22 July 2012 -- Computer Architectures –- part 1/2

Last Name, First Name, Matricola **.....**

Run Length Encoding is a popular image lossless compression algorithm. It is based on the technique that consecutive sequences of repeated strings could be efficiently represented in a shorter way.

In particular,

RULE # 1: it is a method where repeating values (bytes) are replaced by a number (byte) and the byte to repeat.

Example:

AF AF AF AF AF AF AF AF

(8 bytes of same value AF) could be replaced by a counter (with the highest bit set to 0) followed by the byte to repeat:

07 AF (07+1=8 repeats followed by the value to repeat... 0 repeats would not be very useful).

RULE # 2: Other sequences without internal repetitions like:

01 02 03 04 05 06 07

should then be replaced by a count with the highest bit set to 1, to indicate that the 'count' bytes following the counter must be copied, not repeated, i.e. with reference to the previous example:

86 01 02 03 04 05 06 07 would mean: copy the next 06 + 1=7 (a count of 0 would be a bit useless!!) bytes to the output.

Let us assume that an image is stored in an array

```
IMAGE DB N DUP (?)
N EQU 1000
```

has to be compressed into the array

COMPRESSED DB 2*N DUP (?)

by using the RLE algorithm above. It is requested to write a 8086 assembly program to compute the compression and store it in the array COMPRESSED, according to the following rules:

- Item 1: the compression is computed according to rule #1 only, i.e. without recognizing (as for rule #2) the sequences of different symbols.
  Example:
  IMAGE: 01 01 01 02 03 01 03 03 03 03 03 03 02
  COMPRESSED: 02 01 (i.e. the symbol 01 repeated 02+1= 3 times),
  00 02 (i.e. the symbol 02 repeated 00+1= 1 time),
  00 03 (i.e. the symbol 03 repeated 00+1= 1 time),
  00 01 (i.e. the symbol 01 repeated 00+1= 1 time),
  05 03 (i.e. the symbol 03 repeated 05+1= 6 times),
  00 02 (i.e. the symbol 02 repeated 00+1= 1 time).
- Item 2: the compression is computed according to both rule #1 and rule #2 only, i.e. taking care of both repetitions of the same symbol and of sequences of different symbols.
  Example:

22 July 2012 -- Computer Architectures –- part 1/2

Last Name, First Name, Matricola **……**

IMAGE: 01 01 01 02 03 01 03 03 03 03 03 03 02
COMPRESSED: 02 01 (i.e. the symbol 01 repeated 02+1= 3 times),
82 02 03 01 (i.e. a sequence of non repeated symbols made of 02+1= symbols, i.e. 02, 03, 01),
05 03 (i.e. the symbol 03 repeated 05+1= 6 times),
80 02 (i.e. a sequence of non repeated symbols made of 00+1= symbols, i.e. 02)

It is known in advance that the maximum length of a sequence of repeated symbols is no longer than 128 symbols, as well as that the maximum length of a sequence without internal repetitions is also 128 symbols.

Bonus items:

BONUS #1:
The compression ratio is computed as: C=(1- number_byte_Compressed/number_byte_Image) %.
Example: for number_byte_Compressed= 6 and number_byte_Image= 10, we have C=(1-0,6)% = 40%

BONUS #2:
Given the position I of a symbol, find what is that symbol by appropriately scanning COMPRESSED.

Points for each completed item (as uncompleted items could be not evaluated)
Item 1. = up to 22 points;
Item 2. = up to 28 points;

Bonus #1., i.e. to compute the percentage of compression C DB (?) according to the previous definition = up to 2 points.
Bonus #2., i.e. to find the symbol in the I-th position of a symbol in the original image, by scanning its COMPRESSED coding (I DW (?)) = up to 3 points.

It is not required to provide the optimal (shortest, most efficient, fastest, …) solution, but a working and clear solution. The input-output part is not necessary in the class developed solution, but its implementation is mandatory to be discussed at oral exam.

Please use carbon copy and retain one copy for home implementation and debug. Please provide your classroom submitted solution with several explanatory and significant comments. When coming to oral discussion, please mark on your "classroom" copy all modifications. Please also provide an error-free and running release of the solution, as well as with its printed list of instructions. Please consider that the above are necessary but not sufficient requirements to success the exam, since the final evaluation will be based on a number of parameters.

FAILURE TO ACCOMPLISH ALL PREVIOUS NECESSARY REQUIREMENTS WILL CAUSE NO-QUESTION-ASKED AND IMMEDIATE REJECTION.