



עיבוד ספרי של תמונות

Digital Image Processing

נמרוד פלאג

המעבדה לעיבוד אותות ותמונות
הפקולטה להנדסת חשמל, הטכניון

עדכון : אפריל 2010

פרק 1 : מבוא לעיבוד תמונה

- מהי תמונה?

האור המוחזר מ„אובייקט“, לאחר מעבר במערכת אופטית, ונקלט במערך חיישנים (סרט צילום , CCD ...).

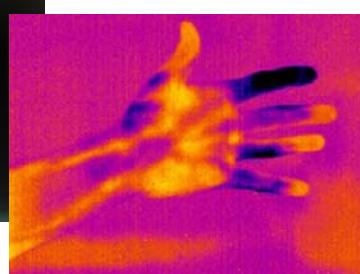
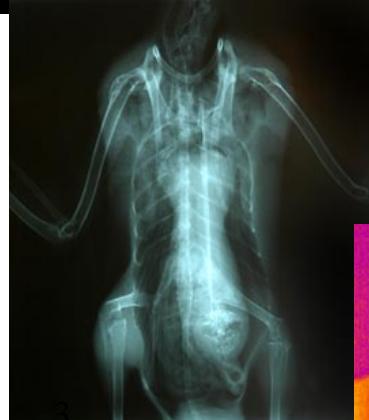
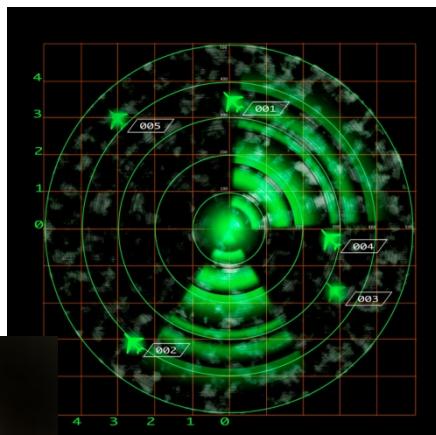
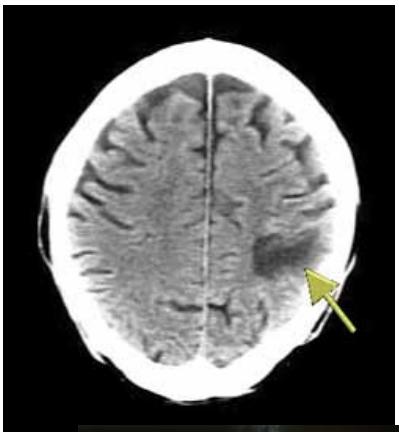
- התמונה היא דו-ממדית, ובה עצמת האור (בהיקות) וצבעו (אורך גל) תלויים במקומות:

פילוג העצמה והצבע של הקריינה

“תמונות” אחרות

- ישנים סוגים רבים של תמונות שאינן נקלות

דוקא במערכת אופטית:



- ראdar

רנטגן

CT

מערך מיקרופוניים

- אור בלתי נראה (IR) ועוד....

אבל מה שאנו אוהבים לראות זה...

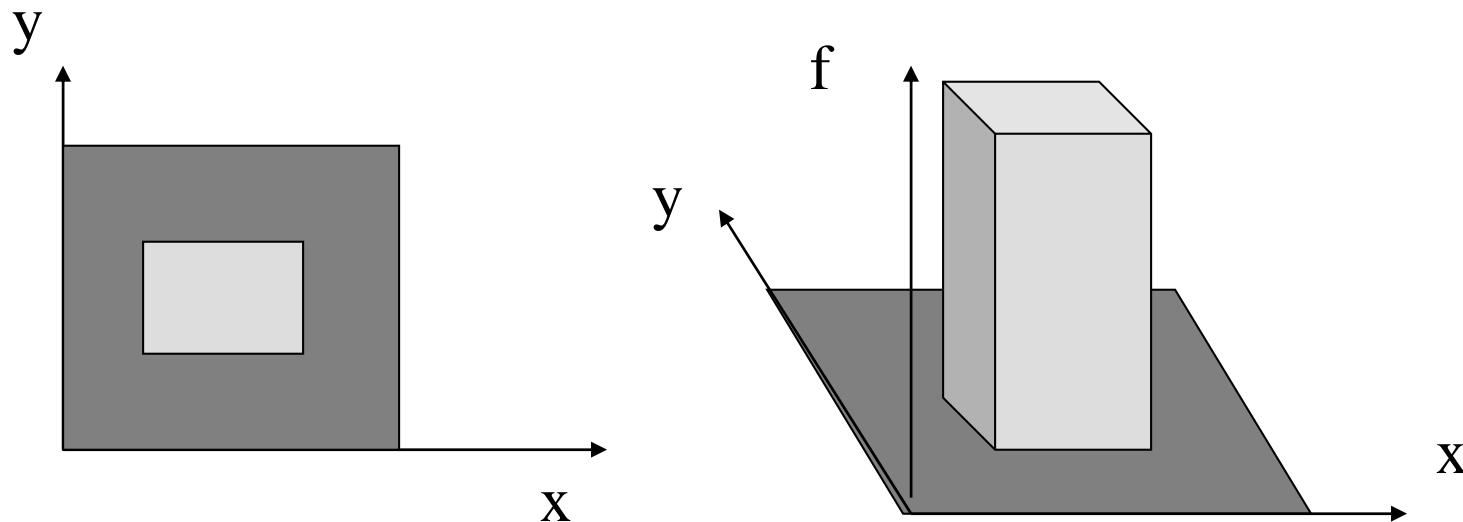


צנזורה!



ייצוג מתמטי של תמונה

- נתאר תמונה כפונקציה של שני משתנים : (y) $f(x,y)$
- x,y הם **הקוֹרְדִּינָטוֹת** במרחב דו-ממדי, והערך של f מבטא (בתמונה אופטית) את **הבהיקות** (עוצמת אור) בנקודה.



ומה בתמונה צבע?

- בתמונה צבע מורכבת כל נקודה (pixel) משלשה צבעי יסוד : **אדום**, **ירוק** ו**כחול** (מסומן RGB) ולכון נתאר כל פיקסל ע"י שלש פונקציות:

$$r(x, y), g(x, y), b(x, y)$$

בזר” נ עוסק בתמונות רמות-אפור, לשם הנקוטות.

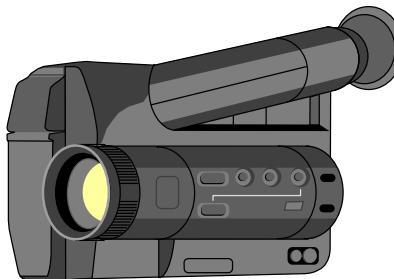
רמות אפור ורציפות

- במקור, האובייקט אותו מצלמים הינו **“רציף”** והוא בשטחו (**“מרחב”**) והו בצבאו (**“בהיקות”**), אך בעבר למחשב חיבים, משיקולים מעשיים **לדגום** את התמונה בנקודות מסוימות (פיקסלים בדים) ובצבאים מסוימים (רמות אפור בדים)
- מתקבל מערך (או מטריצה) של **נקודות מייצגות**

הערה: בדים = דיסקרטי , Discrete

אמצעי קלט לתמונות

- אמצעי הרכשה (Grabbing) לתמונה:
 - מצלמה ספרטית (בדר"כ מבוססת: CCD)
- מדידת מתח/זרם יחסית לכמות פוטוניים, עברור כל פיקסל.
- רזולוציה בסיסית : 640 על 480 - (VGA) רוחוק מסרט צלום.
- אלטרנטיבה : שורת חיישנים עם סריקה מכנית (איטי)
- לבע : שלשה מערכיים + מסננים אופטיים מתאימים.
- קיימים חיישנים מתאימים גם ל IR - רנטגן ועוד .



אמצעי קלט לתמונות (המשך)

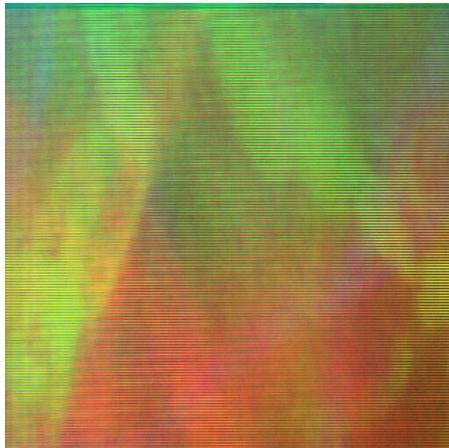
– סורק אופטי:

- עקרון דומה עם הקלות: כיוול נוח, רקע אחיד, תאורה מבוקרת.
- רזולוציית הסריקה נמדדת בדר"כ ב , dpi – ולא עולה על 600 נקודות לאינץ' (כ- 24 נקודות למ"מ)
- רזולוציות גבהות יותר מושגות ע"י אינטרפולציה !



– תמונות סינטטיות:

- תמונות מחושבות בשיטות הדמייה : מכ"מ... , US, MRI, CT
- מדידת תופעה פיזיקלית ותרגום שלה
- לערך דו-ממדי.



אמצעי פלט לתמונות



- **המרת תמונה ספרתית למסך:** (Display)
 - צגי שפופרת קרן-קתרודית (SKU: CRT)
- קרן סורקת המאירת נקודות של חומר פלאורסנטי. (RGB)
- פרמטרים חשובים: מהירות הסריקה, רזולוציית המסך.
- אפשרויות אחרות, לדוגמה: LCD
 - מדפסות:
- **Continuous Tone:** יקרות ודוממות לפיתוח סרט צילום אמיתי.
- **Half-Tone** בדר"כ שימוש במגוון צבעים מוגבל וטכניקת- ליצירת אשליה של רמות אפור/צבע. (Laser, Ink, Wax).

תמונה עם אפקט lena Half-Tone



עיבוד תמונה ותחומיים סמוכים

- **“עיבוד תМОנות”** דן ב”**משימות**“ בהן הקלט והפלט הינם תמונות (ספרתיות או אנלוגיות.).
- **ראיה ממוחשבת** : הקלט תמונה, והפלט אינו תמונה אלא מידע עליה. (Computer Vision).
- **גרפיקה ממוחשבת** : הקלט הינו מידע כלשהו והפלט הינו תמונה. (Computer Graphics).
- **לדוגמא** : מציאת המיפוי התלת-מימדי של גופים בתמונה, לעומת בניית תמונה ריאלית ממידע על גופים תלת-ממדיים במרחב. (Rendering).

מסלול אופיני לתמונה



תחומיים שונים בעיבוד תמונה

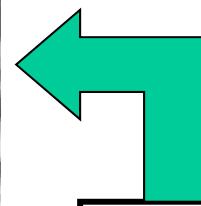
- שיפור: סילוק "קלוקלים" שנוצרו בדגימה.
- שחזור: סילוק קלוקלים שטיבם ידוע.
- ניתוח: זיהוי מרכיבים בדרכ'כ לצורך מע' אוטומטיות.
- דחיסה: ייצוג חסכוני ע"י סילוק מידע לא חיוני.
- בנייה: הרכבה על-סמך מידע חלקית (הדמייה ע"י "חתכים")

שיפור תמונה

קלוקולים אופיניים המתקבלים במהלך דגימת או



צילום התמונה, מפריעים להבחן בפרטים, מטשטשים את האובייקט וכדומה.



תמונה מקור : תצ"א
(Aerial)

ניגודיות נמוכה (Low Contrast)



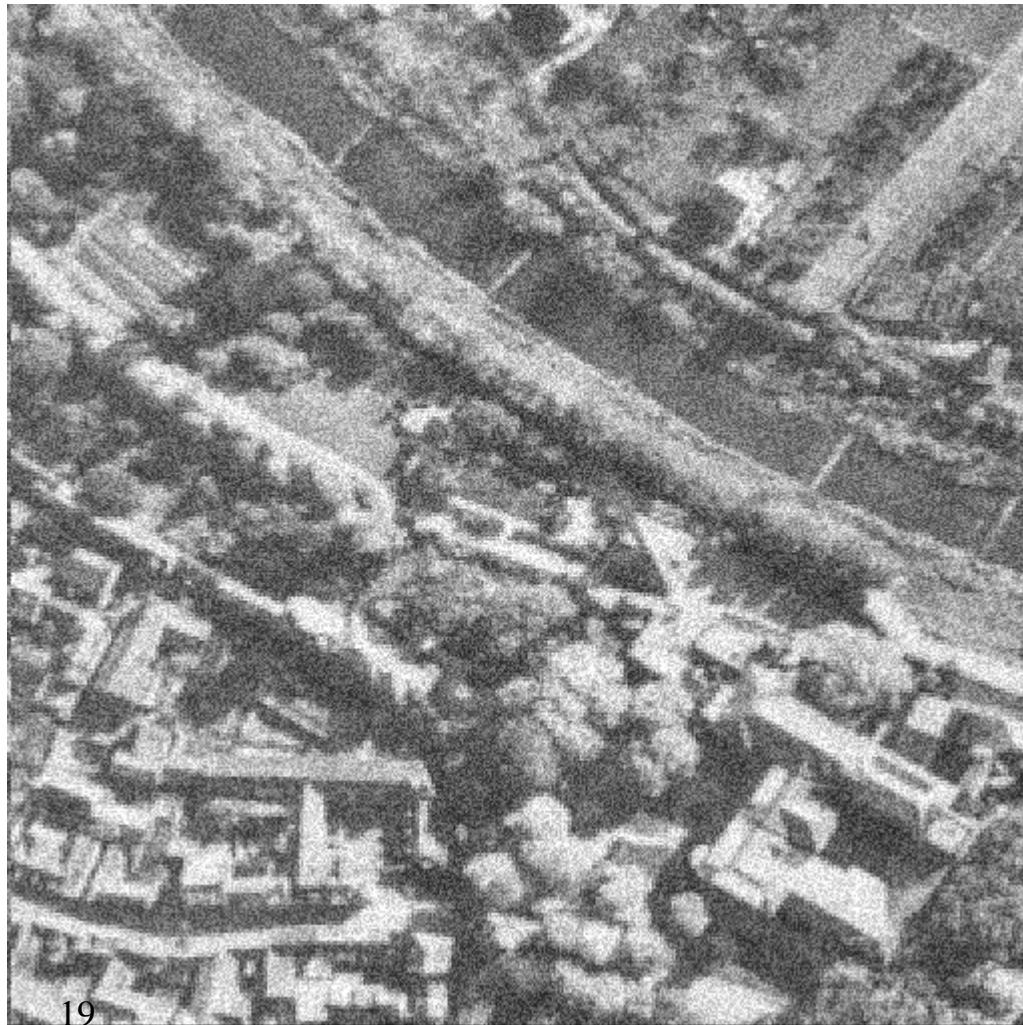
טשטוש(Blur)



מריחה בתנועה (Motion Blur)



רעש אחיד (Uniform Noise)



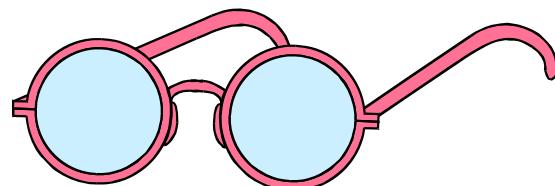
קלוקולים נוספים

- מיקוד לא נכון (De-Focus).
- תזוזה או רעדת המצלמה או הגוף.
- תנודות תרמיות של אוויר (טורבולנציה)
- ניתן לתקן חלק מהבעיות ע"י עיבוד מתאים אך חייבים לזכור:

לא ניתן להוסיף מידע שלא היה קיים בתמונה המקורית !

שיטת לשיפור התמונה

- השיטות רבות ומגוונות ותלויות באופי הקלculus
ובמטרת השיפור!
- מדידת השיפור, פעמים רבות סובייקטיבית (לפי העין), ולעתים ניתנת להערכת מתמטית.



ניתוח תמונה

- הכוונה לשיטות המאפשרות שליפת מידע מתוך תמונה באמצעות מחשבים, ללא צורך המתבונן אנושי.
- דוגמאות:
 - מיון עצמים
 - איתור תקלות במעגלים מודפסים
 - זיהוי כתוב
 - ניהוג רכב אוטומטי



דחיסת תמונה



33,252B 8:1
דחוס



מקור 262,144B = 512 x 512

טכניקות של עיבוד תמונה

- עיבוד אופטי
- עיבוד אלקטרוני - אנלוגי
- עיבוד אלקטרוני - ספרתי

עיבוד אופטי

- אין המרה של **התמונה הרציפה** לנקודות בדיםות
- מספר פועלות ניתנות לביצוע:
 - התמרת פורייה
 - סינון מרחבי
 - חישובי קורלציה, קונבולוציה
- יתרונות: **מהירות** (האור), **מקבילות** (כל הנקודות מחושבות במקביל), **דיוק** ועוד.

עיבוד אופטי (המשך)

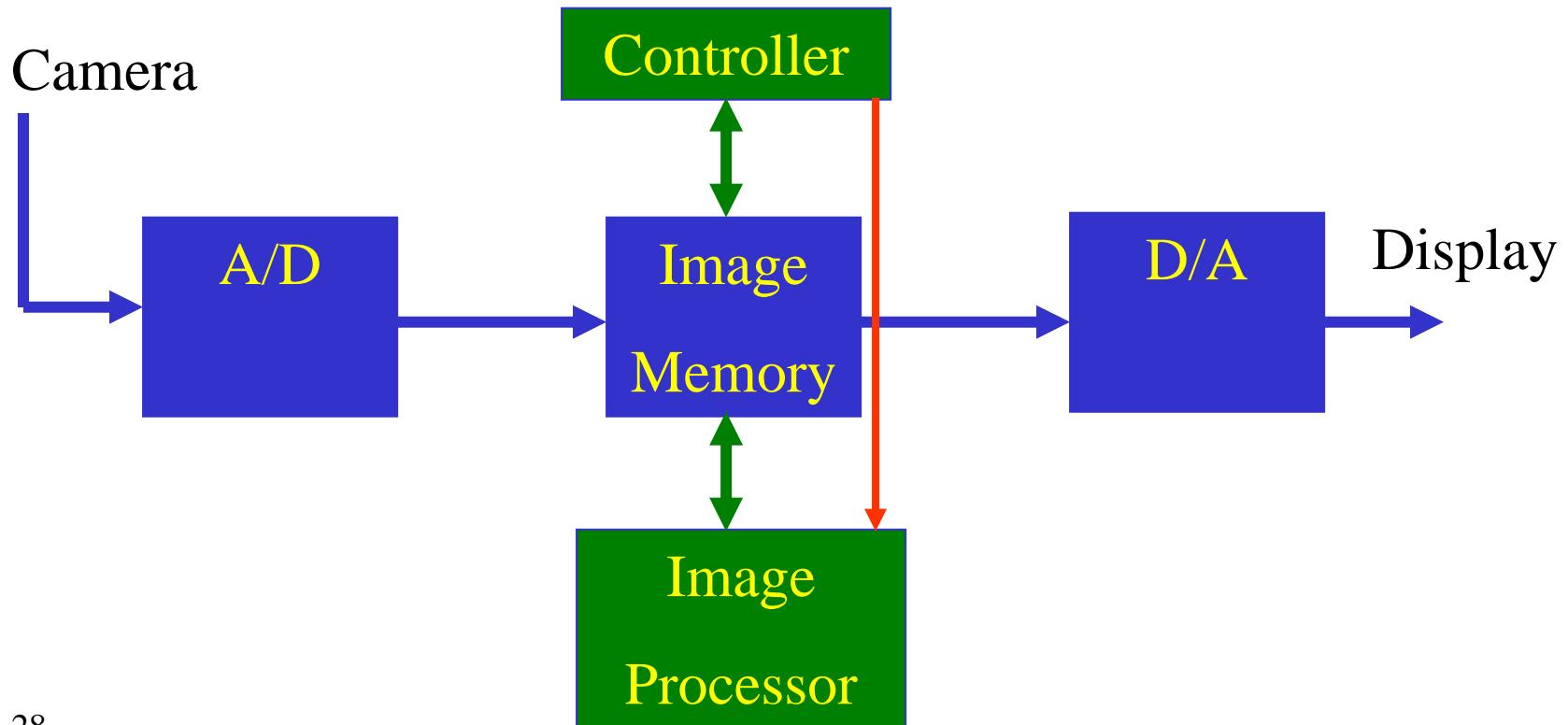
- **מגוון פעולות מוגבל** (חסרות אפילו 4 פעולות היסוד)
- חוסר גמישות מחייב מערכ אופטי ייחודי לכל פעולה.
- דריש אור קוהרנטי (לייזר)
- העברת ושמירת המידע הינם בעויה תיימן מאד.

עבודה אלקטרוני - אנלוגי

- המרה **ל מידע חשמלי**, ביצוג ע"י אות אנלוגי.
- ישנו מספר **שימושים אפשריים**: שיפור קונטרסט, הורדת רעשים ועוד.
- **הឧבודים בזמן-אמת**, שומר על דיקוק גבוה וזול.
- גם כאן יש **בעיות מעשיות** של מערכות חמרה:
חוסר גמישות, רגישות להתיישנות רכיבים,
רגישות לשינויי טמפרטורה, בעיית איסון ועוד.

עיבוד אלקטרוני - ספרתי

- תהליך שמחייב המרת התמונה ע"י דגימה והמרה בחזרה לאות רציף ע"י שייחזורה:



עבוד אלקטרוני - ספרתי (המשך)

- יתרונות:
 - כל פעולה ארכיטקטונית אפשרית
 - אמצעי איסון והעברה פשוטים וזולים
 - גמישות רבה
 - יציבות מערכת
 - מחיר זול
- עדין קיימת בעיות מהירות עבודה אך הולכת וירדת עם התפתחות המחשב (שעון, מיקבול וכו')

אפיון של תמונות

נכשא לאפיון תמונות לפי **מדדים אובייקטיביים**:

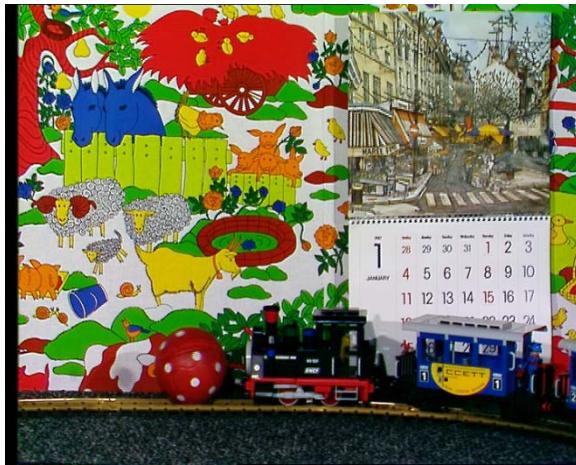
1. צבעוניות
2. בהיקות (Luminance) ובהירות (Brightness)
3. ניגודיות (Contrast)
4. כושר הפרדה (Resolution)
5. תדרים מרחביים
6. רעש (SNR ,MSE)

איפיון: צבעוניות

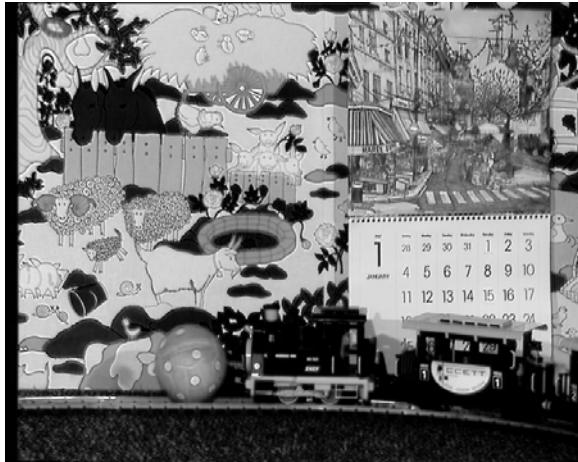
- הצלב מוסיף **מידע וצבעיות**, נעים יותר לעין.
- בדרך"כ ייצוג הצלב **במרחב תלת-מימדי**: בטלוויזיה ע"י אדום, כחול וירוק, (RGB) ובדף ע"י ארגמן, תכלול וצהוב, (CMY) + **שחור**, מסיבות של aicota ומחריר.
- המשמעות היא הכפלת פי שלשה של כמות המידע. בתמונה צבע לעומת תמונה רמות-אפור.

RGB Color Space

Original (color) image



R



G



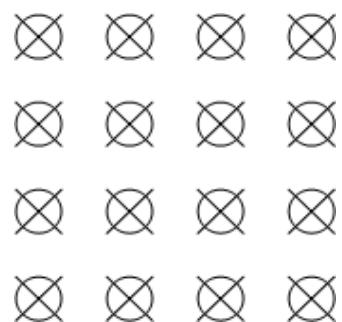
B



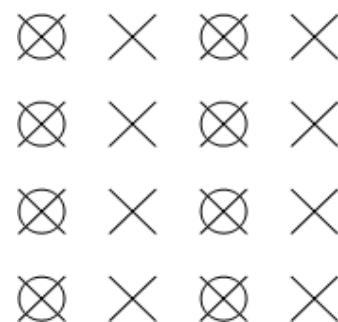
YCrCb Color Space

- Linear transformation of RGB color components
- Y = luminance component (brightness)
Cr,Cb = chrominance components (color)

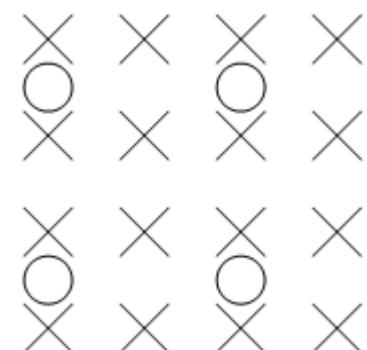
YCrCb 4:4:4



YCrCb 4:2:2



YCrCb 4:2:0

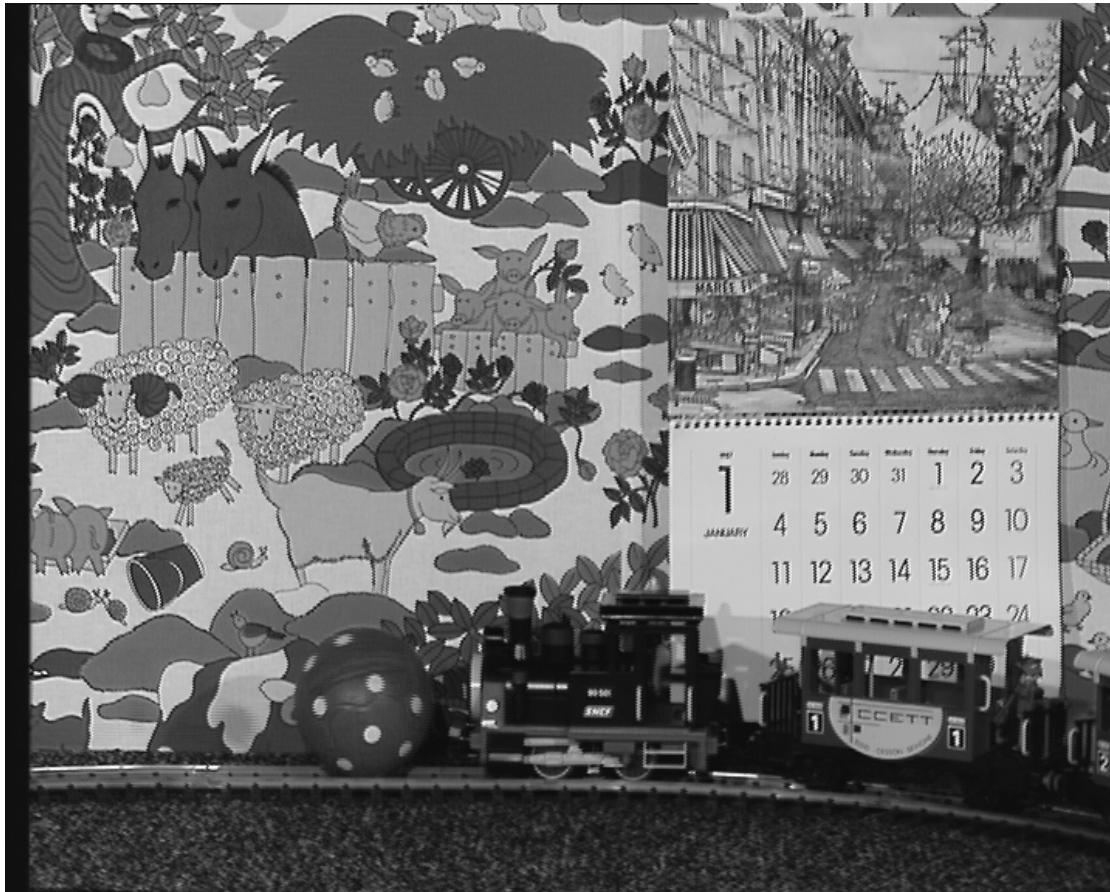


x – luma samples

o – chroma samples

4:2:0 Format - Example

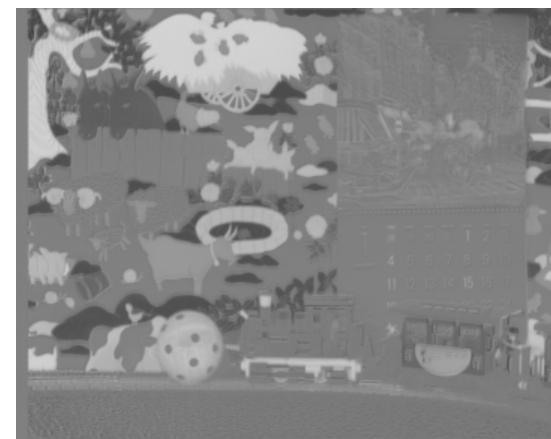
Luminance - Y



Chrominance



Cb



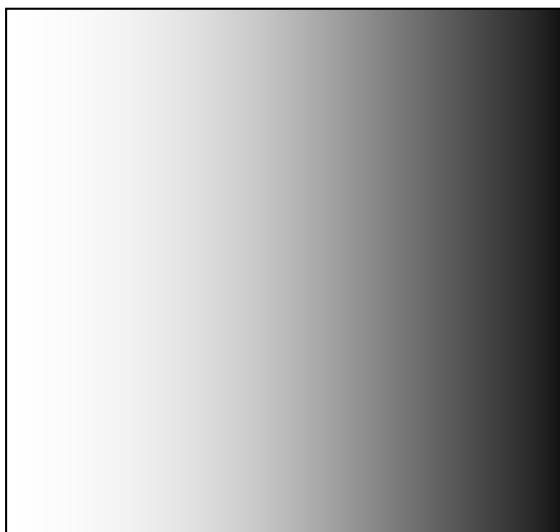
Cr

איפיון II : בהיקות ובהירות

- המידע בתמונה טמון בפילוג הבהירות בה. הבהירות הינה סובייקטיבית ומຕוארת ע"י הגודל האובייקטיבי **בהיקות** : עוצמת האור הנפלטת מיחידת שטח, לכוון הסתכלות.
- **הבהירות** מתארת גם את הצג ותנאי התאורה, וניתן לשנותה כרצונו. הגודל הכמותי המתאר את הבהירות הוא **רמת האפור**, (Gray-Level), הנוטן מקום **מיספרי** בין רמת-השחור לרמת-הלבן המרביות.

איפון II : בהיקות ובהירות (המשך)

- רמות-אפור: בדר"כ רמת-האפור מתוארת ע"י סיביות ולכן תנוע בתחום 0-255 או לחלופין מנורמלת בין 0 ל-1.



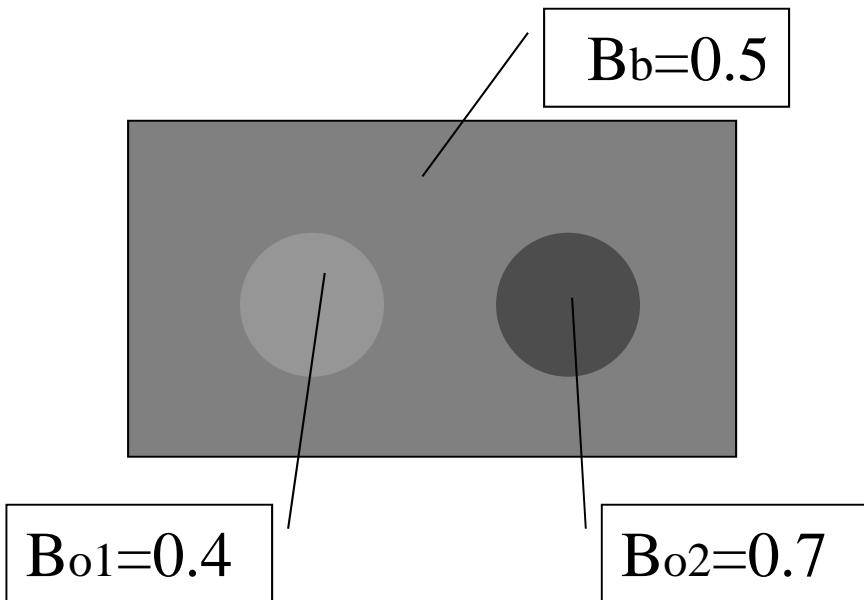
תמונה בה רמות
האפור נעות
בדרוגיות בין לבן
לשחור

איפיון III : ניגודיות

- הניגודיות מציינת את ההבדל בין רמות האפור של אובייקט לאלו של סביבתו.
- יכולת הבדיקה של העין תלואה מאד בניגודיות ורבה פחות בבהירותו !
- ניגודיות מקומית מוגדרת ע"י :

$$Contrast = \left| \frac{B_{object} - B_{background}}{B_{background}} \right|$$

איפיון III : ניגודיות (המשך)



$$C_1 = \left| \frac{B_{o1} - B_b}{B_b} \right| = \frac{0.5 - 0.4}{0.5} = 0.2$$

$$C_2 = \frac{0.7 - 0.5}{0.5} = 0.4$$

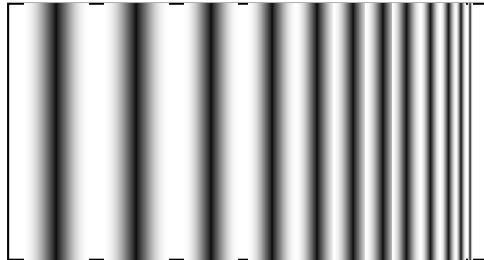
- במקרה זה לעיגול הימני ניגודיות חזקה יותר כלפי הרקע מאשר לימני.

איפיון IV : כושר הפרדה

- כל תמונה ספרטית מתוארת ע"י מערך של נקודות (פיקסלים), המהוות כל אחת **דגם מרחבית** של התמונה האמיתית.
- ככל שמספר הפיקסלים וצפיפותם יהיו גדולים יותר, יהיה ניתן להבחין בפרטים "עדינים" יותר.
- מספר הפיקסלים בכלל כיוון קובע את "**גבול ההפרדה**" של התמונה.
- בפועל, מספר הפיקסלים האקטיביים בתמונה קטן מהגבול התאורטי.

איפיון: IV כושר הפרדה (המשך)

- מדידת ההפרדה נעשית ע"י תמונת קווים מחזוריים:



למדידת ההפרדה האפקית השתמש בקווים אנכיים בצפיפות משתנה, בפילוג סינוסי (כמו כאן) או, בדרך"כ, פשוט בקווים שחור-לבן.

למדידת ההפרדה בתמונה מספיק לתאר כמות קווי הפרדה אפקית ואנכית .

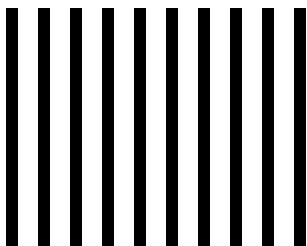
במשיר תצוגה (צג) או הדפסה מתייחסים גם לגודל הפיסי :
צפיפות נקודות ליחידה אורך (DPI).

איפיוו: V תדרים מרחביים

- השימוש ביצוג ע"י **תדר זמן** מוכר היטב באותות חד-ממדיים, ומקובל גם בדו-ממד, בו הוא מייצג **תדר מרחב**.
- טכניקת המעבר היא ע"י התמרת פורייה (Fourier Transform) ודומותיה (הרבות), בצורה הדו-ממדית שלhn.
- תכונות התמורות דו-ממדיות דומות בדר"כ לחדר-ממדיות (אם הן ספרביליות) ויוכרו בהמשך.

איפיוו: V תדרים מרוחביים (המשך)

- דוגמאות של תדר מרוחבי:



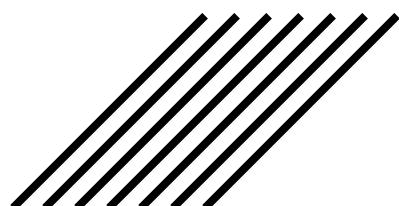
$$f_x = 30 \text{cycles / width}$$

$$f_y = 0 \text{cycles / height}$$



$$f_x = 0 \text{cycles / width}$$

$$f_y = 60 \text{cycles / height}$$



$$f_x = 60 \text{cycles / width}$$

$$f_y = 30 \text{cycles / height}$$

המספרים
מקורבים בלבד!
ניתן לחשב לפי
רוחב הדף, לפי
ס"מ, אינץ' וכו'.

איפיון VI : רعش (לכלוד ?)

- ה „רעש“ (מונח הלקוח מעיבוד אות דיבור...) הינו ערך אקראי ובلتி רצוי, הנוסף בכל אחד משלבי ייצור התמונה ונוסף למידע הרצוי.



תמונה Lena
המקורית, בתוספת
רעש גaussיאי לבן

$$f(x, y)$$

$$g(x, y) = f(x, y) + n(x, y)$$

איפיון VI רעש (המשך)

- איפיון הרעש מתמקד בדרך “ב” בתלותו במקום בתרמונה, ובעצמתו. (מניחים **אי תלות בתמונה עצמה**).
- מניחים (אם אין מידע אחר) אי-תלוות בגורמים ידועים ולכן הרעש הוא אקראי.
- שיערוד **סטטיסטי** של הרעש אפשרי, באמצעות פונקציית פילוג הסתברות, כשהഫילוג הגאוסי (או ה-“נורמלי”) הוא אחד המקובלים.

רעד : פילוג גאוסי (I)

- לפי הפילוג הגaussיאני, ההסתברות שערך המוקומי של הרעד יהיה בין שני ערכים סמוכים $n, n + \Delta n$

הנו :

$$p_{(n)} = \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{n^2}{2\sigma_n^2}} \cdot \Delta n$$

Standard Deviation : סטיית התקן : σ_n

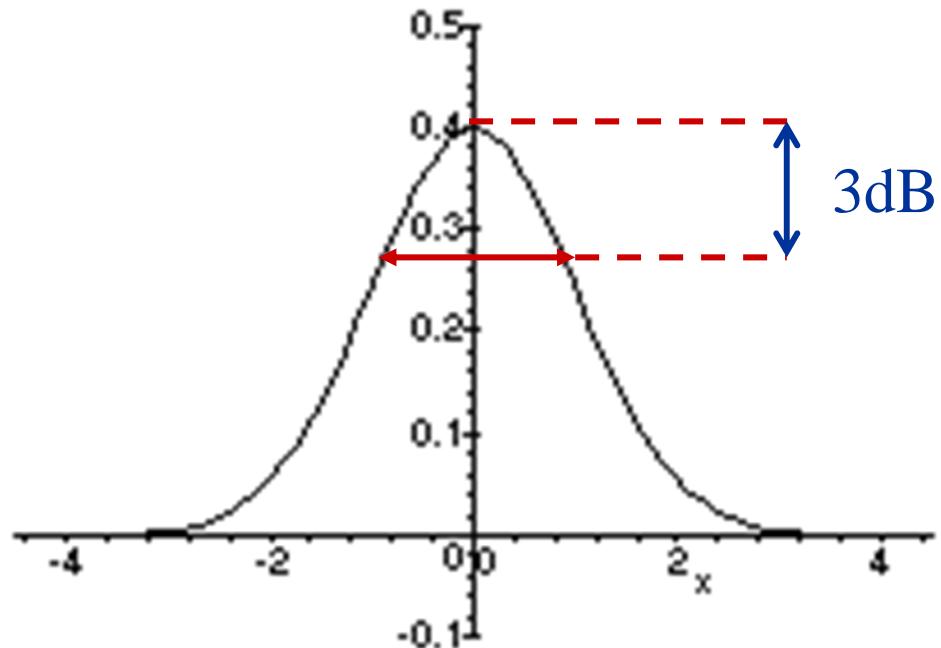
Variance : שוננות : σ_n^2



Carl Friedrich Gauss, 1777-1855
45

רעד : פילוג גאוסי - עקומת פעמו

(II)



רעש : פילוג גאוסי (III)

- **תכונות של רעש גaussiano:**
 - יכול לקבל ערכים חיוביים ושליליים
 - ההסתברות יורדת ככל שערך הרעש עולה
 - הערך הממוצע של משתנה אקראי נקרא **תוחלת**

N_v, N_h
מס' הפיקסלים
בשני הצירים

$$E_n = \frac{1}{N_v N_h} \sum_i \sum_j n_{ij}$$

ומחשב לפי:

$$\sigma_n^2 = \frac{1}{N_v N_h} \sum_i \sum_j (n_{ij} - E_n)^2$$

ואת השונות לפי:

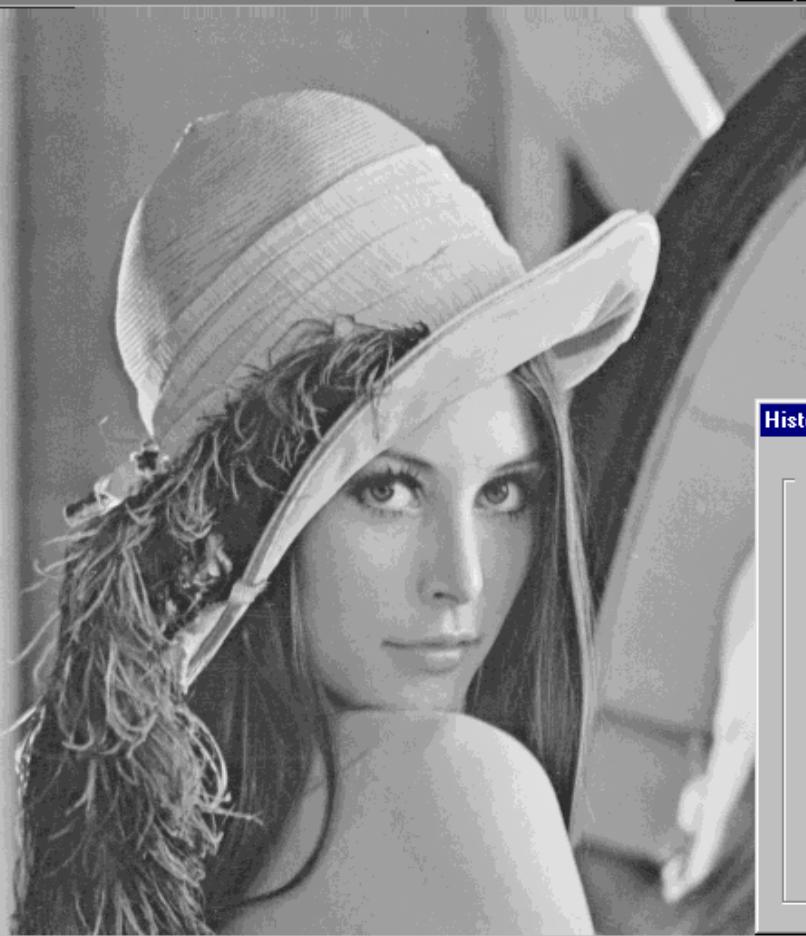
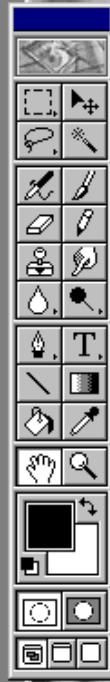
רעש : פילוג גאוסי (IV)

- **תוחלת הרעש** (Expected Value) היא ערך הרעש במרכז הפעמון (הסימטרי)
- **השונות** מייצגת את ממוצע ריבוע המרחק בין ערכי הרעש השונים לבין התוחלת.
- **תוחלת** : עבור משתנה אקראי, ממוצע : עבור משתנה דטרמיניסטי

ההיסטוגרמָה של תמונה

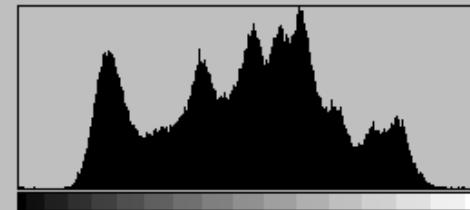
- התפלגות רמות האפור בתמונה מתוארת ע"י ההיסטוגרמָה: צבירת כמות הפיקסלים בכל רמת אפור, גובה כל עמודה מתאר את כמות הפיקסלים.
- בדוגמאות הבאות:
 - תמונה Lena עם הההיסטוגרמָה שלה.
 - תמונה Lena בתוספת רעש יוניפורמי + ההיסטוגרמָה.
 - תמונה חלקה
 - בתוספת רעש גאוסי

lena_bw.TIF @ 100%



Histogram

Channel: Black



OK

Mean: 124.06

Level:

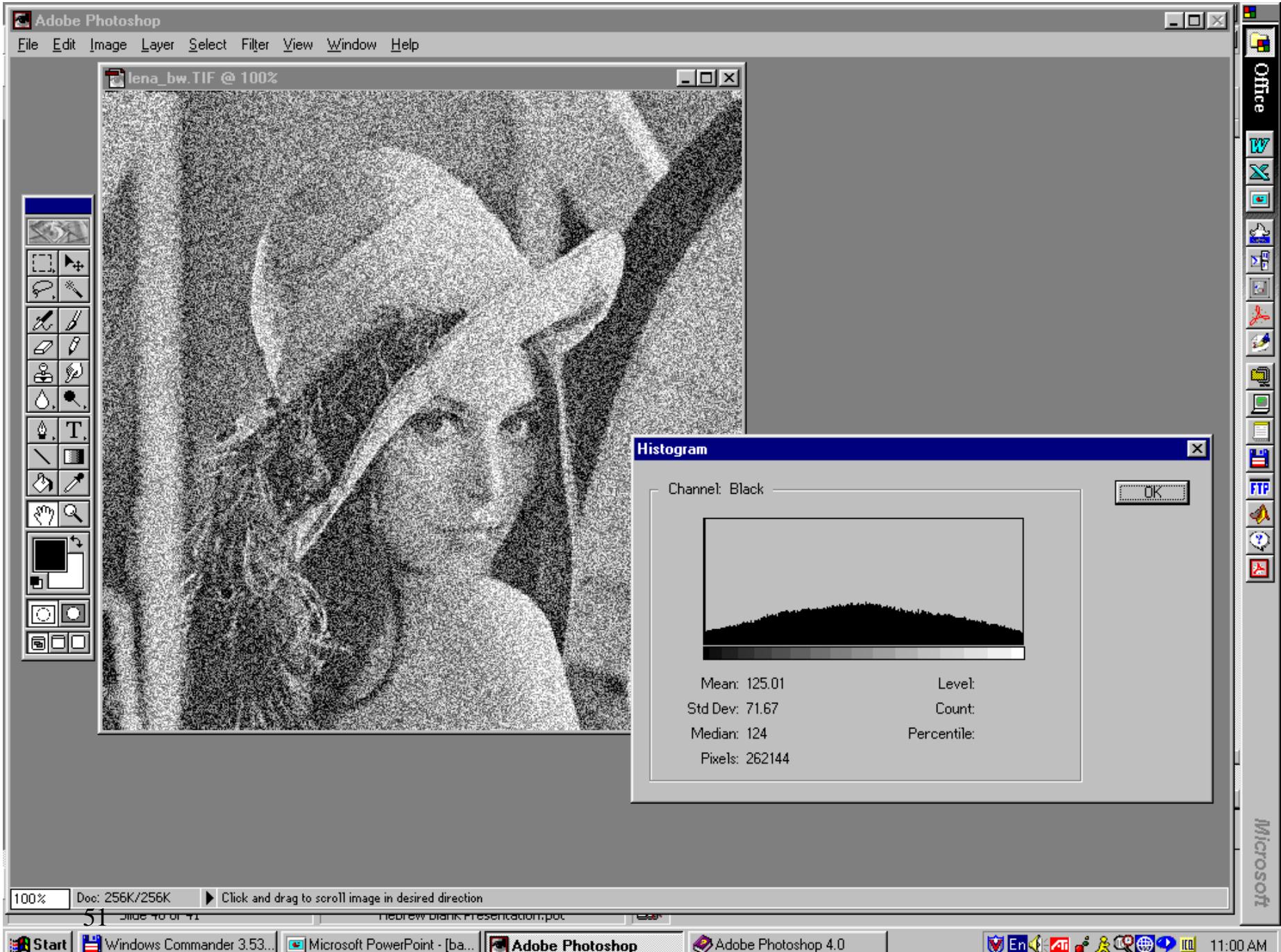
Std Dev: 47.85

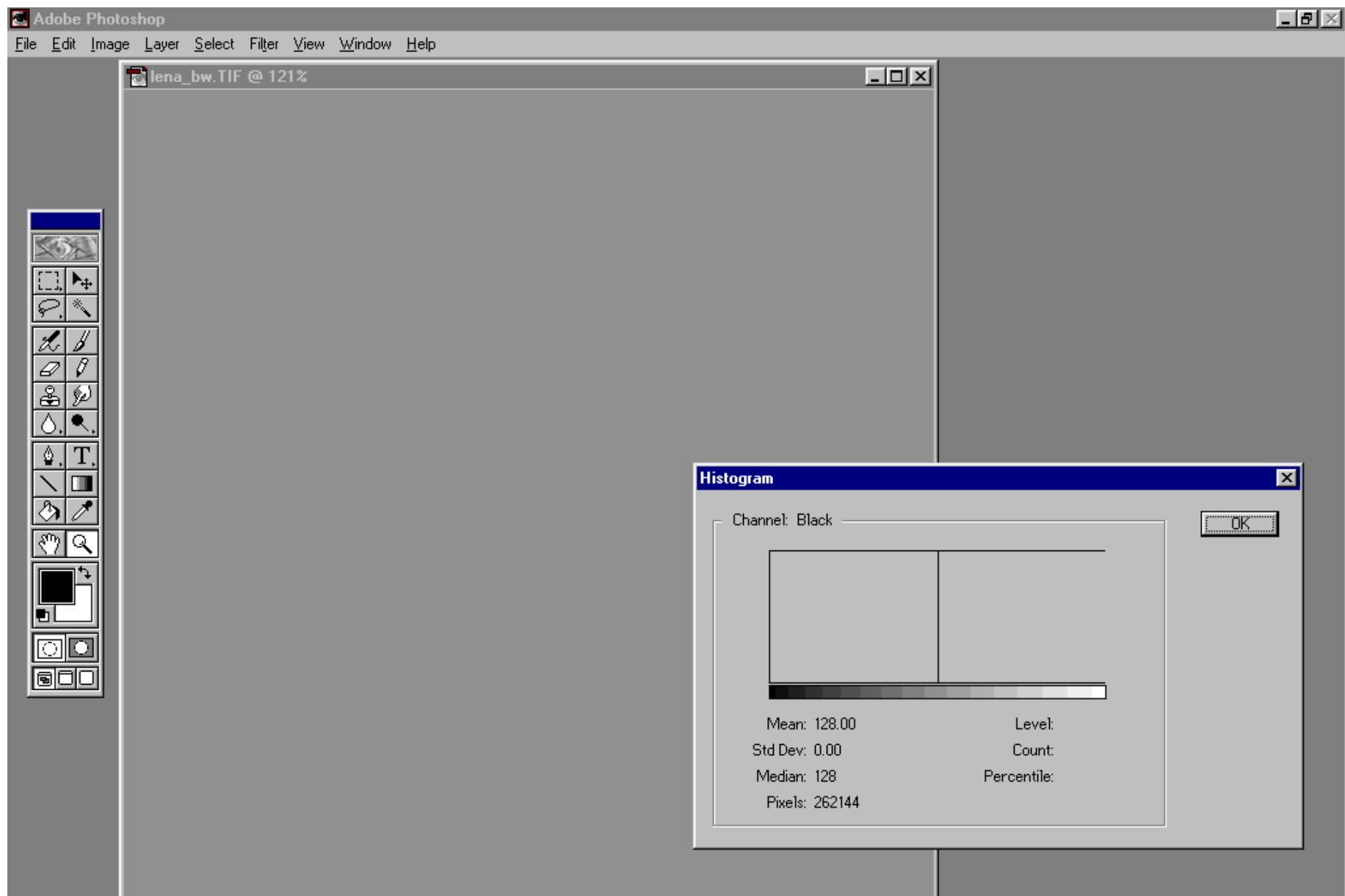
Count:

Median: 129

Percentile:

Pixels: 262144





Doc 256K/OK

 Zoom in

Start

52

32

Commander 3.53..

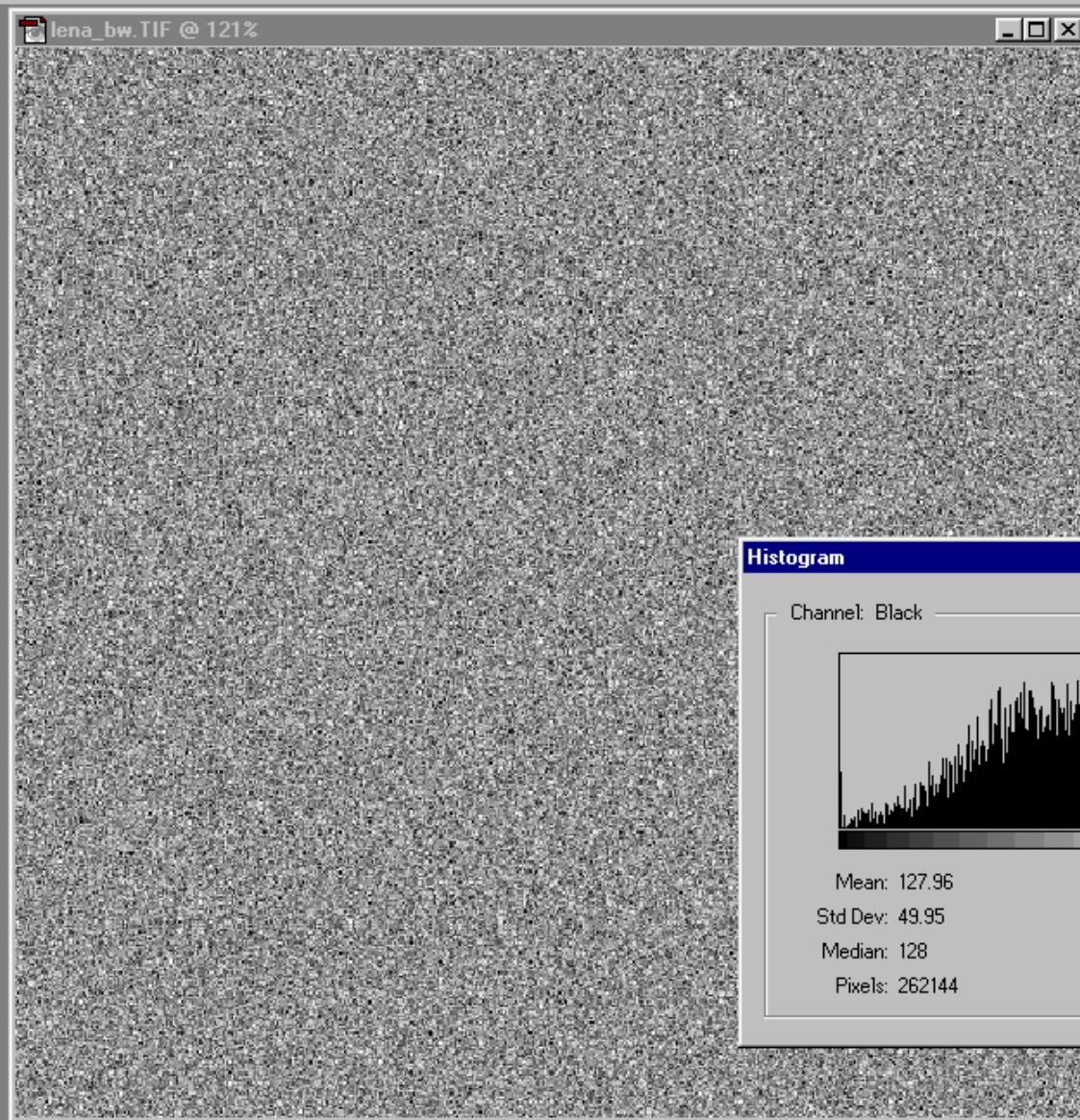
Microsoft PowerPoint - [ba..]

Adobe Photoshop

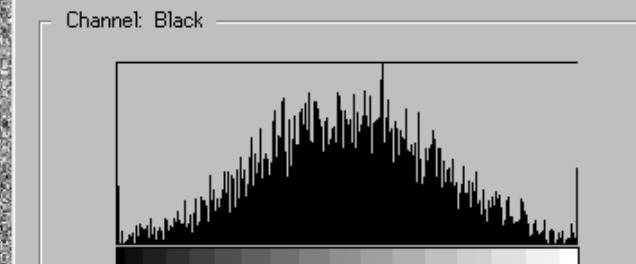
11:41 AM



lena_bw.TIF @ 121%



Histogram



Mean: 127.96

Std Dev: 49.95

Median: 128

Pixels: 262144

Level:

Count:

Percentile:

מדידת רעש בתמונה : MSE

- שגיאה ריבועית ממוצעת – Mean Squared Error –
- הבסיס הוא השוואה בין שתי תמונות (או אותות אחרים כMOV), כאשר אחד "מקור" והשני "תוצאה"
- נמדד את השוני ביניהם, ולפי השוני ניתן איפיון של **דמיון** (=איכות) או **חלופין שגיאה** (=עיות).
- $\mathbf{x} = \{X_i | i = 1, 2, \dots, N\}$ and $\mathbf{y} = \{Y_i | i = 1, 2, \dots, N\}$
are two finite-length, discrete signals (e.g. Images)

$$MSE_{(X, Y)} = \frac{1}{N} \sum_{i=1}^N (X_i - Y_i)^2$$

$e_i = X_i - Y_i$ is the "error" signal

נורמה כללית

- ה"נורמה" היא צורה כללית יותר לבטא זאת :

$$d_{p(X,Y)} = \left(\sum_{i=1}^N |e_i|^p \right)^{1/p}$$
 is the norm l_p

- בעיבוד תמונות בדרך כלל נבטא זאת ע"י יחס האות לרעש – **PSNR**

יחס אות לרעש :

- כמות ההפראה שנוספת לתמונה נמדדת ע"י **יחס בין אות המקורי לרעש הנלווה**. Signal to Noise Ratio.
- נהוג להגדיר את **כמות המידע** ("**אנרגייה**") באות כלשהו, כשונות (Variance) שלו. לפיכך, "גודל" האות יוגדר:

$$\sigma_s^2 = \frac{1}{N_v N_h} \sum_i \sum_j (S_{ij} - E_s)^2$$

Si,j : ערך האות בנקודה (i,j)

E_s : תוחלת בהירות האות

יחס אות לרעש SNR (המשך)

- ואילויחס אות לרעש SNR יוגדר **בסקללה לוגריתמית**:

$$SNR = 10 \log_{10} \left(\frac{\sigma_s^2}{\sigma_n^2} \right) = 10 \log_{10} \left(\frac{\sum_i \sum_j (s_{ij} - E_s)^2}{\sum_i \sum_j (n_{ij} - E_n)^2} \right)$$

- יש לשים לב כי עבר רעש גאוסי לבן בעל ממוצע אפס, בתמונה בעלת מס' **סופי** של נקודות N השאיתה לאפס היא יחסית ל- \sqrt{N} .

יחס אות לרעש : PSNR :

- מדידה המתאימה יותר לתמונות ולמערכת הראייה נקראת Peak-SNR ומחושבת לפי :

$$PSNR = 10 \log_{10} \frac{(2^n - 1)^2}{MSE}$$

כאשר n הוא מספר הסיביות של הדגימות –(Cluster במונה נמצא רבע הערך הגבוה ביותר האפשרי באות).

מגבלות השימוש ב- PSNR

- המדידה דורשת ידיעת תמונה המקור !
- המדידה הינה חסית למדידות מקבילות ואין משמעות למספרים עצם .
- מאיצים רבים מושקעים בפיתוח שיטות מדידת aicota אובייקטיביות כדוגמת התקנים :

ITU-R BT.500-10 –

P.900 –

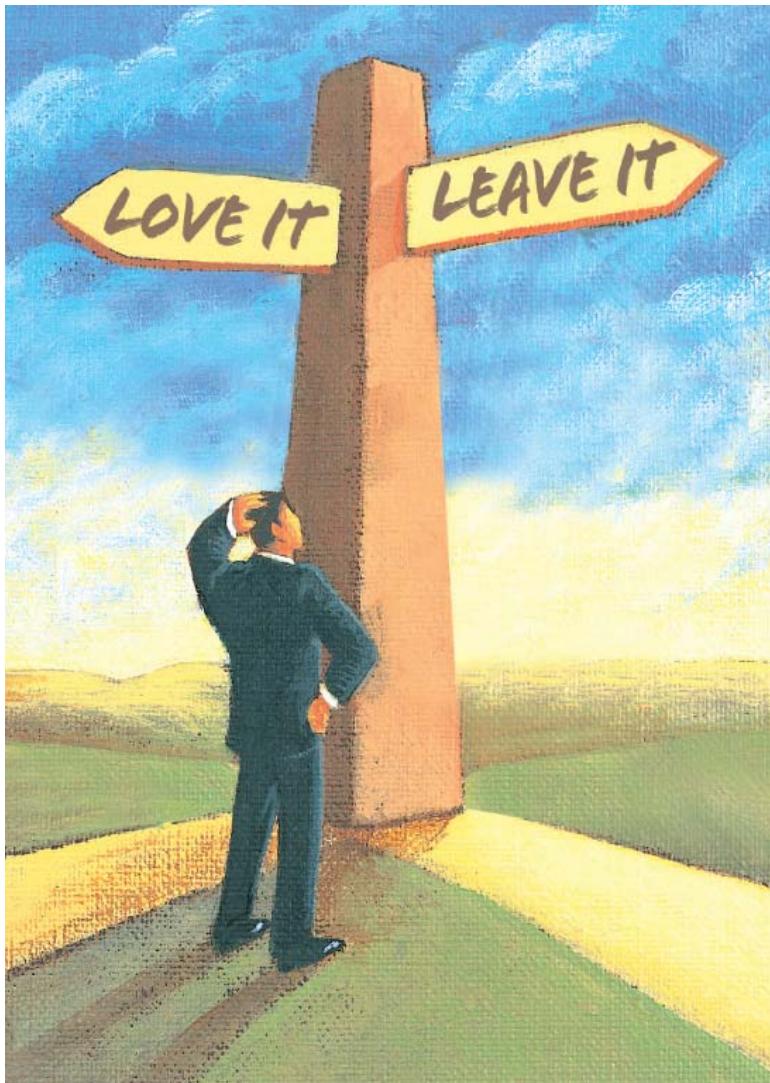
ITU-T Video Quality Expert Group (VQEG) –

תחום דינמי של אות

- התחום הדינמי (Dynamic Range) הוא הערך המרבי של SNR שניתן לקבל בתמונה **עבור רמת רעש נתונה**. כלומר: מספר הפעמים שהרעש **”נכנס“** באות המרבי.



Appendix : MSE and SSIM



Based on:

Zhou Wang and Alan C. Bovik,

Mean Squared Error:

Love It or Leave It?

IEEE Signal Processing Magazine,
January 2009

Why MSE ?

1. It is **simple**. It is parameter free and inexpensive to compute,
only one multiply and two additions per sample.
2. It is also **memoryless**: evaluated at each sample, independent of other samples.
3. All l_p norms are **valid distance metrics** in \mathbf{R}^N ,
 - Nonnegativity: $dp(x, y) \geq 0$
 - Identity: $dp(x, y) = 0$ if and only if $x = y$
 - Symmetry: $dp(x, y) = dp(y, x)$
 - Triangular inequality: $dp(x, z) \leq dp(x, y) + dp(y, z)$.
4. In particular, the $p = 2$ case (proportional to the square root of the MSE) is the **ordinary distance metric** in *N-dimensional Euclidean space*.
5. It has a **clear physical meaning**: it is the natural way to **define the energy** of the error signal. Such an energy measure is preserved after any orthogonal linear transformation (Fourier etc)

The energy preserving property guarantees that the energy of a signal distortion in the transform domain is the same as in the signal domain. This property distinguishes *d2 from the other l_p energy measures, which are not energy preserving*.

Why MSE ? Cont'd

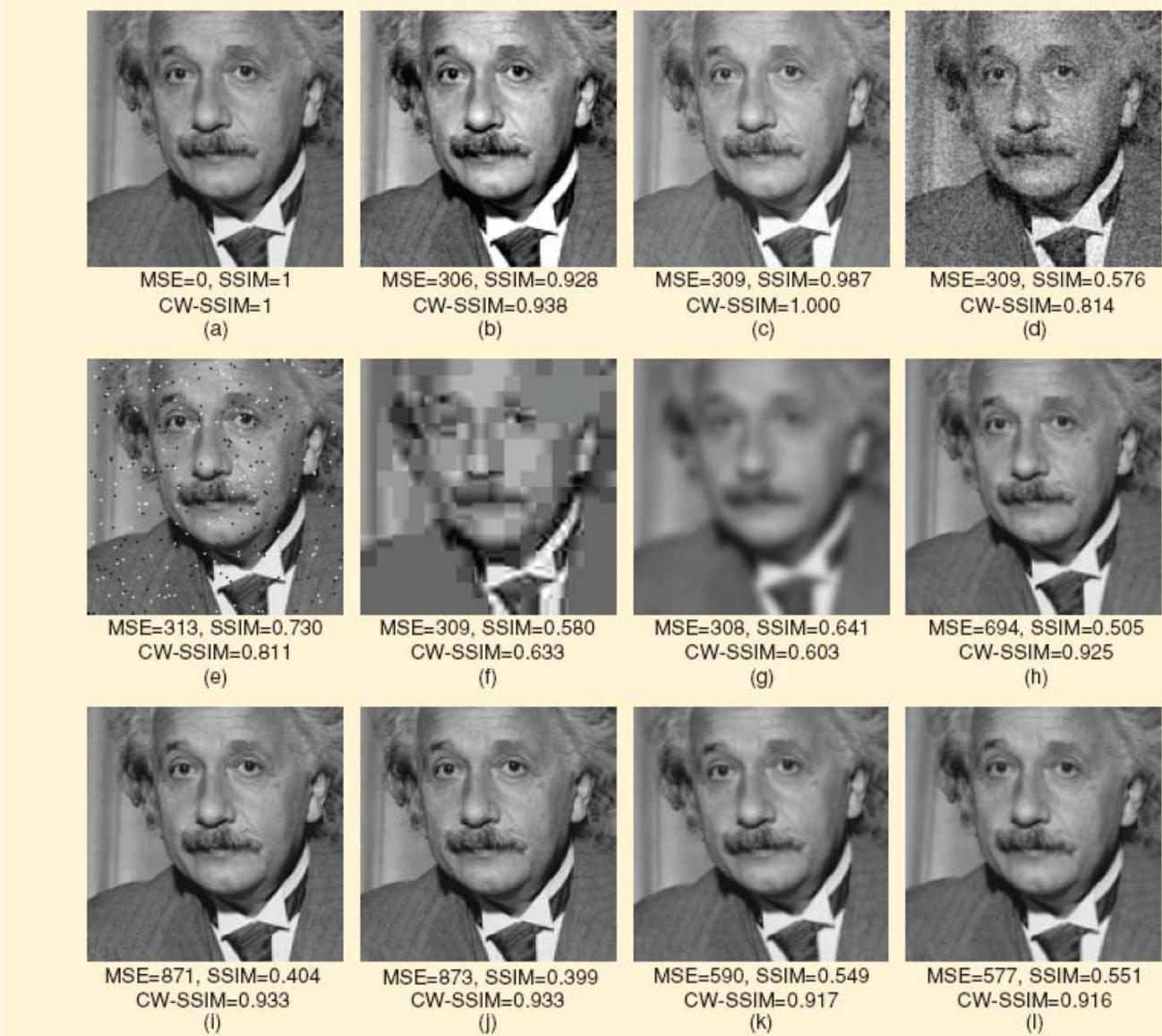
6. The MSE is an excellent metric in the context of **optimization**. The MSE possesses the very satisfying properties of convexity, symmetry, and differentiability. Minimum-MSE (MMSE) optimization problems often have **closed-form analytical solutions**.
 - when they don't, iterative numerical optimization procedures are often easy to formulate, since the gradient of the MSE is easy to compute.
7. MSE is **widely used** simply because it is a **convention**.
 - Historically, it has been employed extensively for optimizing and assessing a wide variety of signal processing applications, including filter design, signal compression, restoration, de-noising, reconstruction, and classification.

Why NOT MSE ?

- The more fundamental issue has been missing: does the MSE really measure signal fidelity?
 - Given all of its above-mentioned attractive features, a signal processing practitioner might choose the MSE if it proved to be a reasonable signal fidelity measure.
- But is that the case?**

Comparison of Image Fidelity Measures

Zhou Wang and Alan C. Bovik,
Mean Squared Error:
Love It or Leave It?
IEEE Signal Processing Magazine ,
January 2009



[FIG2] Comparison of image fidelity measures for "Einstein" image altered with different types of distortions. (a) Reference image. (b) Mean contrast stretch. (c) Luminance shift. (d) Gaussian noise contamination. (e) Impulsive noise contamination. (f) JPEG compression. (g) Blurring. (h) Spatial scaling (zooming out). (i) Spatial shift (to the right). (j) Spatial shift (to the left). (k) Rotation (counter-clockwise). (l) Rotation (clockwise).

An alternative: Structural Similarity

- A framework of **image fidelity measurement** as an image communication problem
- The SSIM proves high effectiveness for measuring the **fidelity** of signals
- The principle underlying the SSIM approach is that the human visual system is highly adapted to extract **structural information** from visual scenes

Basic Form of SSIM

- Suppose that x and y are local image patches taken from the same location of two images that are being compared.
- The local SSIM index measures the similarities of three elements of the image patches:
 - $l(x, y)$ of the local patch **luminances** (brightness values)
 - $c(x, y)$ of the local patch **contrasts**
 - $s(x, y)$ of the local patch **structures**
- These local similarities are expressed using simple, easily computed statistics, and combined together to form **local SSIM**

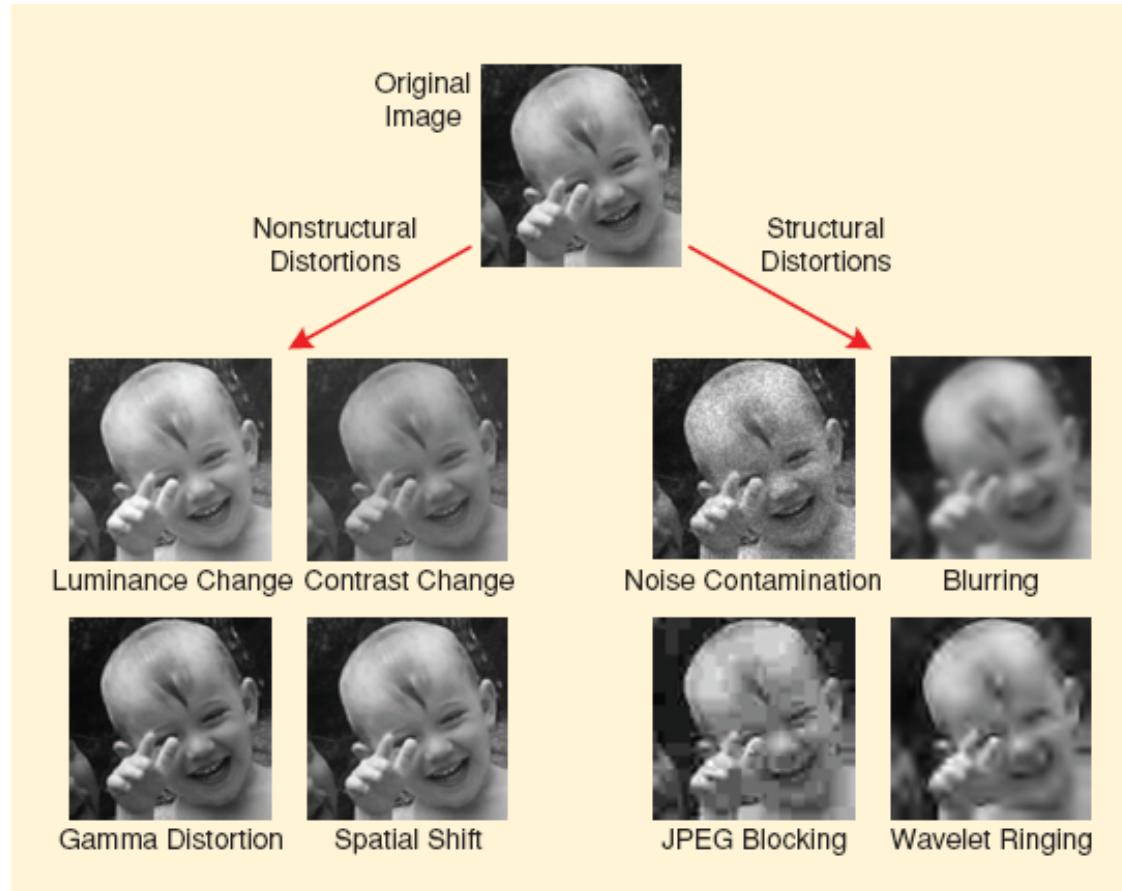
SIMM Calculation

$$S(x, y) = l(x, y) \cdot c(x, y) \cdot s(x, y) = \\ = \left(\frac{2\mu_x\mu_y + C_1}{\mu^2_x + \mu^2_y + C_1} \right) \cdot \left(\frac{2\sigma_x\sigma_y + C_2}{\sigma^2_x + \sigma^2_y + C_2} \right) \cdot \left(\frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \right)$$

Where:

- μ_x and μ_y are (respectively) the local sample **means of x and y**
- σ_x and σ_y are (respectively) the local sample **standard deviations of x and y**
- σ_{xy} is the sample **cross correlation of x and y** after removing their means.
- The items C_1, C_2, C_3 are **small positive constants that stabilize each term**, so that near-zero sample means, variances, or correlations do not lead to numerical instability

Structural vs. Nonstructural Distortions



Example

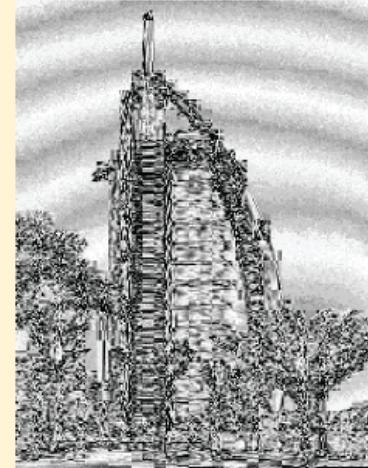
- JPEG induced annoying **pseudo-contouring effects** (in the sky region) and **blocking artifacts** (along the boundaries of the building) that are successfully captured by the SSIM index, yet poorly predicted by the absolute error map.



(a)



(b)



(c)



(d)

[FIG8] Comparison of image fidelity/distortion maps. (a) Reference image. (b) JPEG compressed image. (c) Absolute error map of the distorted image (enhanced for visibility). (d) SSIM index map of the distorted images (enhanced for visibility).