

Τεχνητή Όραση

ΤΠ 7004

Μάθημα 2^ο: Εικόνες και επεξεργασία

Δρ. Θάνος Δεμίρης

Περιεχόμενα 2^{ου} Μαθήματος

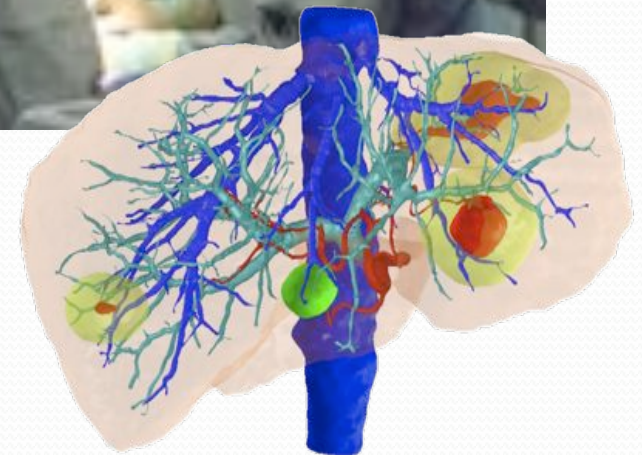
- Επανάληψη 1^{ου} μαθήματος και βασικών εννοιών
- Απλά μοντέλα κάμερας (πρώτη επαφή...)
- Σύντομη αναφορά στο φως και την ηλεκτρομαγνητική ακτινοβολία
- Εικόνες: Ορισμός και τύποι

Repetitorium

Σύντομη επανάληψη του 1^{ου} μαθήματος

Εφαρμογή στην ιατρική

- Ανίχνευση ανατομικών και παθολογικών σχηματισμών σε ακολουθίες εικόνων
- Δημιουργία τρισδιάστατων αναπαραστάσεων
- Ανανέωση εικόνων κατά τη διάρκεια εγχειρήσεων



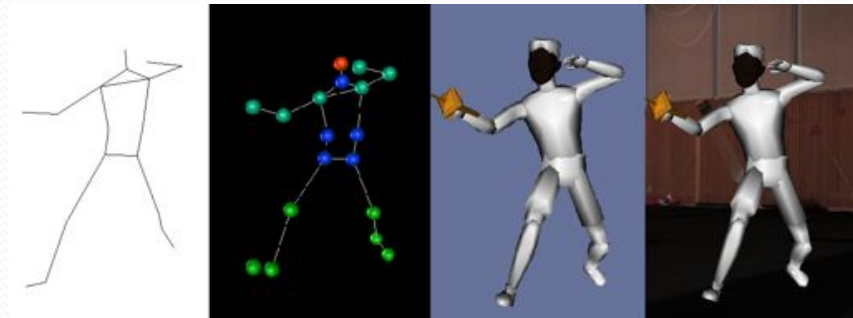
Εφαρμογή στην προβολή πολιτιστικής κληρονομιάς

- Επαυξημένη πραγματικότητα
- Προσδιορισμός θέσης και οπτικού πεδίου με μεγάλη ακρίβεια
- Προβολή συνθετικών αντικειμένων στο πραγματικό περιβάλλον



Εφαρμογές στη διαδραστική τηλεόραση

- Ανίχνευση αθλητών
- Προσδιορισμός της κίνησης των αθλητών
- Προβολή της κίνησης σε τρισδιάστατα μοντέλα
- Σύνθεση του συνθετικού με το πραγματικό
- Γεωμετρικά ορθές προβολές στοιχείων στο χώρο



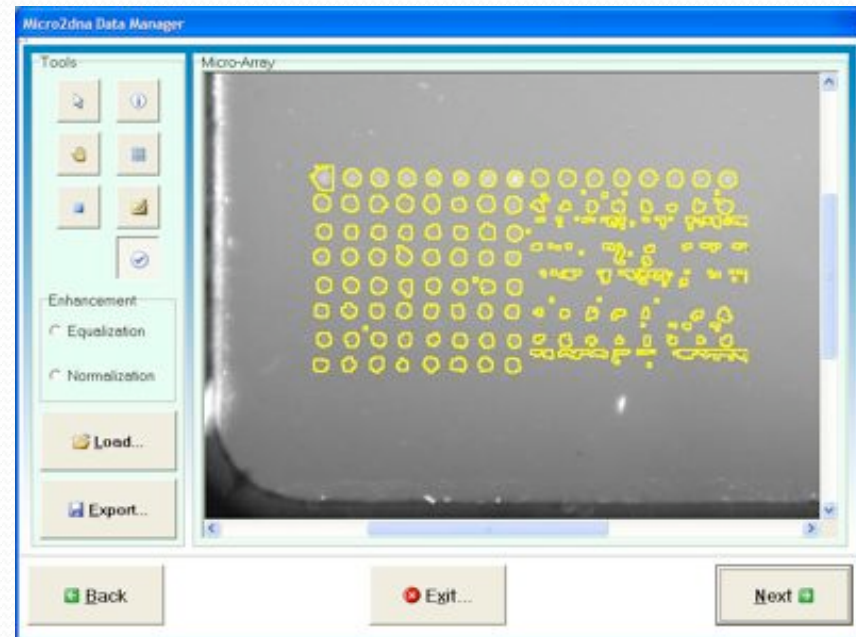
Εφαρμογή στην αρχιτεκτονική

- Φωτο-ρεαλισμός
βασισμένος σε ανάλυση
σκηνής και εντοπισμό
όλων των πηγών φωτός
- Προβολή συνθετικών
στον πραγματικό χώρο
με τη σωστή σκίαση

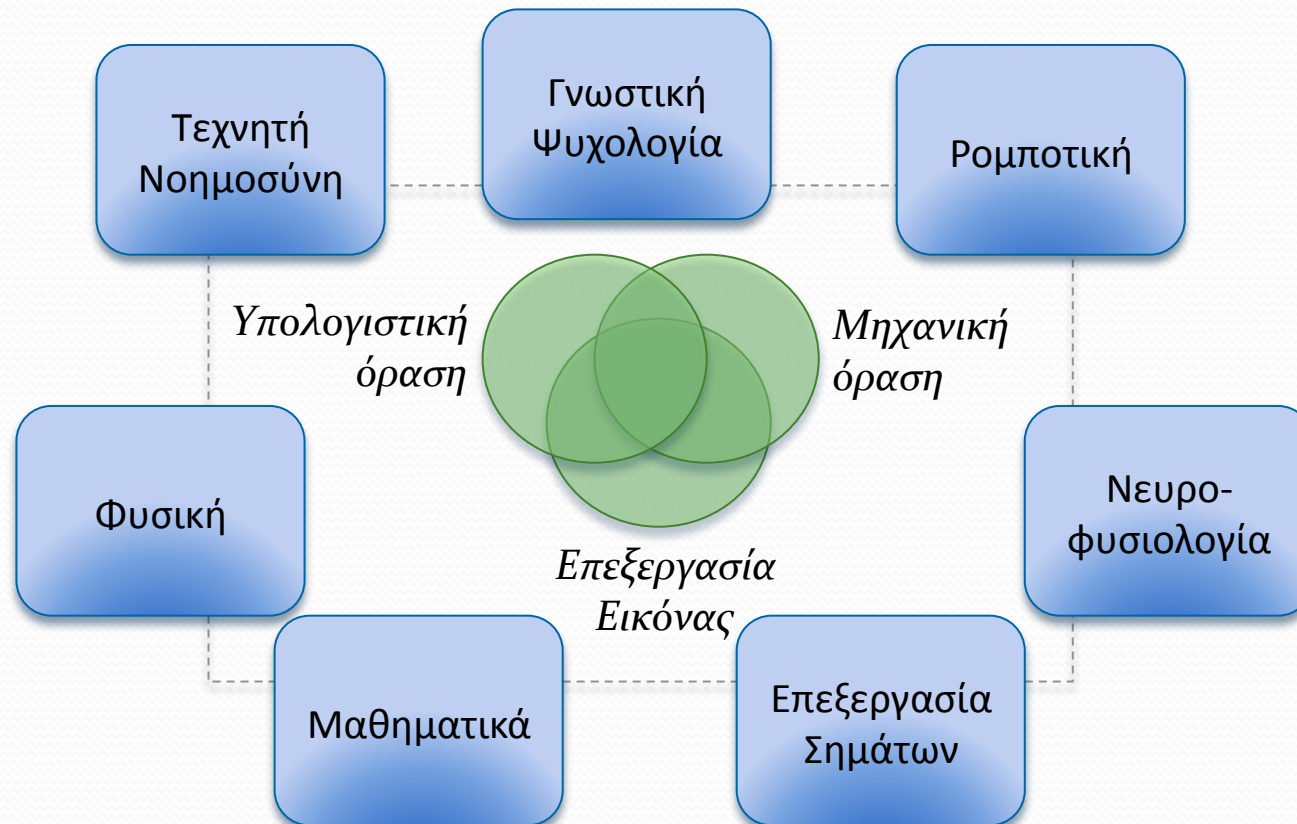


Εφαρμογές στη γενετική

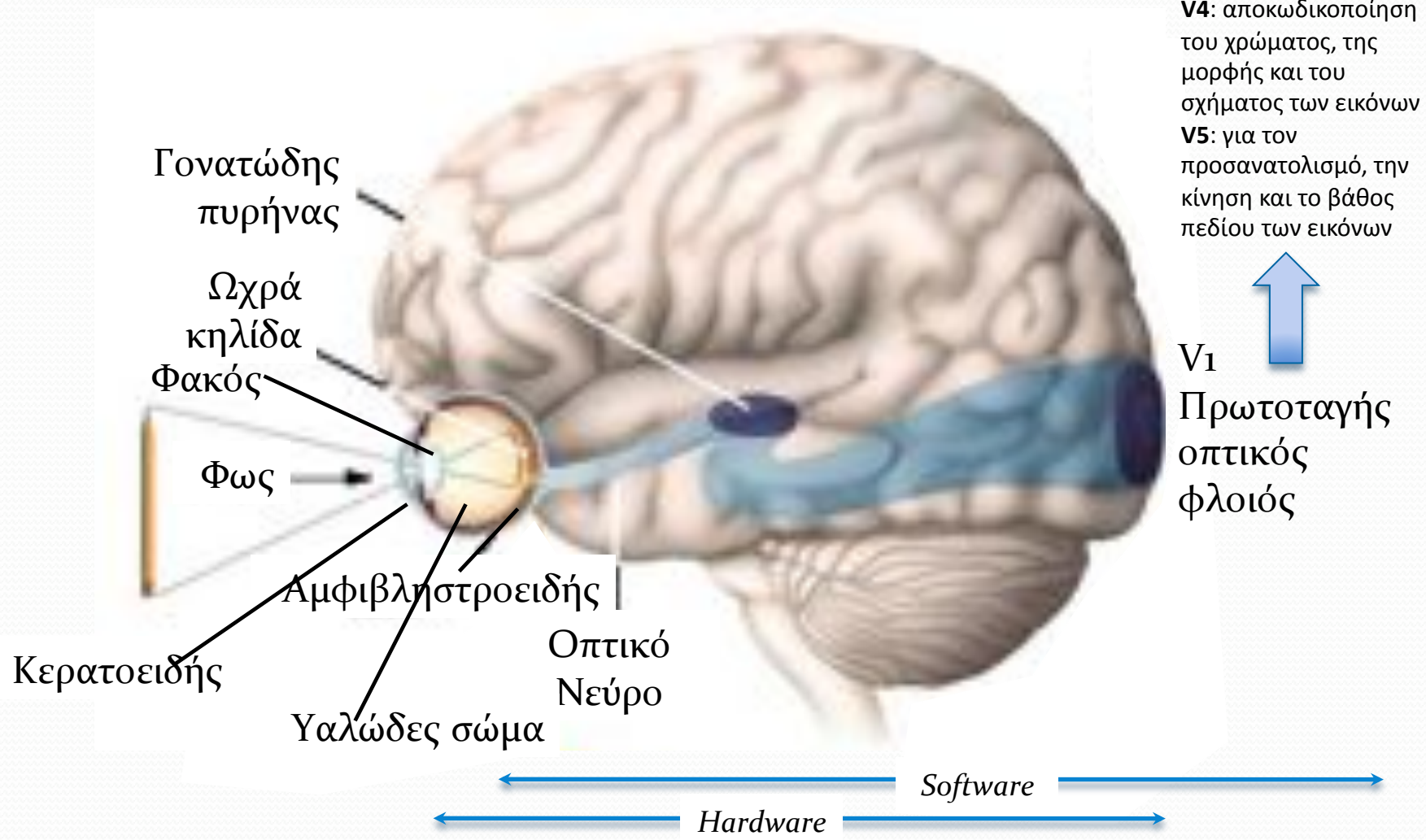
- Ανάλυση επαναληπτικών προτύπων
- Προσέγγιση τιμών σημείων εντοπισμού μεταλλάξεων



Η θέση της τεχνητής όρασης στην έρευνα και τεχνολογία



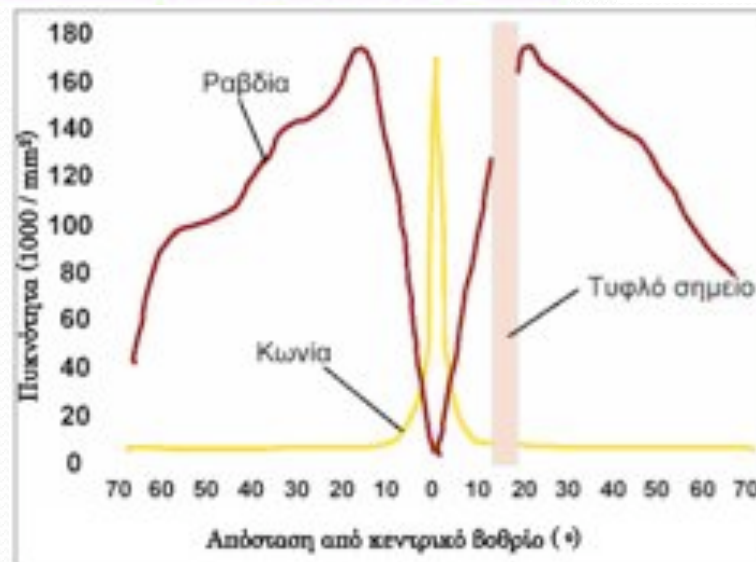
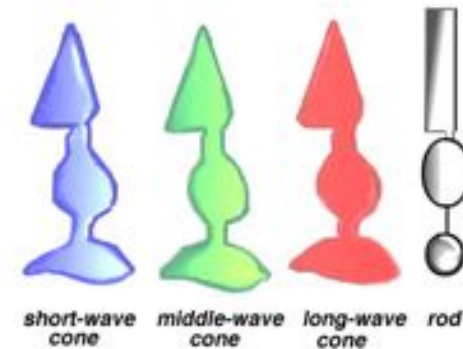
Όραση στα θηλαστικά



Ανώτερα κέντρα
εγκεφάλου αναγκαία
για την αντίληψη π.χ.
V4: αποκωδικοποίηση
του χρώματος, της
μορφής και του
σχήματος των εικόνων
V5: για τον
προσανατολισμό, την
κίνηση και το βάθος
πεδίου των εικόνων

Φωτοϋποδοχείς: Κωνία και ραβδία

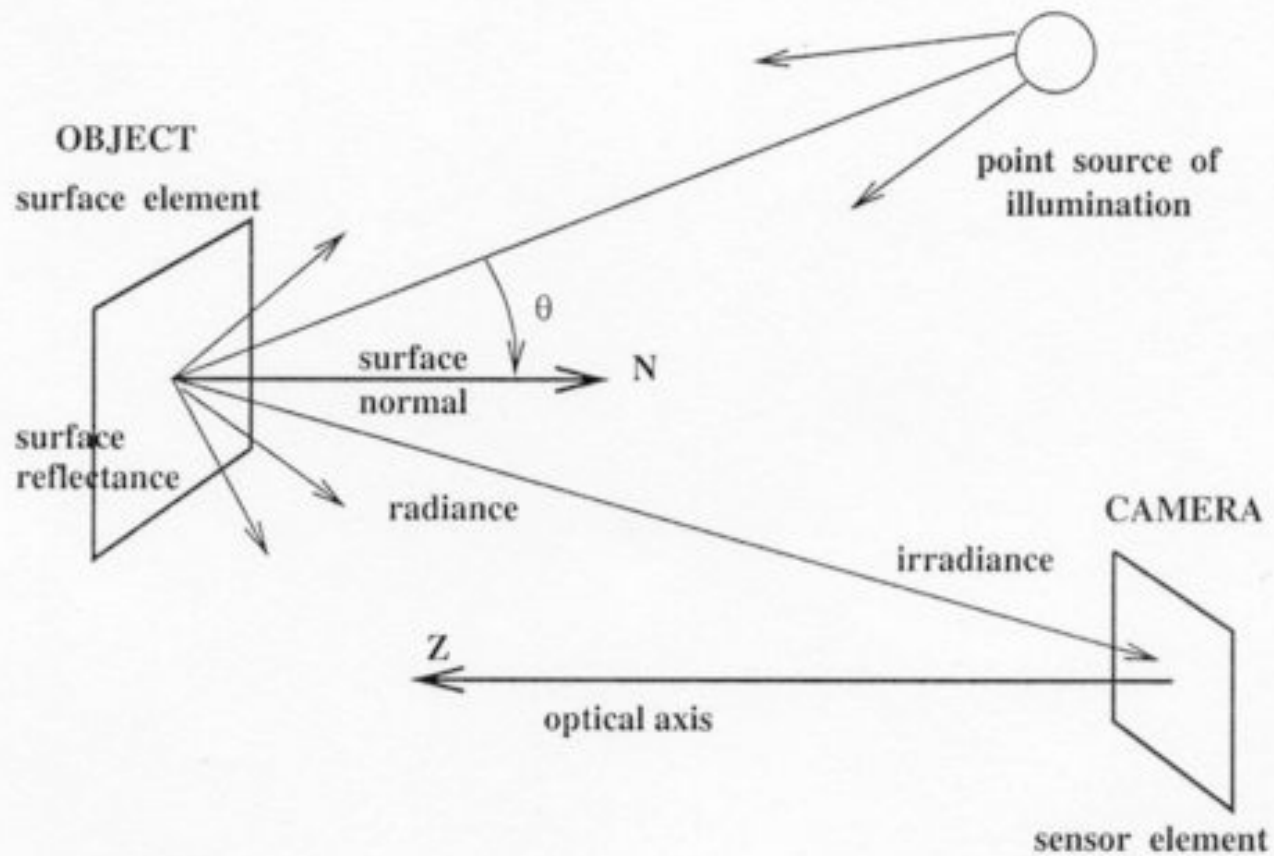
Χρώμα	Περιοχή μηκών κύματος (nm)	Περιοχή συχνοτήτων (Hz)
Κόκκινο	~ 630–700 nm	~ 476–429 $\times 10^{12}$ Hz
Πορτοκαλί	~ 590–630 nm	~ 510–476 $\times 10^{12}$ Hz
Κίτρινο	~ 560–590 nm	~ 535–510 $\times 10^{12}$ Hz
Πράσινο	~ 500–560 nm	~ 600–535 $\times 10^{12}$ Hz
Μπλε	~ 440–500 nm	~ 680–600 $\times 10^{12}$ Hz
Ιώδες (Μοβ)	~ 400–440 nm	~ 750–680 $\times 10^{12}$ Hz



Κάμερες και Εικόνες

Ορισμός και τύποι

Δημιουργία εικόνας

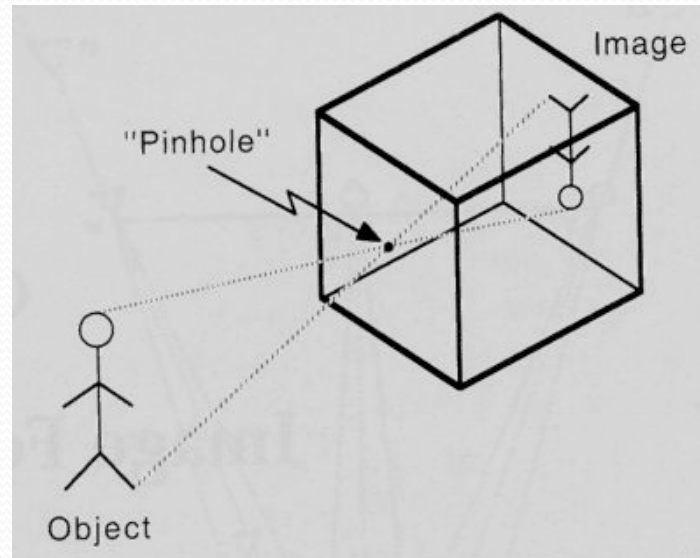


Οι 2 βασικές ιδιότητες στο σχηματισμό εικόνων

- Γεωμετρία
 - Σε ποιο σημείο της εικόνας θα προβληθεί ένα σημείο του χώρου
- Φωτισμός
 - Η φωτεινότητα ενός σημείου της εικόνας αποτελεί συνάρτηση των πηγών φωτός και των ιδιοτήτων επιφανειών

Απλή κάμερα οπής (Pinhole)

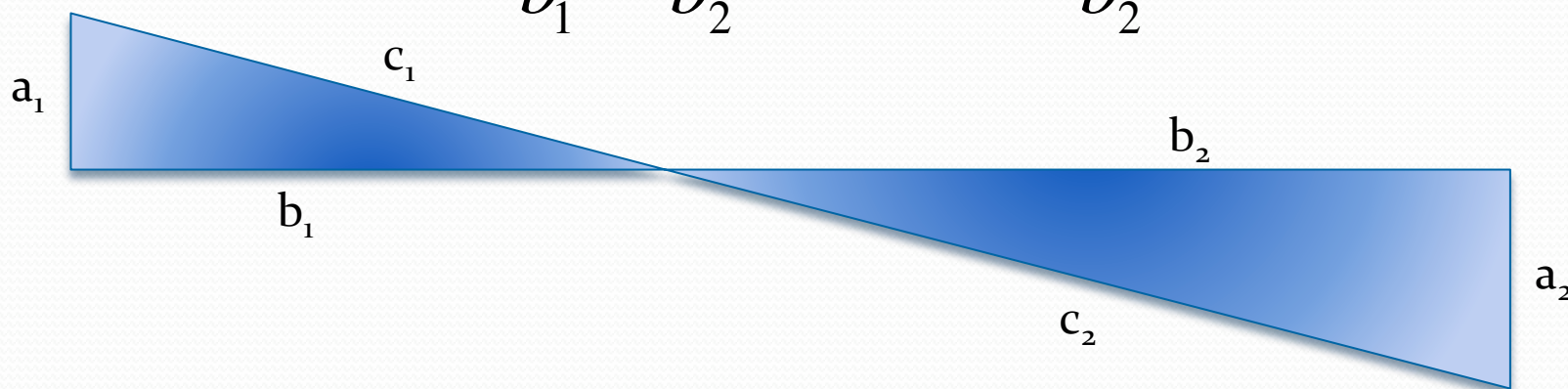
- Ο πιο απλός μηχανισμός προβολής μίας 3-διάστατης σκηνής σε δύο διαστάσεις
- Ευθείες γραμμές φωτός διαπερνούν μία οπή και εμφανίζονται (ανάποδα) στην επιφάνεια προβολής



Όμοια τρίγωνα

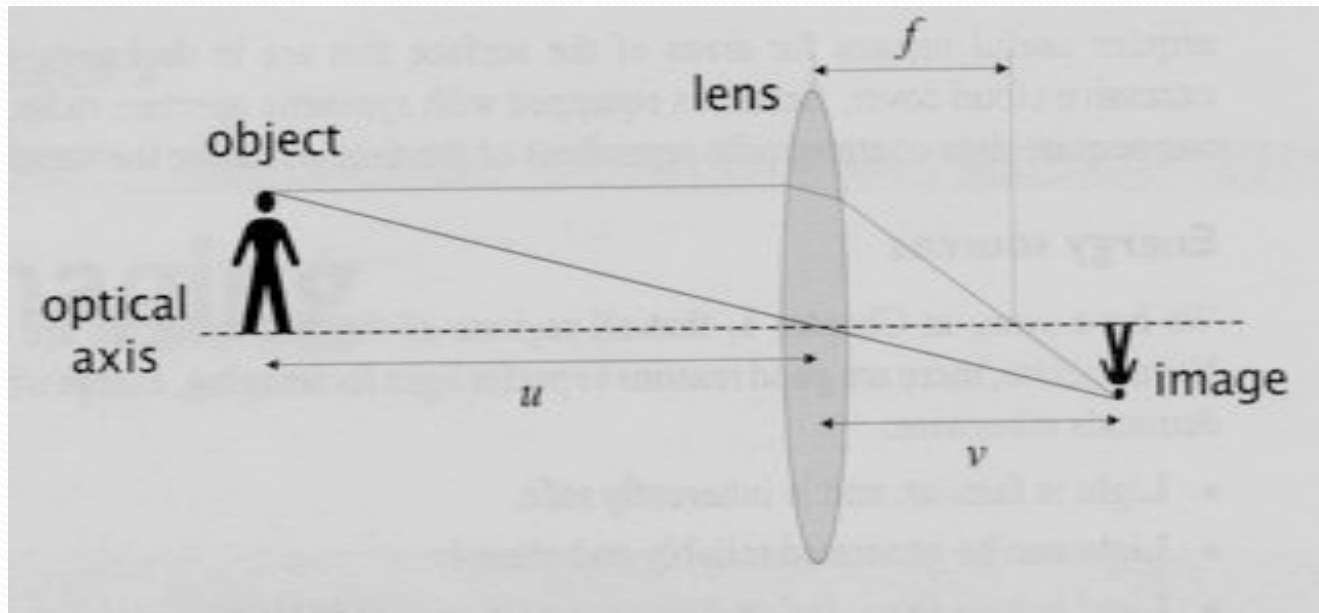
- Δύο τρίγωνα είναι όμοια όταν έχουν ακριβώς τις ίδιες γωνίες και κατά συνέπεια το ένα είναι ένα πολλαπλάσιο του άλλου.
- Τότε ισχύει:

$$\frac{c_1}{b_1} = \frac{c_2}{b_2} \Leftrightarrow c_1 = c_2 \frac{b_1}{b_2}$$



Οπτική της κάμερας με φακό

- Οι φακοί χρησιμοποιούνται ώστε να συγκεντρώνουν δέσμες φωτός από τη σκηνή στο αντίστοιχο σημείο της εικόνας



Οπτική της κάμερας

- Παράμετροι φακού
 - Τύπος φακού
 - Εστιακή απόσταση
 - Οπτικό πεδίο
- Φωτομετρικές παράμετροι
 - τύπος, ένταση και κατεύθυνση φωτισμού
 - ιδιότητες αντανάκλασης των επιφανειών του χώρου
- Γεωμετρικές παράμετροι
 - τύποι προβολής
 - θέση και κατεύθυνση της κάμερας στο χώρο
 - αλλοιώσεις προοπτικής από τη χρήση κάμερας

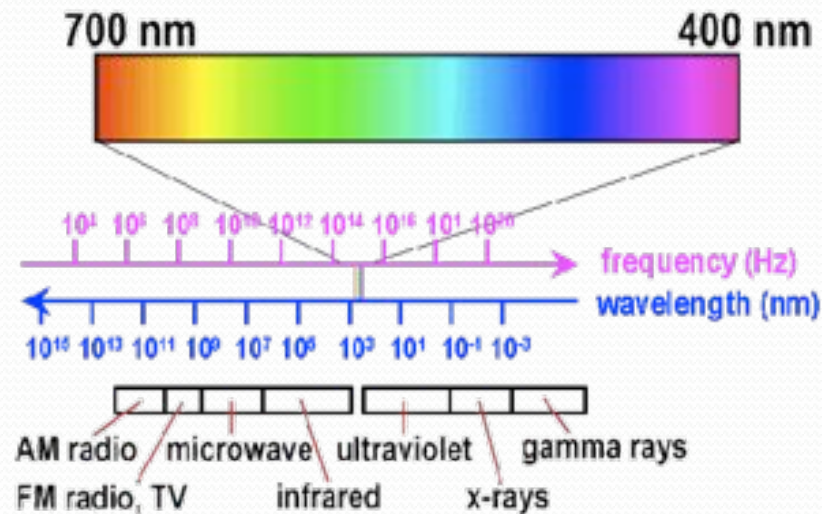
Οπτικό πεδίο

- d : διάμετρος φακού
- Η διάμετρος του φακού d και η εστιακή απόσταση f (focal length) ορίζουν το οπτικό πεδίο (**field of view w**)
- Ένα είδος ένδειξης για το τμήμα του 3-διάστατου χώρου που καλύπτει η κάμερα:

$$\tan w = d / 2f$$

Τι είναι το φως;

- Το ορατό τμήμα του φάσματος της ηλεκτρομαγνητικής ακτινοβολίας
- Το μήκος κύματός του κυμαίνεται ανάμεσα στα 400 και 700 νανόμετρα περίπου



Μικρά μήκη κύματος

- Διαφορετικού μήκους κύματος ακτινοβολίες έχουν διαφορετικές ιδιότητες
- Π.χ. το τμήμα του φάσματος που αντιστοιχεί σε x-rays περιέχει επαρκή ενέργεια ώστε να διαπερνά όγκους υλικών



Μεγάλα μήκη κύματος

- Μεγάλες ποσότητες υπέρυθρης ακτινοβολίας εκπέμπονται από θερμά αντικείμενα (η μέτρηση της τα κάνει «ορατά» ακόμη και στο σκοτάδι



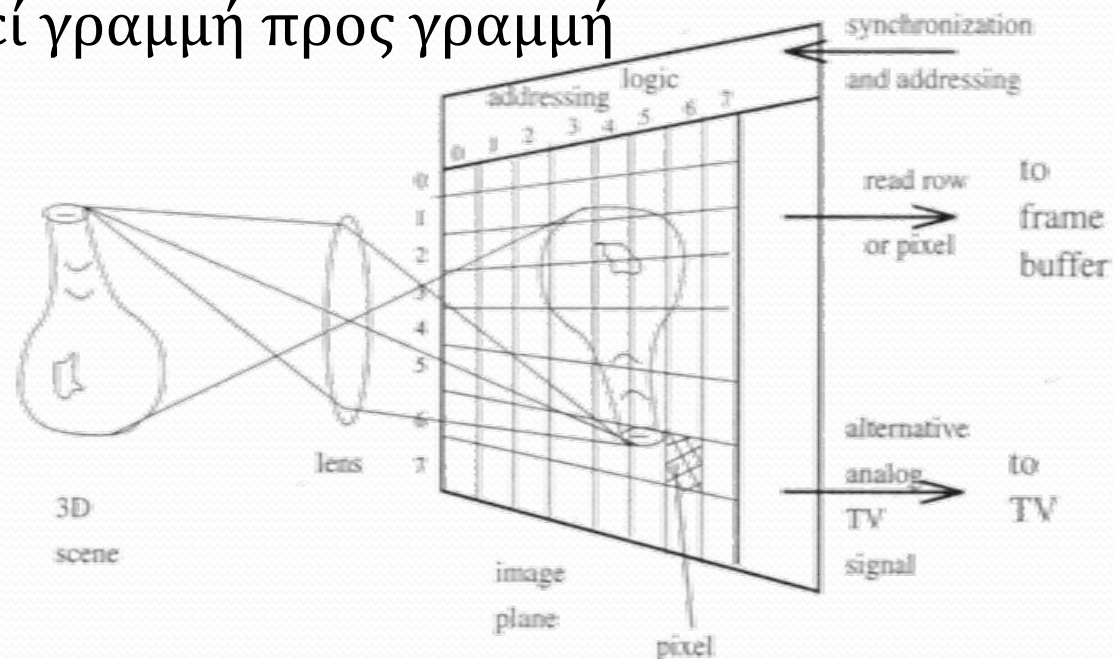
Εικόνες βάθους

- Ουσιαστικά πίνακες με τιμές βάθους για κάθε αντικείμενο στην εικόνα
- Με τη χρήση υπερήχων ή λέιζερ



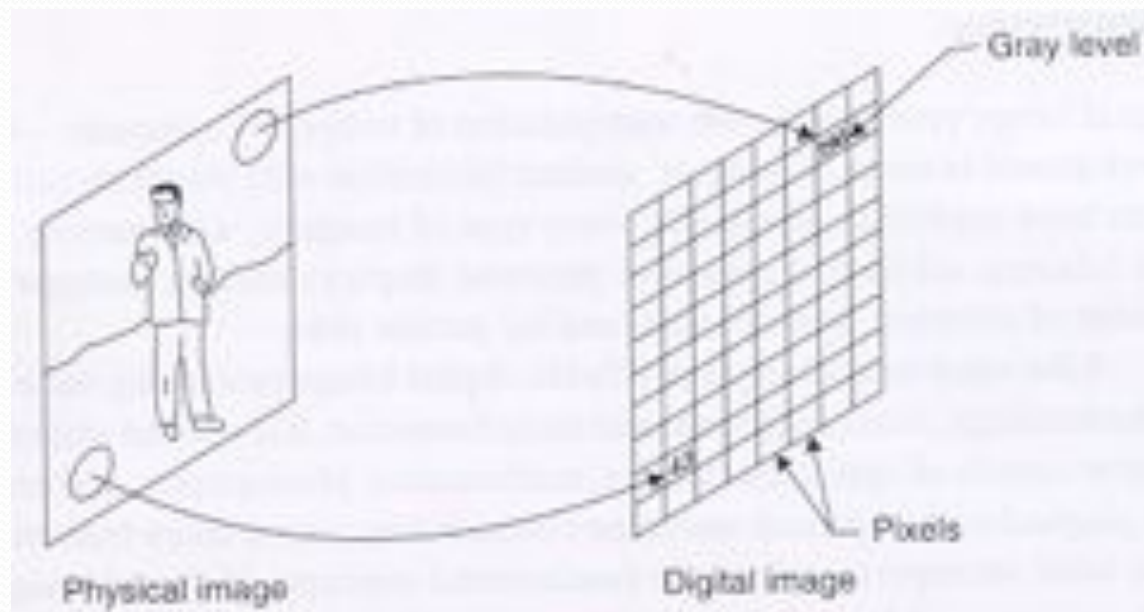
CCD (Charged-Coupled Device)

- Μικροσκοπικές κυψέλες στερεάς κατάστασης μετατρέπουν την ενέργεια του φωτός σε ηλεκτρική ενέργεια
- Η εικόνα είναι ένα είδος ψηφιακής μνήμης που μπορεί να διαβαστεί γραμμή προς γραμμή

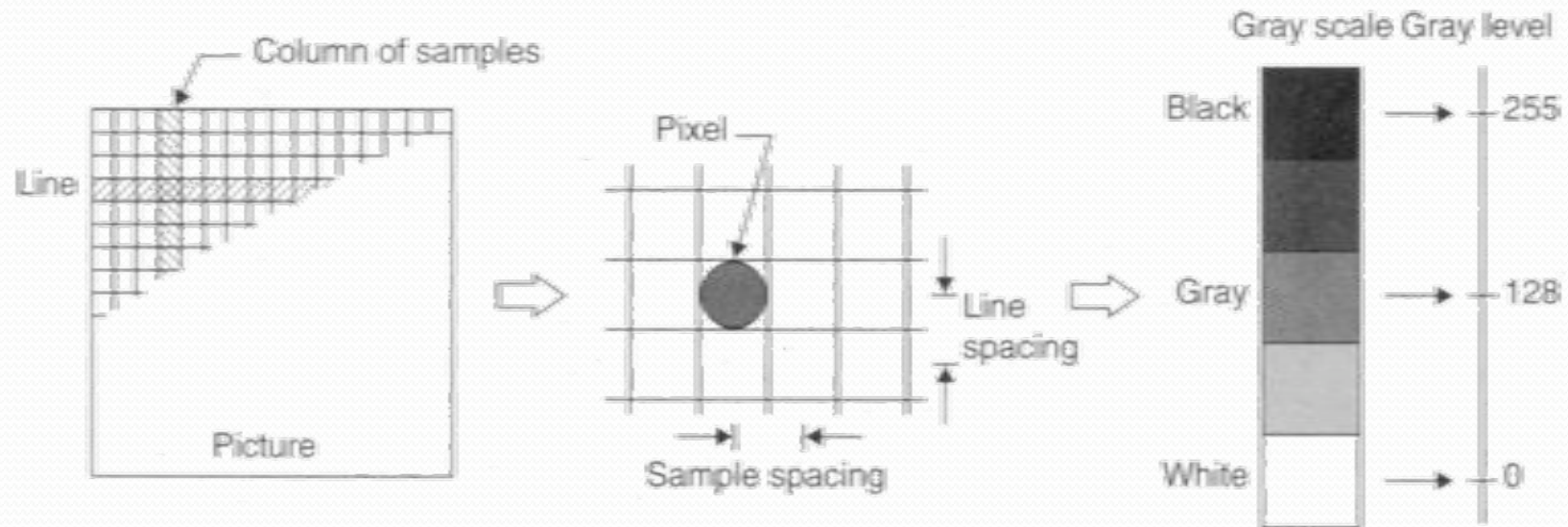


Ψηφιοποίηση της εικόνας

- **Δειγματοληψία (Sampling)** Μέτρηση ενός σήματος σε διακριτά σημεία
- **Κβαντισμός (Quantization)** αναπαράσταση της μέτρησης με έναν ακέραιο αριθμό



Ψηφιοποίηση της εικόνας



Κβαντισμός Εικόνας

- 256 gray levels (8bits/pixel) 32 gray levels (5 bits/pixel) 16 gray levels (4 bits/pixel)



- 8 gray levels (3 bits/pixel) 4 gray levels (2 bits/pixel) 2 gray levels (1 bit/pixel)



Δειγματοληψία

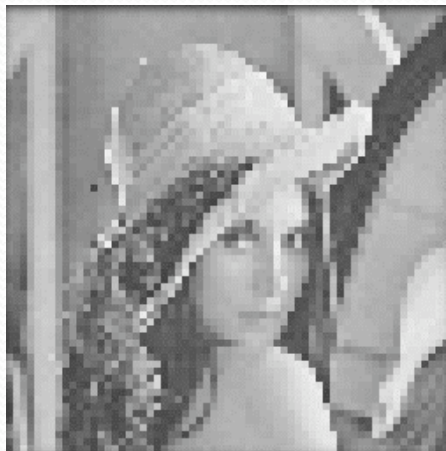
original image



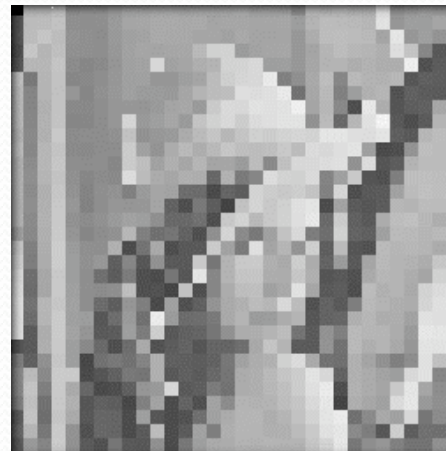
sampled by a factor of 2



sampled by a factor of 4



sampled by a factor of 8



Και λίγη ιστορία...



- 'lena' ή 'lenna' = digitized Playboy centerfold, Νοεμβρίου 1972 (Lenna τυπώθηκε στο Playboy, Lena είναι η σουηδική έκδοση. Το πλήρες όνομα: Lena Soderberg)
- Ο Alexander Sawchuk (επίκουρος καθηγητής, ηλεκτρολόγος μηχανικός, στο USC Signal and Image Processing Institute (SIPI)) ψηφιοποίησε την εικόνα το καλοκαίρι του 1973
- Η Lenna τιμήθηκε στο 50^ο συνέδριο IS&T στη Βοστώνη το Μάιο του 1997

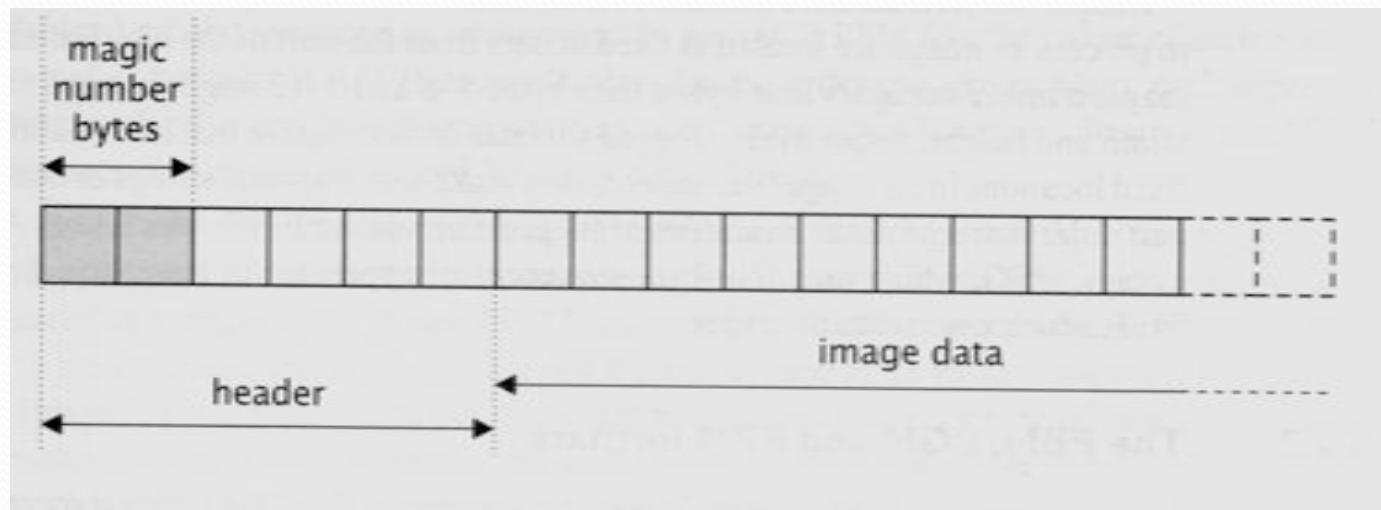
Η ψηφιακή εικόνα

- Μια ψηφιοποιημένη εικόνα καταλήγει να είναι ένας πίνακας με αριθμούς
- Κάθε αριθμός αντιπροσωπεύει τη φωτεινότητα στο συγκεκριμένο σημείο (ενδεχόμενα σε ένα κανάλι)
- N: # γραμμών, M: # στηλών, Q: # διαφορετικών επιπέδων φωτεινότητας
 - $N = 2^n$, $M = 2^m$, $Q = 2^q$ (q ο αριθμός bits/pixel, εναλλακτικά και ως K)
 - Αποθηκευτικές απαιτήσεις: $N \times M \times Q$ (π.χ., $N=M=1024$, $q=8$, 1MB)

$$\begin{array}{cccc} f(0,0) & f(0,1) & \dots & f(0,M-1) \\ f(1,0) & f(1,1) & \dots & f(1,M-1) \\ \dots & \dots & \dots & \dots \\ f(N-1,0) & f(N-1,1) & \dots & f(N-1,M-1) \end{array}$$

Μορφοποίηση αρχείων εικόνας (formats)

- Τα περισσότερα ακολουθούν την παρακάτω διάταξη
- Στην κεφαλίδα υπάρχει τουλάχιστον η πληροφορία των διαστάσεων και του μεγέθους κάθε pixel
- Οι περισσότερες κεφαλίδες ξεκινούν με ένα “magic number” – μία σύντομη ακολουθία – χαρακτηριστική για τη μορφοποίηση





Common image file formats

- GIF (Graphic Interchange Format) -
- PNG (Portable Network Graphics)
- JPEG (Joint Photographic Experts Group)
- TIFF (Tagged Image File Format)
- PGM (Portable Gray Map)
- FITS (Flexible Image Transport System)

Το μέγεθος της Λένα...

- Title: lena_std.tif
- Width: 512 pixels
- Height: 512 pixels
- Bits per pixel: 32 (RGB)
- TIFF: 772 KB
- JPEG: 40 KB
- GIF: 228 KB (8-bit color)
- RAW: 768 KB
- PNG: 720 KB



Επεξεργασία εικόνων

Βασικές αρχές

Γιατί να επεξεργαστεί κανείς εικόνες;

- Συμπύεση δεδομένων
- Βελτιώσεις και διορθώσεις
- Ακριβείς μετρήσεις (αποστάσεις, μέτρηση όγκου σε ακολουθίες)
- Αναγνώριση περιεχομένου
 - Από μία σειρά ψηφίων χωρίς καμία σημειολογία
 - Η χειρωνακτική επεξεργασία εικόνων έχει πολλά αρνητικά
 - Περιορισμένη ακρίβεια
 - Εξαιρετικά χρονοβόρα
 - Μετατροπή σε διαφορετικές μορφές, με στόχο τη στατιστική ή άλλη μαθηματική επεξεργασία, όπως και την καλύτερη οπτικοποίηση της πληροφορίας (π.χ. τρισδιάστατη αναπαράσταση)

Είδη επεξεργασίας

ΣΤΟ ΧΩΡΟ

- Σημειακή επεξεργασία εικόνων
 - Η επεξεργασία αφορά μεμονωμένα pixel, χωρίς να λαμβάνει υπόψη της ομάδες ή άλλα χαρακτηριστικά
- Αριθμητική επεξεργασία
 - Ειδική κατηγορία σημειακών
 - Π.χ. Αλγεβρικές πράξεις ανάμεσα σε εικόνες
- Γεωμετρική επεξεργασία
 - Συμπληρωματικές στις σημειακές
 - Δεν σχετίζονται με τη φωτεινότητα, αλλά με τη θέση
 - Π.χ. περιστροφή

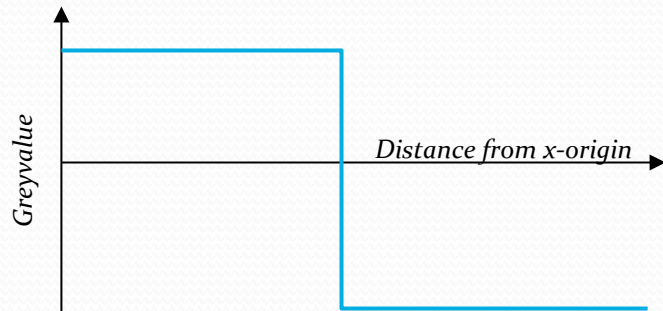
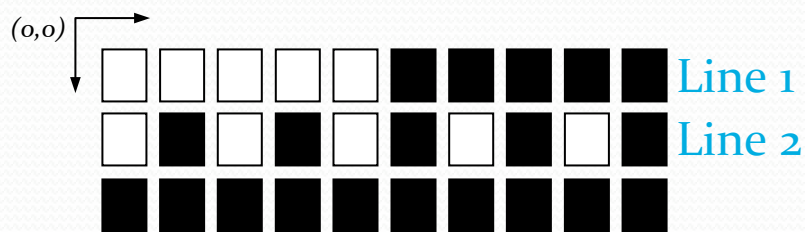
Είδη επεξεργασίας (συνέχεια)

- Επεξεργασία περιοχών
 - Γνωστή και ως επεξεργασία με μάσκες
- **ΣΤΟ ΧΩΡΟ ΣΥΧΝΟΤΗΤΩΝ**

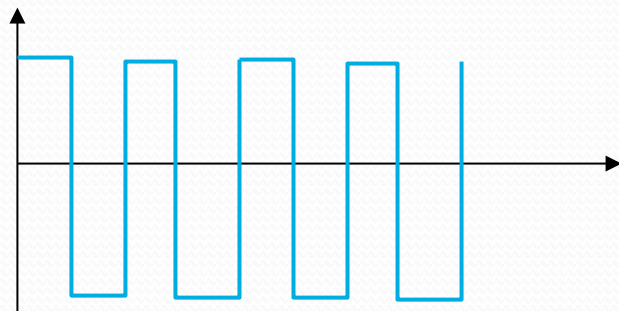
Πριν την επεξεργασία

- Εργαλεία εκτίμησης της κατάστασης
 - Προφίλ
 - Ιστόγραμμα

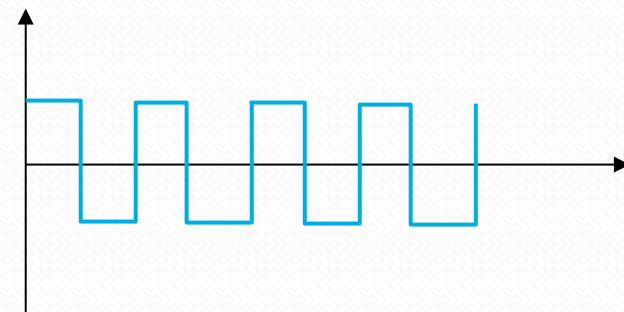
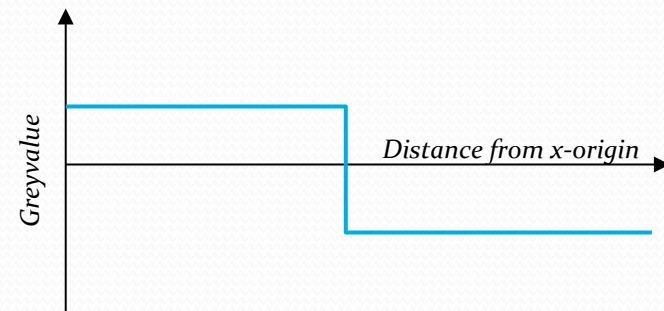
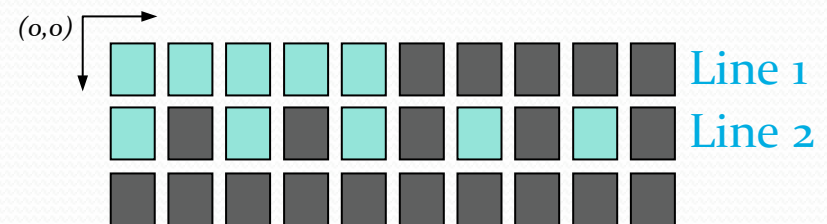
Προφίλ γραμμών



Line 1

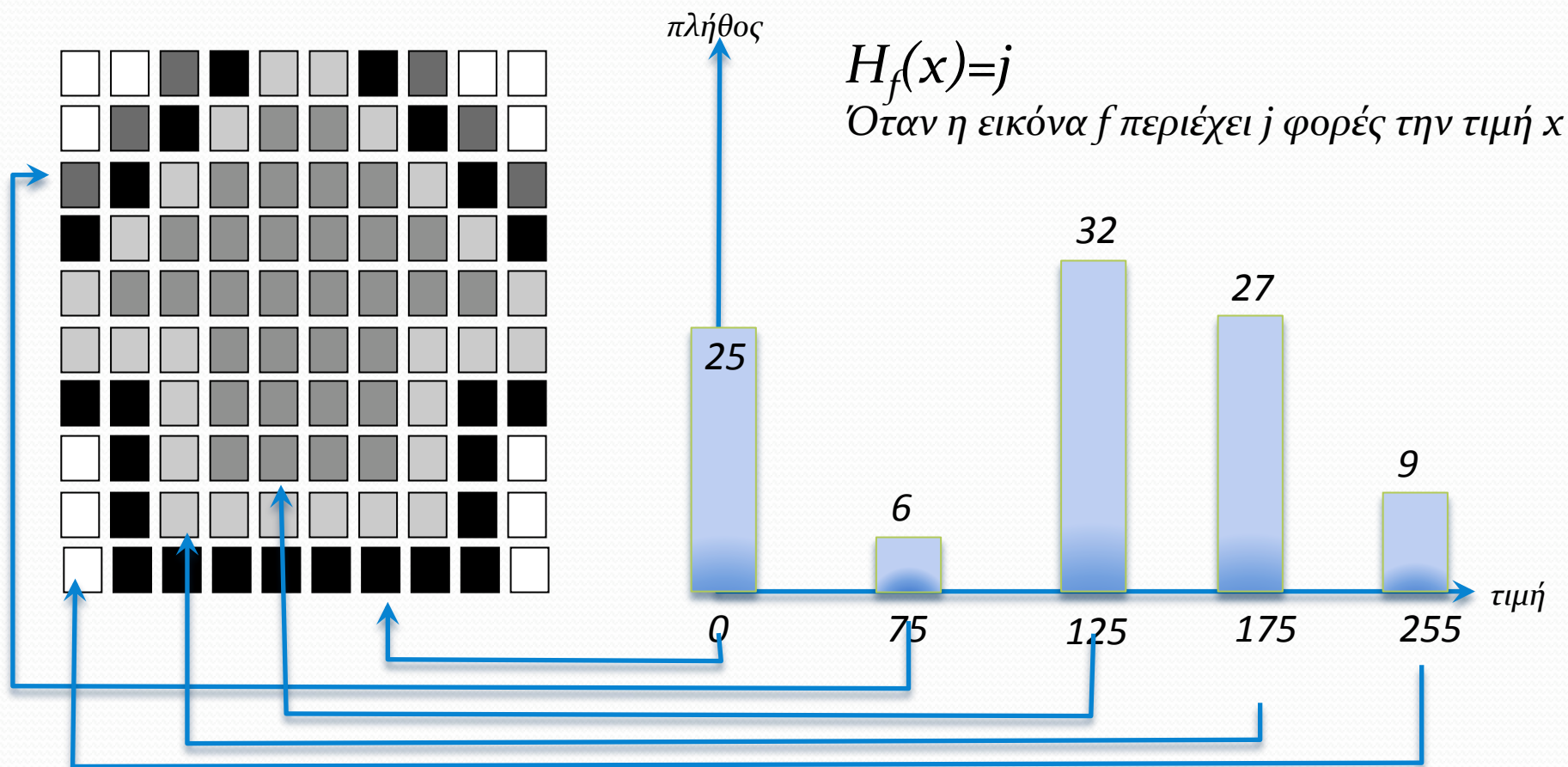


Line 2



Το ιστόγραμμα

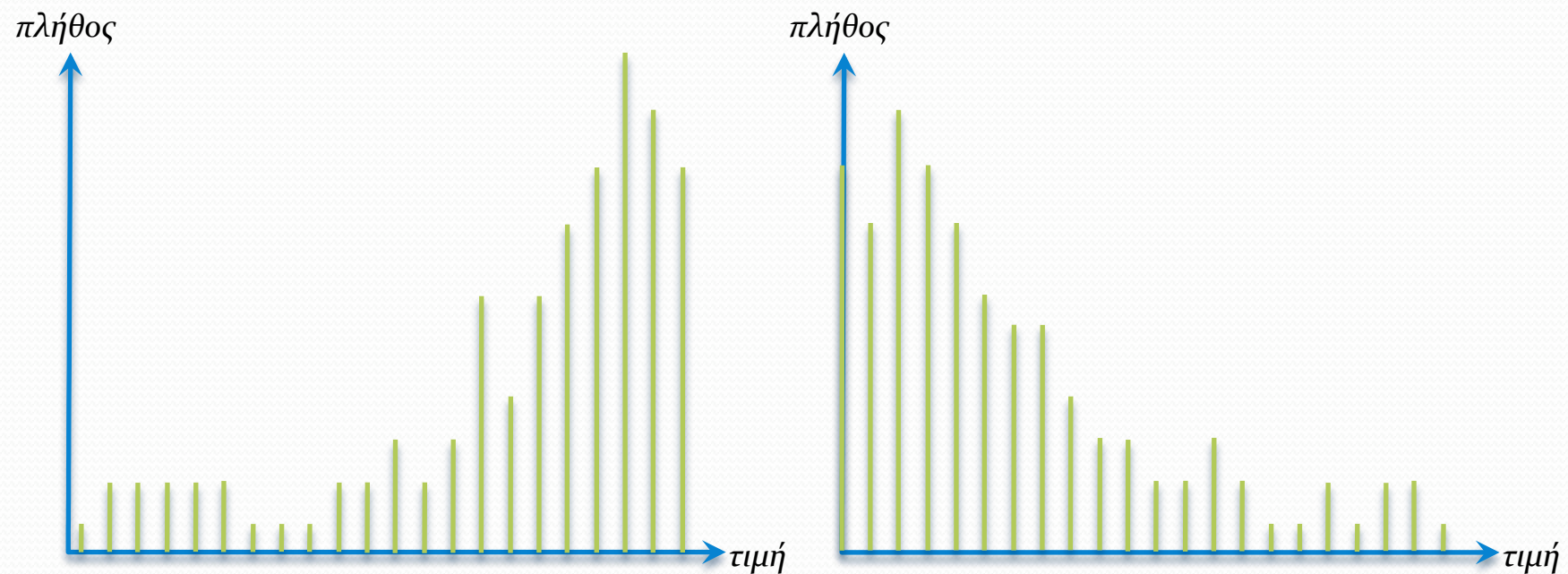
- Μία στατιστική έκφραση της εικόνας



Υπολογισμός του ιστογράμματος

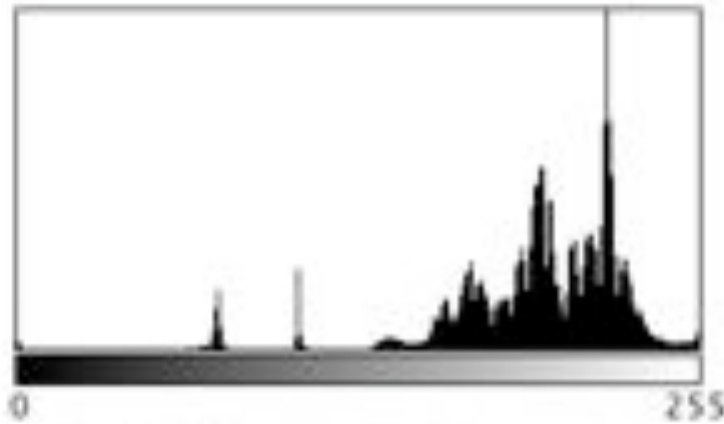
- Άνοιξε την εικόνα (φόρτωσε στη μνήμη)
- Μετάτρεψέ τη σε μονοχρωματική
- Διάβασε το μέγεθος και την κωδικοποίησή της από την κεφαλίδα
- Δημιούργησε ένα πίνακα με τόσα στοιχεία όσα επιτρέπει η κωδικοποίηση
- Για κάθε σημείο στην εικόνα
 - Άντλησε την τιμή του σημείου
 - Πρόσθεσε ένα στην αντίστοιχη θέση στον πίνακα

Τι μας λένε τα ιστογράμματα;

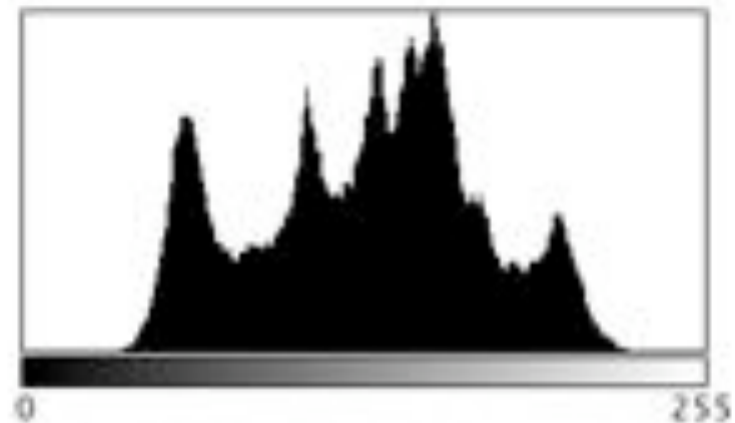


Πόσο είναι το άθροισμα του πλήθους του ιστογράμματος;

Ποια εικόνα προτιμάτε να επεξεργαστείτε;



Count: 92000
Mean: 194.102
StdDev: 35.508
Min: 0
Max: 255
Mode: 221 (4211)



Count: 262144
Mean: 128.229
StdDev: 42.925
Min: 26
Max: 237
Mode: 154 (3211)

Οι απαντήσεις



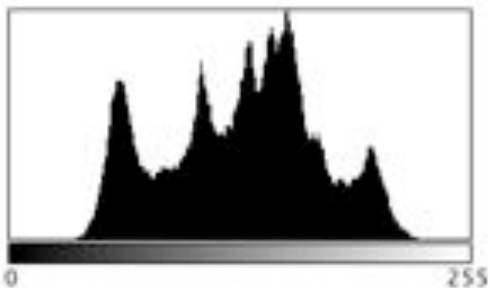
Η θεωρία της σημειακής επεξεργασίας

- Εφαρμογή μίας συνάρτησης σε κάθε ένα σημείο της εικόνας χωρίς διακρίσεις
- Μία «αφαιρετική» γενική προσέγγιση:
 - $g(\mathbf{n}) = h[f(\mathbf{n})]$
 - \mathbf{n} : τονισμένα γράμματα συχνά συμβολίζουν διανύσματα, εδώ τις συντεταγμένες ενός σημείου της εικόνας
- Γραμμική επεξεργασία:
 - $g(\mathbf{n}) = Pf(\mathbf{n}) + L$

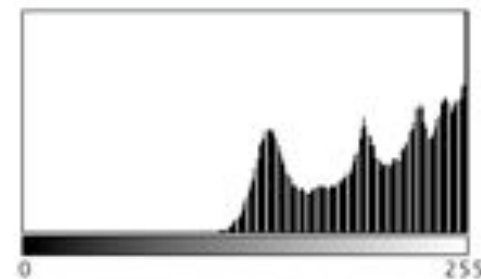
Πρόσθεση σταθεράς



+ 100 =



Count: 262144
Mean: 128.229
StdDev: 42.925
Min: 26
Max: 237
Mode: 154 (3211)

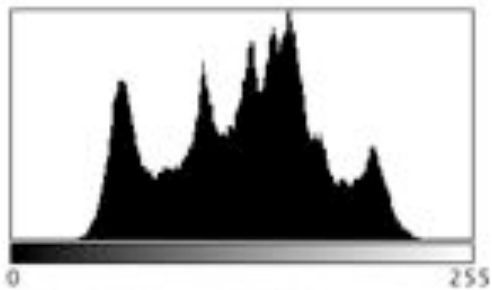


Count: 262144
Mean: 214.904
StdDev: 41.676
Min: 100
Max: 255
Mode: 255 (74016)

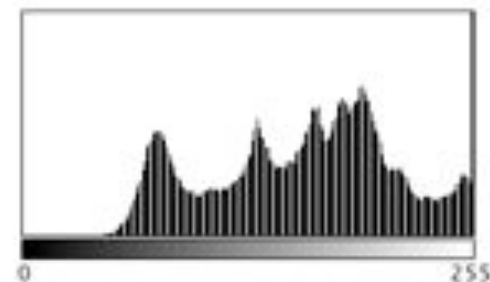
Πολλαπλασιασμός



$\times 1.25 =$



Count: 262144	Min: 26
Mean: 128.229	Max: 237
StdDev: 42.925	Mode: 154 (3211)



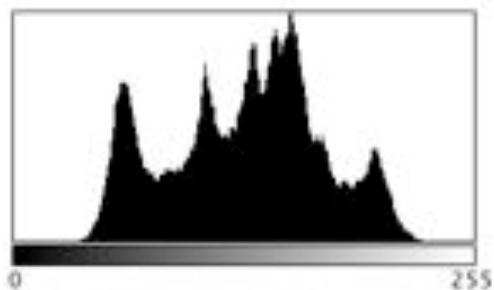
Count: 262144	Min: 33
Mean: 160.173	Max: 255
StdDev: 53.204	Mode: 255 (8699)

Το αρνητικό ως γραμμική επεξεργασία

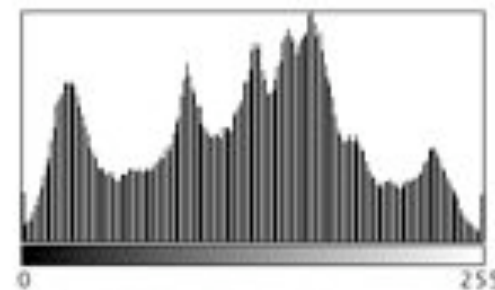
- Κατά βάση αφαιρούμε κάθε τιμή από τη μέγιστη δυνατή τιμή
- Με τον πολλαπλασιασμό με -1 δημιουργούμε ένα είδωλο του ιστογράμματος
- Προσοχή σε εικόνες με τιμές χωρίς πρόσημο!!!
- $g(\mathbf{n}) = -f(\mathbf{n}) + (K-1)$



Το κανονικοποιημένο ιστόγραμμα



Count: 262144 Min: 26
Mean: 128.229 Max: 237
StdDev: 42.925 Mode: 154 (3211)



Count: 262144 Min: 0
Mean: 122.984 Max: 255
StdDev: 62.151 Mode: 160 (3211)

Άπλωμα ιστογράμματος

- Οι ακραίες τιμές του ιστογράμματος μεταφέρονται στην ελάχιστη και τη μέγιστη δυνατή τιμή
- Οι υπόλοιπες «απλώνουν» δημιουργώντας μεγαλύτερη ευκρίνια στην εικόνα

- $A = \min \{f(\mathbf{n})\}$
- $B = \max \{f(\mathbf{n})\}$
- $PA + L = 0$
- $PB + L = K - 1$

$$P = \left(\frac{K - 1}{B - A} \right)$$

$$L = -A \left(\frac{K - 1}{B - A} \right)$$

$$g(n) = FSHS[f(n)] = \left(\frac{K - 1}{B - A} \right) [f(n) - A]$$

Κατηγορία 2: αριθμητική επεξεργασία

- Πράξεις μεταξύ εικόνων
- Τυπική εφαρμογή η μέση τιμή διαφορετικών λήψεων της ίδιας εικόνας για περιορισμό του θορύβου
 - Από τον αισθητήρα
 - Από τα ηλεκτρονικά
 - Από τη μετάδοση
 - Κ.ο.κ.



Σε μεγέθυνση



Το αποτέλεσμα



Γεωμετρική επεξεργασία

- Μετασχηματισμός συντεταγμένων και όχι φωτεινότητας
 - $g(\mathbf{n}) = f(\alpha(\mathbf{n}))$
- Περιστροφή της εικόνας
- Μεγέθυνση, σμίκρυνση (ζουμ)
- Είναι προφανές ότι δεν επαρκεί ο γεωμετρικός μετασχηματισμός
 - Μερικά σημεία πέφτουν «ανάμεσα» σε pixel, άλλα εκτός εικόνας
 - Τεχνικές: Nearest neighbor, bilinear

