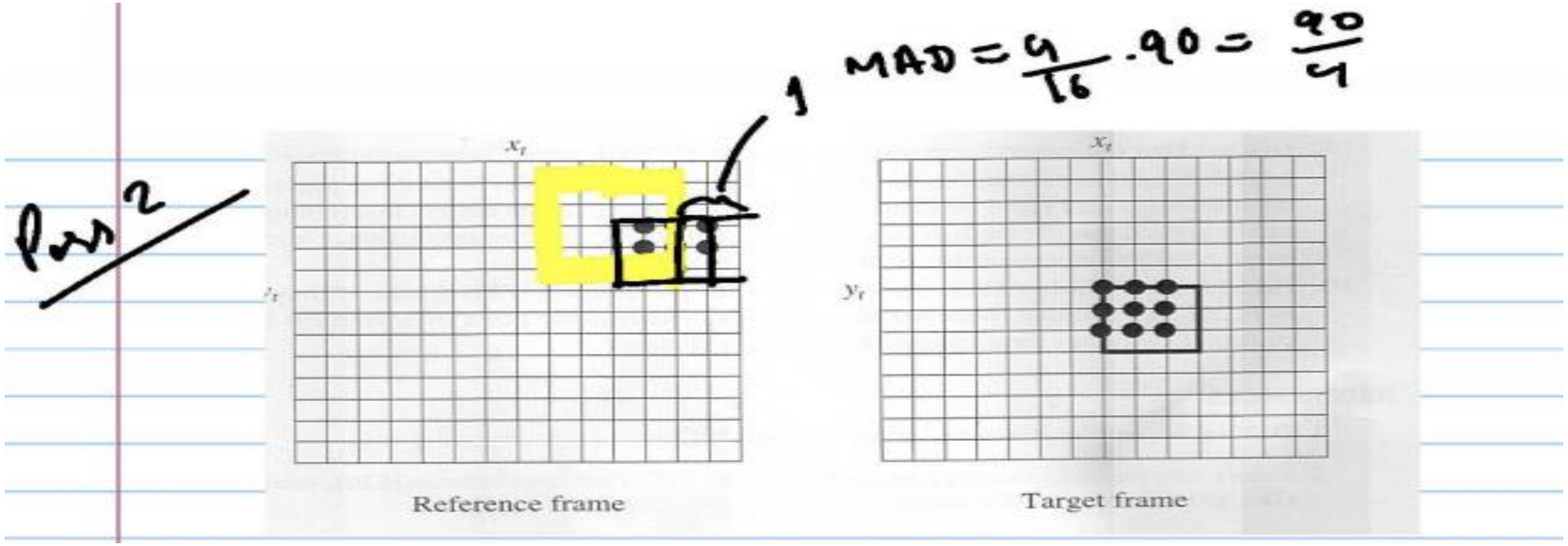
<u>Pass1</u>

Let us reduce P by 2 and consider an area of $[\frac{P}{2}] \text{x} [\frac{P}{2}]$ = 3x3 with the same reference point

blocks #1 → MAD(u,v) = $\frac{9x90}{16}$ = $\frac{810}{16}$

blocks #2 → MAD(u,v) = $\frac{5x90}{16}$ = $\frac{450}{16}$

So the new reference point is at (x,y) = (10,3)



Pass2

Reference point at : (x,y) = (10,3)

Search area: $[\frac{\mathrm{P}}{4}]\mathrm{x}\ [\frac{\mathrm{P}}{4}]$ = 2x2

The smallest MAD is obtained for the incomplete block with reference point (12,3)

Pass3

Search area: $[\frac{\mathrm{P}}{8}]\mathrm{x}\ [\frac{\mathrm{P}}{8}]$ = 1x1

In this case one of the blocks will be at the correct location, giving MAD = $\frac{1x90}{16}$ = 5.625

And $(\Delta x, \Delta y) = (4{,}3)$