

# Κεφάλαιο 1

## Εισαγωγή

### 1.1 Περιγραφή του Προβλήματος

Οι κωδικοποιητές βίντεο είναι αναγκαίοι για να μειώσουν τον τεράστιο αριθμό δεδομένων που έχει ένα βίντεο. Σε οπτικά σήματα με ισχύ μεγαλύτερη των 35dB δεν είναι εύκολο να παρατηρηθεί διαφορά από το ανθρώπινο μάτι. Έτσι επιλέγεται να εισαχθεί ομοιόμορφος θόρυβος για να επιτευχθεί μεγαλύτερος λόγος συμπίεσης θυσιάζοντας την ποιότητα. Η κωδικοποίηση βίντεο υλοποιείται τόσο σε επίπεδο λογισμικού αλλά και υλικό. Οι προσωπικοί υπολογιστές συνήθιζαν να χρησιμοποιούν λογισμικό για την αναπαραγωγή βίντεο αλλά τα τελευταία χρόνια η αύξηση των αναλύσεων του βίντεο (1080p, 4K) καθιστά αδύνατη ή πολύ δαπανηρή την αποκωδικοποίηση ενός συμπιεσμένου βίντεο από έναν επεξεργαστή γενικής χρήσης. Έτσι βλέπουμε υλικό ειδικά σχεδιασμένο (hardware accelerators) για κωδικοποίηση/αποκωδικοποίηση βίντεο που βοηθά τον κύριο επεξεργαστή.

Χρονιά	Κωδικοποιητής	Διάφορες Χρήσεις	Λόγος Συμπίεσης
1993	MPEG-1	VCD	26:1
1995	MPEG-2	DVD	31:1
1999	MPEG-4	DivX, XVID	200:1
2003	H.264	BluRay, DVB-TS	400:1
2013	H.265	next generation H.264	550:1

Πίνακας 1.1: Κυριότεροι κωδικοποιητές βίντεο

Σήμερα όπου υπάρχει βίντεο υπάρχει και κωδικοποίηση. Όπως φαίνεται στον πίνακα ?? με το πέρασμα του χρόνου οι κωδικοποιητές γίνονταν ολοένα και πιο αποτελεσματικοί αυξάνοντας όμως κατά πολύ την πολυπλοκότητα τους. Πλέον μπορούμε με λίγα δεδομένα να έχουμε βίντεο καλής ποιότητας σε πολύ μεγάλες αναλύσεις. Ενδεικτικά για ένα ασυμπίεστο βίντεο ανάλυσης DVD (720x480) με

διάρκεια 20sec και ρυθμό ανανέωσης 30Hz χρειαζόμαστε χώρο 296MB.Κάνοντας συμπίεση με τον H.264 θα έχουμε ένα αρχείο 0.8MB.

## 1.2 Συμβολή της Εργασίας

Στην εργασία αυτή γίνεται χρήση της κβαντοποίησης διανυσμάτων στον κωδικοποιητή H.264. Είναι μια τεχνική συμπίεσης δεδομένων με απώλειες (lossy compression), όπου έχει δοκιμαστεί ελάχιστα παλαιότερα στο βίντεο και φαίνεται πως μπορεί να δώσει λύσεις στον λόγο συμπίεσης καθώς επίσης και στην μείωση της πολυπλοκότητας του αποκωδικοποιητή. Ο τρόπος που μειώνεται η πολυπλοκότητα μπορεί να μετατρέψει ακριβά ενεργοβόρα κυκλώματα σε φθηνές μνήμες χωρίς απαιτήσεις ισχύος.

## 1.3 Διάρθρωση της Διπλωματικής Εργασίας

Στο Κεφάλαιο 1 παρουσιάζεται το ψηφιακό βίντεο, με ποιές μορφές αυτό αποθηκεύεται, πώς οι κωδικοποιητές το αντιμετωπίζουν και ποιές είναι οι κύριες τεχνικές που χρησιμοποιούνται για την συμπίεση του.

Στο Κεφάλαιο ?? γίνεται μία σύντομη εισαγωγή στον τομέα της Θεωρίας Πληροφοριών. Είναι αναγκαίο να εξηγηθεί τι είναι η πληροφορία καθώς και η εντροπία αυτής. Επίσης θα εξηγηθεί αναλυτικά ο κύριος αλγόριθμος clustering k-means καθώς και οι όποιες βελτιστοποιήσεις προέκυψαν στην πορεία της παρούσας διπλωματικής.

Στο Κεφάλαιο ?? παρουσιάζονται τα βήματα που έγιναν για την παραγωγή των codebooks και την επιπλέον κατηγοριοποίηση τους με στόχο την μείωση της εντροπίας.

Στο Κεφάλαιο ?? παρουσιάζεται η επέμβαση που έγινε στον H.264 έτσι ώστε να χρησιμοποιεί τα codebooks για να κάνει την κωδικοποίηση/αποκωδικοποίηση.

Στο Κεφάλαιο ?? παρουσιάζονται τα αποτελέσματα του VQ-H.264 και συγκρίνονται με τις επιδόσεις του JM-H.264.

## Κεφάλαιο 2

# Ψηφιακό βίντεο και τεχνικές συμπίεσης

### 2.1 Εισαγωγή

Σε αυτό το κεφάλαιο γίνεται μία εισαγωγή στους τρόπους που οι κωδικοποιητές βίντεο διαχειρίζονται τα δεδομένα και τους τρόπους που χρησιμοποιούν για να τα συμπίεσουν. Επίσης εδώ θα μας γίνει ξεκάθαρο γιατί το βήμα της κβαντοποίησης (εισαγωγή θορύβου) είναι αναγκαία για να έχουμε τόσο μεγάλους λόγους συμπίεσης

### 2.2 Συστατικά στοιχεία του βίντεο

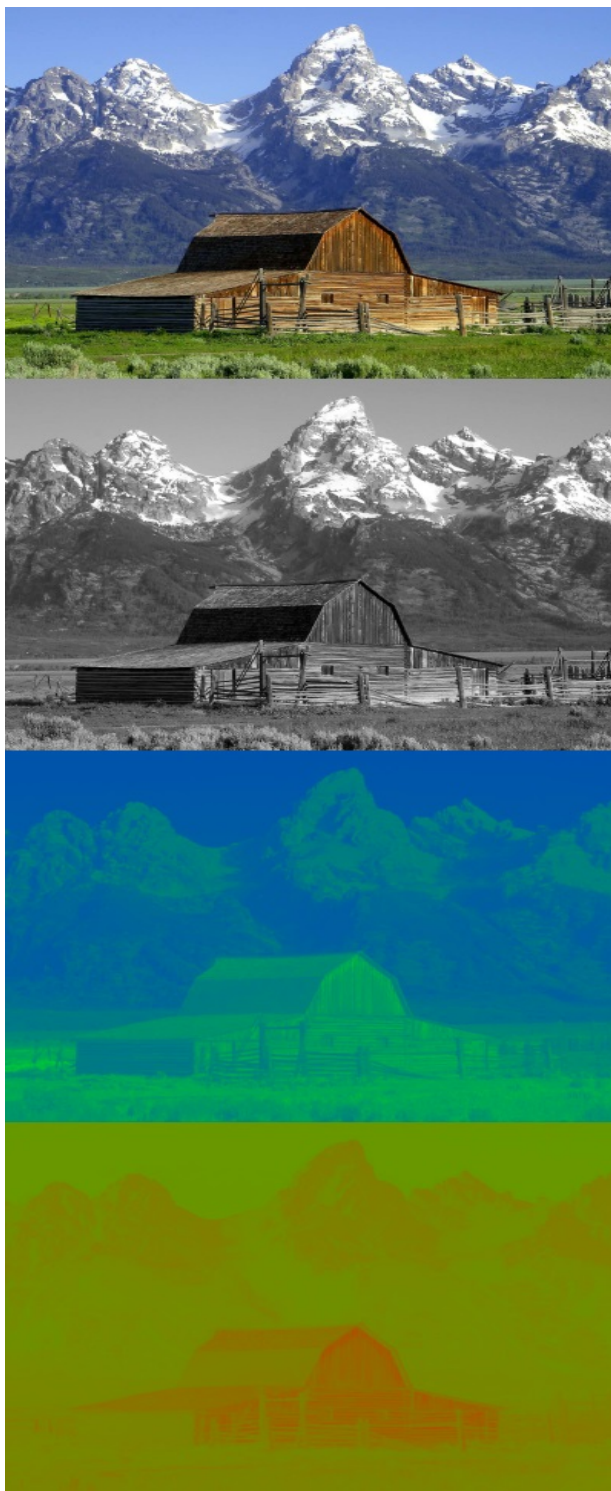
Το ψηφιακό βίντεο αποτελείται από μία σειρά καρτέ που αναπαράγονται με σταθερό ρυθμό (συνήθως 25Hz ή 30Hz). Το κάθε καρτέ απεικονίζεται σε ένα χώρο χρωμάτων που ονομάζεται YUV, όπου το Y είναι η φωτεινότητα και το U,V το χρώμα. Η κάθε συνιστώσα στο ανθρώπινο μάτι φαίνεται στο Σχήμα 1.1.

Σε κάθε σημείο του βίντεο(pixel) αντιστοιχίζεται μια τέτοια τιμή YUV. Υπάρχουν διάφοροι τρόποι που γίνεται αυτή η αντιστοίχιση με κυριότερες αυτές του YUV420 και YUV444. Η πρώτη μας λέει πως κάθε για κάθε τέσσερα pixel υπάρχουν τέσσερις τιμές Y μία τιμή U και μια τιμή V ενώ η δεύτερη πως για κάθε τέσσερα pixel υπάρχουν τέσσερις τιμές για την κάθε συνιστώσα π.χ για ένα καρτέ ανάλυσης 720\*480 τύπου YUV420 έχουμε  $720 * 480 = 345600$  στοιχεία Y,  $720 * 480 / 4$  στοιχεία U και  $720 * 480 / 4$  στοιχεία V αρά σύνολο 518400. Ο τρόπος αποθήκευσης του βίντεο γίνεται πάντα ανά καρτέ και υπάρχουν δύο τρόποι, ο packed και ο planar. Στον πρώτο το YUV για κάθε pixel αποθηκεύεται με την σειρά ενώ στον δεύτερο και επικρατέστερο αποθηκεύεται πρώτα όλο το Y και μετά ακολουθούν τα U,V.

Κάθε pixel έχει και ένα βάθος (bitdepth), δηλαδή ποιο είναι το εύρος τιμών που παίρνει. Οι σύγχρονοι κωδικοποιητές υποστηρίζουν 8-14bits έτσι συνεχίζοντας το παράδειγμα μας από την προηγούμενη παράγραφο και έχοντας συνολικά 518400 στοιχεία για ένα καρτέ και υποθέτοντας ότι το bitdepth 8 τότε συνολικά χρειαζόμαστε  $518400 * 8 \text{bits} \approx 0.5 \text{MB}$ . Το πιο κοινό bitdepth που χρησιμοποιείται από τους σημερινούς κωδικοποιητές είναι αυτό τον 8bits.

## 2. Ψηφιακό βίντεο και τεχνικές συμπίεσης

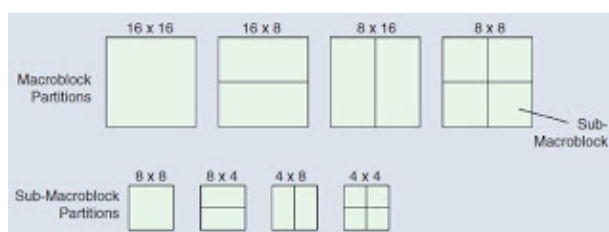
---



Σχήμα 2.1: Στην πρώτη εικόνα φαίνεται ένα καρέ με όλες τις συνιστώσες ενώ παρακάτω ξεχωριστά το YUV για το ίδιο καρέ.

## 2.3 Οργάνωση των pixels από τους κωδικοποιητές

Η πλειοψηφία των σημερινών κωδικοποιητών ομαδοποιούν τα pixels. Η πιο συνηθισμένη ομαδοποίηση είναι αυτή του macroblock, όπου το κάθε καρέ διαιρείται σε μικρά πλακίδια σταθερής διάστασης  $N \times N$  pixels όπου συνήθως  $N=16$ . Έτσι το καρέ του παραδείγματος μας αποτελείται από  $518400/(16*16) = 2025$  macroblocks. Το κάθε macroblock μπορεί να σπάσει σε blocks και το κάθε block σε subblocks όπως φαίνεται στο Σχήμα 1.2.



Σχήμα 2.2: Διαμερισμός του macroblock στον H.264.

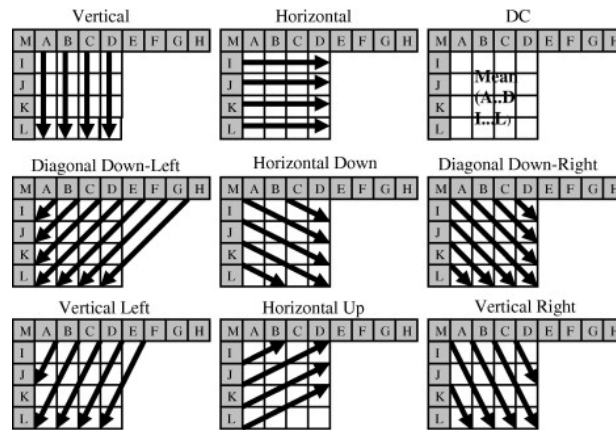
Μετά τον χωρισμό σε macroblocks οι κωδικοποιητές ομαδοποιούν τα καρέ σε Intra και Inter όπου τα τελευταία χωρίζονται σε P και B. Όλες οι παραπάνω κατηγοριοποιήσεις έχουν να κάνουν με τον μηχανισμό του Motion Estimation όπου ο encoder δημιουργεί διαφορές pixels (residuals). Τα residuals έχουν μικρότερη ενεργεία οπότε είναι οι διάφοροι μετασχηματισμοί που εφαρμόζονται αργότερα στην αλυσίδα μας δίνουν μικρότερη ενεργεία και επιπλέον περισσότερα μηδενικά.

- Intra η αλλιώς I είναι τα καρέ που χρησιμοποιούν πληροφορία μόνο από το τρέχων καρέ για να βγάλουν τα residuals. Αυτό γίνεται απλά παίρνοντας τις τιμές των γειτονικών block και εφαρμόζοντας μια πράξη π.χ μέσος όρος όπου το αποτέλεσμα αυτής αφαιρείται από τις τιμές των pixel του τρέχων block. Υπάρχουν διάφορες πράξεις που μπορούν να γίνουν όπως φαίνεται και στο Σχήμα 1.3 και δοκιμάζονται κάθε φορά όλες μέχρι να καταλήξουμε στην καλύτερη. Η απόδοση της συμπίεσης είναι η χειρότερη σε αυτήν την κατηγορία αλλά χωρίς αυτά το βίντεο δε θα μπορούσε να ξεκινήσει γιατί δεν θα είχαμε όλη την πληροφορία.
- P καρέ είναι αυτά τα οποία ψάχνουν το καλύτερο block για να κάνουν διαφορές από κάποιο αριθμό προηγούμενων καρέ όπως φαίνεται στο Σχήμα 1.4
- B καρέ είναι αυτά τα οποία ψάχνουν το καλύτερο block για να κάνουν διαφορές από κάποιο αριθμό προηγούμενων αλλά και επόμενων καρέ όπως φαίνεται στο Σχήμα 1.4. Αυτά έχουν την καλύτερη απόδοση συμπίεσης αλλά η πολυπλοκότητα αποκωδικοποίησης είναι η μεγαλύτερη.

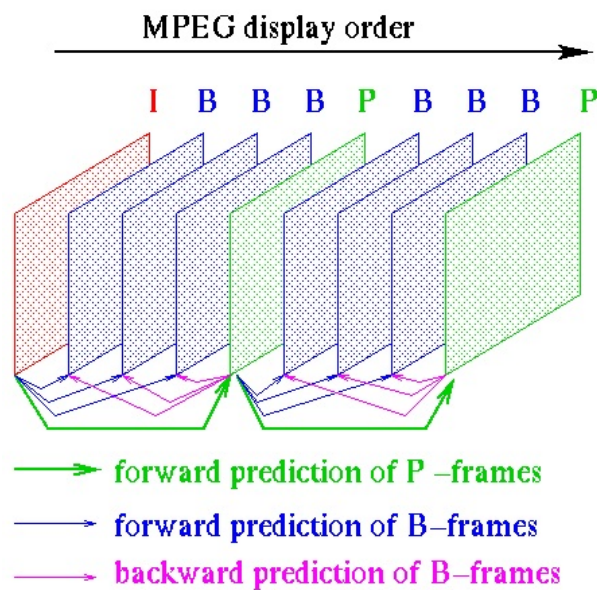
Ένα επιπλέον στοιχείο των encoder είναι το group of pictures (GOP) όπου αυτό χαρακτηρίζει την σειρά με την οποία τα είδη των καρέ τοποθετούνται. Το GOP

## 2. Ψηφιακό βίντεο και τεχνικές συμπίεσης

Ξεκινάει από ένα I-frame συνεχίζει με P,B και περιοδικά έρχεται ένα I. Το GOP εξαρτάται από την εφαρμογή και μπορεί να διαφέρει κατά πολύ. Ένα τυπικό φαίνεται στο Σχήμα 1.4



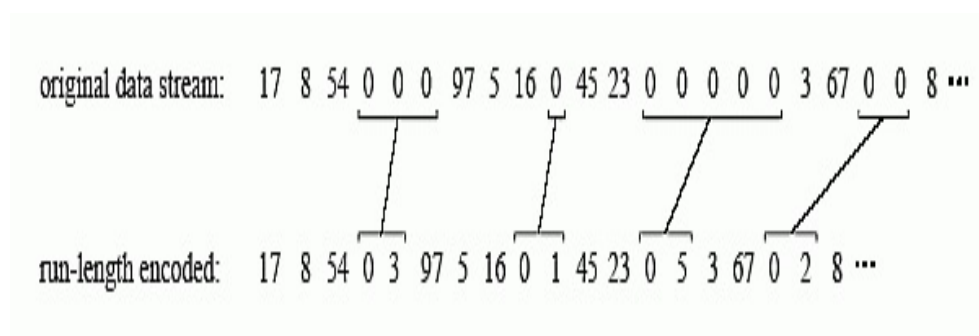
Σχήμα 2.3: Τρόποι που χρησιμοποιούνται στον H.264 για Intra prediction



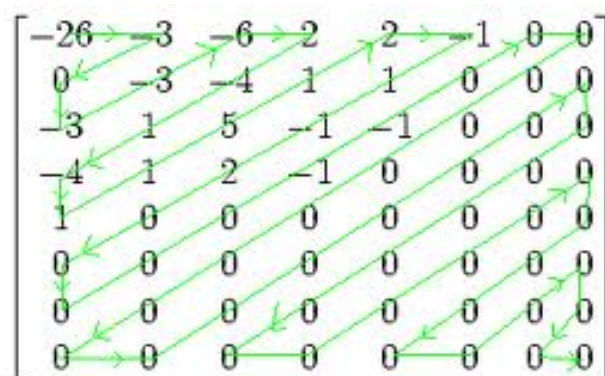
Σχήμα 2.4: Inter prediction στον mpeg και απεικόνιση ενός ενδεικτικού GOP

## 2.4 Μετασχηματισμός

Αφού ο encoder έχει κατασκευάσει τα residuals για το macroblock τότε το επόμενο βήμα είναι ένας μετασχηματισμός οποιού συνήθως είναι ο DCT. Αυτό συμβαίνει γιατί ο μετασχηματισμός μας επιστρέφει πολλά 0 αν το σήμα μας είναι χαμηλής ενέργειας, όπως συμβαίνει με τα residuals. Έτσι χρησιμοποιώντας κάποιο ειδικό σύμβολο μπορούν να απεικονιστούν πολλά μηδενικά του μετασχηματισμού και να μειώσουμε το data rate μας όπως φαίνεται στο Σχήμα 1.5. Η σάρωση του macroblock δεν γίνεται γραμμικά αλλά με την μέθοδο του zigzag όπως στο Σχήμα 1.6 γιατί έτσι έχει παρατηρηθεί πως έχουμε παραπάνω μηδενικά στην σειρά.



Σχήμα 2.5: Παράδειγμα ένωσης μηδενικών



Σχήμα 2.6: Σάρωση ZigZag



## 2.5 Κβαντοποίηση

Η κβαντοποίηση είναι το σημείο που σε κάθε encoder εισάγεται ο θόρυβος, θα μπορούσε να θεωρηθεί και ο DCT,iDCT λόγω ακρίβειας αλλά εκεί το σφάλμα είναι παρά πολύ μικρό. Μετά τον μετασχηματισμό τα residuals έχουν αντικατασταθεί με συντελεστές οι οποίοι είναι πραγματικοί αριθμοί. Η τακτική που εφαρμόζεται είναι να γίνει ακέραια διαίρεση ο συντελεστής της κάθε συχνότητας με έναν σταθερό αριθμό διαφορετικό για κάθε συχνότητα Σχήμα 1.7. Επομένως για τον δεδομένο πίνακα κβαντοποίησης με τους συντελεστές του Σχήματος 1.8 έχουμε τα αποτελέσματα του Σχήματος 1.9 έτοιμα προς κωδικοποίηση με κάποιον entropy encoder και εγγραφή.

$$Q = \begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix}.$$

Σχήμα 2.7: Παράδειγμα πίνακα κβαντοποίησης για μετασχηματισμό 8x8

$$G = \begin{matrix} & & & \xrightarrow{u} & & & & \\ \begin{bmatrix} -415.38 & -30.19 & -61.20 & 27.24 & 56.13 & -20.10 & -2.39 & 0.46 \\ 4.47 & -21.86 & -60.76 & 10.25 & 13.15 & -7.09 & -8.54 & 4.88 \\ -46.83 & 7.37 & 77.13 & -24.56 & -28.91 & 9.93 & 5.42 & -5.65 \\ -48.53 & 12.07 & 34.10 & -14.76 & -10.24 & 6.30 & 1.83 & 1.95 \\ 12.12 & -6.55 & -13.20 & -3.95 & -1.88 & 1.75 & -2.79 & 3.14 \\ -7.73 & 2.91 & 2.38 & -5.94 & -2.38 & 0.94 & 4.30 & 1.85 \\ -1.03 & 0.18 & 0.42 & -2.42 & -0.88 & -3.02 & 4.12 & -0.66 \\ -0.17 & 0.14 & -1.07 & -4.19 & -1.17 & -0.10 & 0.50 & 1.68 \end{bmatrix} & \downarrow v. \end{matrix}$$

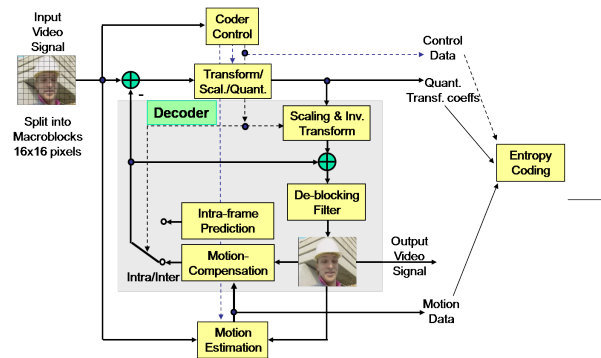
Σχήμα 2.8: Παράδειγμα συντελεστών DCT 8x8

$$B = \begin{bmatrix} -26 & -3 & -6 & 2 & 2 & -1 & 0 & 0 \\ 0 & -2 & -4 & 1 & 1 & 0 & 0 & 0 \\ -3 & 1 & 5 & -1 & -1 & 0 & 0 & 0 \\ -3 & 1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Σχήμα 2.9: Παράδειγμα συντελεστών DCT 8x8



Παρακάτω βλέπουμε την δομή του H.264 οπου επίτηδες έχω παραλείψει να αναφερθώ σε κάποια κομμάτια τους αφού δεν κρίνονται αναγκαία για την κατανόηση αυτής της διπλωματικής.



Σχήμα 2.10: H.264 structure

## 2.6 Αποκωδικοποίηση βίντεο

Για την αποκωδικοποίηση κάνουμε όλα τα βήματα προς τα πίσω.

- Μετά την entropy decoding πολλαπλασιάζουμε τα νούμερα με τα στοιχεία του πίνακα κβαντοποίησης. Έτσι παίρνουμε τους DCT συντελεστές αλλά λανθασμένους λόγω της ακέραιας διαίρεσης.
- Έπειτα κάνουμε iDCT και παίρνουμε τα residuals.
- Τελευταίο βήμα είναι να κάνουμε το Motion Decomposition βρίσκοντας τα δύο block που χρησιμοποιήθηκαν για να παραχθούν τα residuals και τα προσθέτουμε για να πάρουμε τα ανακατασκευασμένα pixels.