<u>ראייה אנושית גישה חישובית</u>

<u>שיעור 1, 02/03</u>

ראייה ממוחשבת- לגרום למכונה לראות.

ראייה חישובית- להבין איך עובדת מערכת הראייה האנושית.

יש אינטרקציה בין שני התחומים.

מה זה 'להבין'?

1. פרדיגמת התא הבודד single cell paradigm-

מציאת תא אחד ספציפי שעושה כל מטלה. יובל וויזל מצאו כי יש במוח תאים שמגיבים לראייה של קווים באוריינטציה מסויימת ולא לקווים באוריינטציה אחרת. התפיסה הייתה שישנם תאים שמגיבים לקווים אופקיים, אנכיים, לפינות, ובאותו אופן תאים מורכבים יותר שמגיבים לפרצופים וכו'. כונתה בלעג תיאוריית grandmother cell- קיום של תא שמגיב אך ורק כשרואים את סבתא.

המחשבה שלכל פעולה יש תא ספציפי שאחראי עליה היא לא סבירה. יש המון מטלות שאנחנו מסוגלים לבצע וזה נראה לא הגיוני לתכנן מערכת כך.

2. פרדיגמת הcircuit-

אוסף של תאים ויזואלים שעוזרים לבצע מטלה. תת קבוצה של התאים מגיבים. יש הרבה עניין בתיאוריה הזה. ניסו לאפיין קבוצות כאלו של נוירונים. לא היו הרבה הצלחה, יש מספר אקספוננציאלי של תתי קבוצות וקשה לבחון את פעולתן.

3. מודולים של חשיבה-

נפוצה בעשור האחרון. FMRI (functional magnetic resonance imaging), מכשיר מגנטי שמודד תהודה. ניתן באמצעותו לקבל מפה של פעילות במוח. כאשר איזור מסויים במוח פעיל יותר מאחרים, המוח מזרים יותר דם לאותו איזור במוח, יש יותר חמצן ויש שינוי בתהודה המגנטית. אפשר למדוד את הפעילות במוח כתוצאה מגירוי. למשל ניתן לראות כי חלקים אחרים במוח פעילים כשמציגים לנבדק תמונות של אנשים לעומת תמונות של בתים. התפיסה היא שניתן למפות את האיזורים השונים האלו.

4. גישה חישובית-

David Marr - Vision: A Computational Approach.

"לנסות להבין תפיסה על ידי חקר של נוירונים בלבד, זה כמו לנסות לחקור תעופה של ציפור על ידי חקר של נוצות בלבד", צריך להבין גם את ה'אוירודינמיקה'.

נדבר על תפיסת תנועה, תפיסת צבע ותפיסת צורה.

דוגמה לתפיסת תנועה- אליפסה מסתובבת על הציר, לפעמים נראית כמסתובבת קשיחה על הציר ולפעמים

נראית כאילו היא מתרחבת. סביבה יש נקודות סביב האליפסה, אופן התזוזה שלהן תורם לאשליה. נרצה לבדוק אם יש תא מסויים שפעיל בשלב הראשון של הסרטון ולא פעיל בשלב השני, אך גם אם היינו מוצאים תא כזה, קבוצת תאים כאלו או איזור במוח כזה, זה לא היה מספק לנו הסבר מלא לתופעה.
דוגמה לתפיסת צבע- לוח שחמט שמוטל עליו צל מעמוד, שני ריבועים שהם בדיוק באותו צבע אבל נראים בצבע אחר לגמרי. נניח ונמצא תא שמגיב למשבצות בהירות ותא אחר שמגיב למשבצות כהות, אבל זה לא היה מסביר את התופעה.
דוגמה לתפיסת צורה- שני שולחנות, נדמה שהשולחנות אינם באותם מימדים גיאומטרים, השמאלי ארוך וצר והימני כמעט מרובע. מבחינה גיאומטרית השולחנות זהים. גם כאן שלושת הגישות שהזכרנו לא היו נותנות הסבר לתופעה, גם אם היו מצליחות לאפיין את החלק הרלוונטי במוח.

<u>3 הרמות של הסבר לפי Marr</u>

1. הרמה החישובית:
   באה לענות על השאלות- מה הבעיה שמערכת הראייה מנסה לפתור? למה הבעיה קשה? אילו הנחות על העולם הופכות את הבעיה לקלה יותר? מה ההגיון שמאחורי החישוב?
2. הרמה האלגוריתמית:
   מה הייצוג של הקלט? מה הייצוג של הפלט? מה האלגוריתם שמעביר את הקלט לפלט?
3. רמת המימוש:
   איך החישוב והאלגוריתם ממומשים בחומרה?

התיאוריה תקפה לכל תופעה שקשורה לחישוב. למשל נרצה להבין איך עובד מחשבון.
הרמה החישובית: בעיה שהמחשבון מנסה לפתור היא בהנתן a,+,b, לחשב את a+b.
הרמה האלגוריתמית: ניתן לייצג בספרות ערביות או בספרות רומיות. הייצוג הרומי יותר קשה לביצוע פעולות חשבון.
רמת המימוש: אפשר לממש בלפטופ, בטלפון ועוד, כל צורה דורשת מימוש שונה.

אנחנו נדבר רק על הרמה החישובית. תאי העצב והנוירונים רלוונטים לרמת המימוש. יש כאלו שמאמינים שאין קשר בין הרמות ויש כאלו שמאמינים שחייבים להבין את הרמה החישובית לפני שפונים לחקור את רמת המימוש, או להפך.

נחזור לסרטון האליפסה.
מה הבעיה שמערכת הראייה מנסה לפתור? חישוב תנועה בהנתן רצף תמונות.
צביעה של סרטי שחור לבן בצבע: נרצה לבנות אלגוריתם שמקבל סימוני צבע לפריים הראשון ואז עוקב אחרי כל פיקסל וממשיך לצבוע אותו בכל הפריימים בצבע המתאים. בפועל קשה מאד לעשות את זה - חלקים שונים

בפריים זזים בצורות שונות וזו בעיה שתחום הראייה הממוחשבת טרם הצליח לפתור במלואה, אך מערכת הראייה האנושית עושה היטב. נרצה להבין את ההנחות על העולם שבני אדם משתמשים בהם כשהם מפענחים את הסרט ונוכל להעזר בהן כדי להבין את תופעת האליפסה.

לוח השחמט.

מה הבעיה שמערכת הראייה מנסה לפתור? חישוב רמת הצבע של אובייקטים בעולם.

זו נדמית להיות בעיה קלה, נוכל לבדוק אמפירית את רמת הצבע, אבל יש תופעה של הצללה ותאורה שמערכת הראייה מתחשבת בה. למשל אנחנו יודעים שצבע הכביש הוא אפור גם כשמוטל עליו צל והוא נצבע בצבע כמעט שחור.

השולחנות.

מה הבעיה שמערכית הראייה מנסה לפתור? חישוב צורה של אובייקטים בעולם.

התמונה היא בדו-מימד והעולם הוא תלת-מימדי, חסרה לנו הקוארדינטה השלישית כדי לפענח את התמונה בצורה מדוייקת. נצטרך להניח הנחות על העולם כדי לעשות את החישוב הזה, נרצה ללמוד מהן ההנחות האלו.

<u>שיעור 2, 09/03</u>

נתחיל את הנושא של תפיסת תנועה. נדבר על תופעות בתפיסת תנועה.

פסיכופיזיקה של תפיסת תנועה- מושג שהוגדר ב-1860 על ידי גוסטב פכנר, נסיון להגדיר תורה מדעית סגורה של הקשר בין העולם האמיתי לתפיסה שלנו את העולם האמיתי. פסיכולוגיה (Ψ)- התופעה התפיסתית, פיזיקה (Φ)- גירוי פיזיקלי.

למשל נשמיע צליל בעוצמות שונות. ניתן למדוד את עוצמת הצליל על ידי מדידת האמפליטודה של הגל וניתן לשאול כמה חזק שומעים.

הרעיון הוא למצוא קשר ביניהם ולנתח אותו, כאשר לרב אין קשר לינארי.

בהקשר של תפיסת תנועה, נוכל להקרין סרט (צ'רלי צ'פלין בשחור לבן) ולבחון את התגובה לתנועתיות בו. זה בעייתי כי התנועה בסרט מאד מורכבת וכך גם התפיסה שלנו. מוטב לבודד תופעה פשוטה יותר ולחקור אותה.

הפשטה ראשונה: נחקור רק תנועות דו-מימדיות אחידות במרחב התמונה. אפשר לתאר כך את העולם הפיזיקלי Φ באמצעות (v_x, v_y) בלבד.

הפשטה שנייה: אובייקטים מאד פשוטים- קווים.

נדבר על ניסויים שנעשו על ידי Hans Wallach בשנים 1920-1930.

בדק קווים ולא נקודות כי עם נקודות התוצאות צפויות ולכן לא מעניינות. עם קווים עולה בעיה הנקראת 'בעיית הצמצם'. הכיוון אליו הקו זז יראה לנו אחרת כתלות בחריר שדרכו אנחנו מסתכלים על התנועה.
בניסוי הראשון וולאך לקח קו אחד גדול מאד (לכאורה אינסופי) והזיז אותו למעלה ולמטה. נראה שהקו זז באלכסון. כשמדובר על קו אינסופי, אם נזיז אותו למעלה-למטה, הצדה או באלכסון.
http://web.mit.edu/persci/demos/Motion&Form/mini.html

בעיית הצמצם (Wallach)- נגדיר $M(\theta,\ v)$ סרט של קו אינסופי בזווית θ שנע בכיוון v. נפרק את v לשני רכיבים $v^{\perp}, v^{\theta}$ כאשר v^{θ} מקביל לקו ו- $v^{\perp}$ ניצב לקו. אז $M(\theta,\ v^{\theta}{}_1, v_1{}^{\perp}) = M(\theta,\ v_2{}^{\theta}, v_2{}^{\perp})$ לכל $v_1{}^{\theta}, v_2{}^{\theta}$.
אפשר לחשוב על זה כאילו נקבל את אותו סרט עבור כל ה-vים שה- $v^{\perp}$ שלהם שווה. הסיבה לכך היא סימטריה, נוכל להזיז קודם לפי $v^{\perp}$ ואז לפי v^{θ}, קודם בניצב לקו ואז במקביל (מחיבור וקטורים). התזוזה ב- $v^{\perp}$ זהה והתזוזה ב-v^{θ} היא חסרת משמעות כי הקו אינסופי (הקו סימטרי להזזות המקבילות לו).

שאלה תפיסתית- איך $M(\theta,\ v^{\theta}, v^{\perp})$ יתפס?
יש אינסוף וקטורים שיכולים לייצג את התנועה של הקו. בפועל נראה שהקו זז באלכסון, כלומר $M(\theta,\ 0,\ v^{\perp})$, בלי תזוזה מקבילה.

הגדרה: 'המהירות הנורמלית'.
עבור קו בזווית θ שנע במהירות $v = v^{\perp} + v^{\theta}$, המהירות הנורמלית היא $v^{\perp}$.
תופעה: קו בזווית θ שנע במהירות v נתפס כנע במהירות הנורמלית $v^{\perp}$.
טענה: מתוך כל המהירויות $v = v^{\perp} + v^{\theta}$ המהירות $v^{\perp}$ היא האיטית ביותר. ממשפט פיתגורס מתקיים כי לכל v אחר $|v|^2 = |v^{\perp}|^2 + |v^{\theta}|^2 \geq |v^{\perp}|^2$.
אפשרות נוספת היא שהתנועה $v^{\perp}$ היא היחידה שאין לה תנועה סימטרית.

ניסוי נוסף, מביטים על קוים אלכסונים שזזים, דרך חלונות שונים. דרך כל צורת חלון כיוון התנועה נראה שונה.
זה נקרא Barberpole Illusion (עמוד הספר).
http://www.liv.ac.uk/~marcob/Trieste/barberpole.html

תופעה: כאשר קו נצפה דרך חלון, התנועה הנתפסת תלויה בצורת החלון.
קשה יותר להסביר זאת מתמטית.

הגדרה: עבור קו שנע במהירות v בזווית θ ומוסתר על ידי קו בזווית θ_2, הטרמינטור היא הצומת בין הקו לבין מה שמסתיר אותו. מהירות הטרמינטורים היא ההטלה של v בכיוון θ_2.

תופעה: קו שמוסתר על ידי חלון, אם יש מהירות טרמינטורים דומיננטית, אז הקו יתפס כנע במהירות הזו.

הגדרה: עבור קו בזווית θ שנע במהירות v_ϕ , נפרק אותו ל $v_\phi = v_\phi{}^\perp + v_\phi{}^\theta$. קו האילוץ (Constraint Line) היא קבוצת כל המהירויות $v = v_\phi{}^\perp + v^\theta$ (קו האילוץ מקביל לקו המקורי θ , במרחק המהירות הנורמלית).

נתבונן בשני קווים, הראשון בזווית θ_1 ומהירות $v_{\phi 1}$, והשני בזווית θ_2 ומהירות $v_{\phi 2}$.

IOC היא נקודת המפגש של שני קווי האילוץ. נקודת ה-VA vector average הוא הממוצע הוקטורי של המהירויות הנורמליות.

<u>שיעור 3, 16/03</u>

נדבר על מודלים חישוביים, אלגוריתמים לחישוב תנועה. המטרה היא להבין את תפיסת התנועה אצל בני האדם ברמה החישובית.

קלט: $I(x,s,t)$ כמות האור שמגיעה בנקודה x,y בזמן t.

נניח שיש רק שתי נקודות זמן $I_1(x,s) = I(x,s,t_1),\ I_2(x,s) = I(x,s,t_2)$.

נפשט את הבעיה עוד יותר- חישוב תנועה במימד אחד.

קלט: $I_1(x),\ I_2(x)$

(x הקואורדינטה, $I(x)$ האור בקואורדינטה)

פלט: $\Delta x = v\Delta t$ התנועה בין 2 הפריימים.

יש לבצע הנחה כלשהי על העולם שתקשר בין מה שאנחנו רוצים לחשב לקלט שלנו.

הנחה שימור הבהירות (Constant Brightness Assumption):

$I_1(x) = I_2(x+v)$.

ההנחה הזו מאפשרת לנו לפתור את הבעיה, גם אם היא לא תמיד מדוייקת. נניח כי שינויים בבהירויות שאנחנו רואים נובעים מתזוזת אובייקטים ולא משינויים בתאורה או שינויים בצבע של האובייקט.

נוכל להגדיר פונקציה SSD, sum of square difference.

$$SSD(v) = \sum_x (I_1(x) - I_2(x+v))^2$$

חישוב תנועה יהיה $v = argmin_v SSD(v)$.

אלגוריתם אפשרי:

עבור מספר סופי של מהירויות v_i

נחשב $SSD(v_i)$

נבחר את v^* כך שייתן את המינימום

יתרונות:

קל מאד למימוש, מימוש מקבילי.

חסרונות:

אנחנו מוצאים את המינימום מבין המספר הסופי שבדקנו, אך לא בהכרח את המינימום האמיתי.

נרצה למצוא את המינימום בצורה אנליטית. לשם כך נרצה שפונקציית ה-SSD תהיה פונקציה פשוטה כפונקציה של v. אין לנו שליטה על הסיבוכיות של $I(x+v)$ כפונקציה של v, אז נעשה קירוב.

הגדרה: $\tilde{SSD}(v)$ הוא קירוב ל- $SSD(v)$ כאשר מבצעים קירוב טיילור מסדר ראשון ל $I_2(x+v)$.

$$I_2(x+v) = I_2(x) + v \cdot \frac{d\,I_2}{d\,x}|_x$$

$$\tilde{SSD}(v) = \sum_x (I_1(x) - I_2(x) - v \cdot \frac{d\,I_2}{d\,x})^2 .$$

כש-v מאד קטן הקירוב יותר מדוייק.

הערה: המינימום של $\tilde{SSD}(v)$ ניתן על ידי הפתרון של המשוואה הבאה $(\sum(\frac{d\,I_2}{d\,x})^2)_v = -\sum_x (I_2 - I_1)\frac{d\,I_2}{d\,x}|_x$.

הוכחה: $\frac{d\,\tilde{SSD}}{d\,v} = -2\sum_x (I_1(x) - I_2(x) - \frac{vdI_2}{dx})\frac{dI_2}{dx}$.

$$\frac{d\tilde{SSD}}{dv} = 0 \Rightarrow (\sum_x (\frac{dI_2}{dx})^2)v = -\sum_x (I_2(x) - I_1(x))\frac{dI_2}{dx}$$

בתנאים מסויימים, למשל מהירויות קטנות, $SSD(v) = \tilde{SSD}(v)$ ולכן שני האלגוריתמים יתנו בדיוק אותה תוצאה.

שני האלגוריתמים מניחים ששינויים בבהירות נובעים מתזוזת אובייקטים וכן שכל התמונה זזה ביחד, ולא רק חלק מהאובייקטים בה.

הרחבה לדו-מימד.

קלט: $I_1(x,y) = I_1(p)$, $I_2(x,y) = I_2(p)$, p הוא וקטור.

v וקטור התנועה.

הנחת שימור הבהירות: $I_2(p+v) = I_1(p)$.

$$SSD(v) = \sum_p (I_2(p+v) - I_1(p))^2 .$$

$\tilde{SSD}(v) = \sum_p (I_1(p) - I_2(p) - \frac{dI_2}{dp} \cdot v)^2$.

$I_2(p+v) \approx I_2(p) + \frac{dI_2}{dp} \cdot v$

הערה: המינימום של $\tilde{SSD}(v)$ מתקבל על ידי פתרון מערכת המשוואות הבאה

$$Mv^* = b$$

$$M = \sum_p (\frac{dI_2}{dp})(\frac{dI_2}{dp})^T$$

$$b = \sum_p (I_2(p) - I_1(p))\frac{dI_2}{dp}$$

הוכחה: איך לגזור פונקציות לפי וקטורים.

$$\frac{d\tilde{SSD}}{dv} = 2\sum (I_1(p) - I_2(p) - \frac{dI_2}{dp} \cdot v)^T \cdot \frac{dI_2}{dp}$$

מתי יש פתרון יחיד ל $Mv = b$?

M היא מטריצה 2×2 , הפיכה כאשר הדרגה שלה היא 2, $rank(m) = 2$. לא הפיכה כאשר הדרגה שלה היא 1 או 0.

מתי יתקיים כי $rank(m) = 0$?

כאשר $I_2(p) = c$. כאשר התמונה קבועה יש אינסוף פתרונות v שיקיימו שימור בהירויות.

מתי יתקיים $rank(m) = 1$?

אם לכל p מתקיים $\frac{dI}{dp}|_p = \alpha d$

$m = \sum \alpha d d^T$ (מטריצה מדרגה 1)

כאשר M מדרגה 1 יש אינסוף פתרונות v^* , קבוצת כל הפתרונות מהווה תת-מרחב מדרגה 1 במרחב הדו-מימדי.

תזכורת: זהו קו האילוץ.

וולאך הראה זאת בצורה פסיכופיזית וגיאומטרית, כעת הראנו על ידי אלגברה לינארית. כלומר הבעיה אינה קשורה לתפיסה של בני האדם אלא נובעת מחוק שימור הבהירות ולכן יש אינסוף אפשרויות לתנועה שכולן יושבות על קו אחד, קו האילוץ.

לוקאס קנדה משתמש בהנחה זאת.

<u>שיעור 4, 20/04</u>

אם אנחנו מניחים ש-$I(p,t)$ היא כמות האור בנקודה p בעין/מצלמה בזמן t, אז $I_1(p) = I(p, t_1)$

$I_2(p) = I(p, t_2) = I(p, t_1 + \delta)$.

$I_2(p) - I_1(p) = I(p, t_1 + \delta) - I(p, t_1) \simeq \frac{dI}{dt}\delta$

$$SSD = \sum_p (\delta \frac{dI}{dt} + v\frac{dI}{dp})^2$$

נניח $\delta = 1$

$$= \sum_p (\frac{dI}{dt} + v\frac{dI}{dp})^2 = \sum_p (\frac{dI}{dt} + u\frac{dI}{dx} + v\frac{dI}{dy})^2$$

u,v נעלמים.

סיכום- אלגוריתם לחישוב תנועה.

הקלט הוא שתי תמונות I1,I2 והפלט הוא u*,v* שמביאים למינימום את ה-SSD. הם מייצגים איך התמונה זזה. מחשבים בכל נקודה את הנגזרות לפי x,y,t. יוצרים מטריצה M בה סוכמים את הנגזרות בכל הנקודות, יוצרים וקטור b ופותרים את מערכת המשוואות שקיבלנו.

יכול להיות מצב שהמטריצה M אינה הפיכה, למשל היא יכולה להיות מדרגה 0 (רק אפסים) או מדרגה 1 (הנגזרת לפי x תהיה שונה מ-0 אבל לפי y תהיה 0). אז אפשר להוסיף קבוע קטן λ למטריצה וכך להבטיח שהיא תהיה הפיכה.

קבלת החלטות בייזיאניות.

דוגמה: מדחום מראה שהטמפרטורה כרגע בחוץ היא 21°.

נרצה לדעת האם זה נכון. המדידה נכונה עד כדי שגיאה.

בנוסף, יש לנו ידע מוקדם על מזג האוויר בירושלים ב-20/04 לאורך השנים. נוכל לראות שבממוצע בתאריך הזה הטמפרטורה היא 25° עם סטיית תקן של 3°.

אז מה באמת הטמפרטורה כרגע בחוץ? נרצה לחבר את הידע המוקדם ואת המדידות בצורה אופטימלית.

שיערוך בייזיאני- באמצעות חוק בייס, $P(A|B) = \frac{P(A)P(B|A)}{P(B)}$.

A- מה הטמפרטורה האמיתית בחוץ כרגע

B- המדחום מדד 21 מעלות

לכל טמפרטורה שתנתן, נחשב מה הסיכוי שזו הטמפ' בחוץ כרגע אם המדחום מדד 21 מעלות.

ההסתברות ל-A זה הידע המוקדם.

ההסתברות ל-B בהנתן A זה נראות.

אם y המדידה של המדחום, x הטמפ' האמיתית, אז $y \sim N(x, \sigma^2)$.

$A \sim N(u_A, \sigma_A{}^2)$.

$$P(A|B) = \frac{1}{P(B)}$$

גאוסיאן- $\frac{1}{\sqrt{2\pi\sigma_A{}^2}} e^{-\frac{1}{2\sigma_A{}^2}(A-u_A)^2}$

אם $\sigma = 0$ אז המדחום מושלם.

טענה: אם $A \sim N(u_A, \sigma_A{}^2)$, $y|A \sim N(A, \sigma^2)$ אז $A|y \sim N(u, \sigma_2{}^2)$

$u = \frac{\frac{1}{\sigma^2}y + \frac{1}{\sigma_A{}^2}u_A}{\frac{1}{\sigma^2}+\frac{1}{\sigma_A{}^2}}$ $\sigma_2{}^2 = \frac{1}{\frac{1}{\sigma^2}+\frac{1}{\sigma_A{}^2}}$.

הערה: u הוא ממוצע משוקלל של (1) המדידה במדחום (2) התוחלת של הידע המוקדם.

כאשר $\sigma \to 0$ נקבל כי $u \to y$ (המדחום מדוייק, נסתמך עליו)

כאשר $\sigma \to \infty$ נקבל כי $u \to u_A$ (המדחום כמעט אקראי, נסתמך על הידע המוקדם)

כאשר $\sigma_A \to 0$ נקבל כי $u \to u_A$ (כאשר יש יום שהטמפ' בו היא תמיד קבועה, נסתמך עליה)

כאשר $\sigma_A \to \infty$ נקבל כי $u \to y$ (כאשר המידע המקדים כמעט אקראי, נסתמך על המדידה הנוכחית)

שאלה פילוסופית: המדחום מדד 21 מעלות, המשערך הבייזיאני ניחש 23 מעלות. הסתבר שהטמפרטורה האמיתית היא באמת 21 מעלות. האם השערוך הבייזיאני הוא 'אשליה'?

השערוך יהיה נכון בממוצע על פני הרבה ניסויים.

ה-λ שהזכרנו קודם שמוסיפים לאלכסון המטריצה M יכולה להקבע על ידי השונויות σ, σ_A האלו.

http://www.cs.huji.ac.il/~yweiss/Rhombus/rhombus.html

מעוינים זזים ימינה-שמאלה או באלכסון?

משתנה בעובי המעוינים וברמת הקונטרסט מול הרקע.

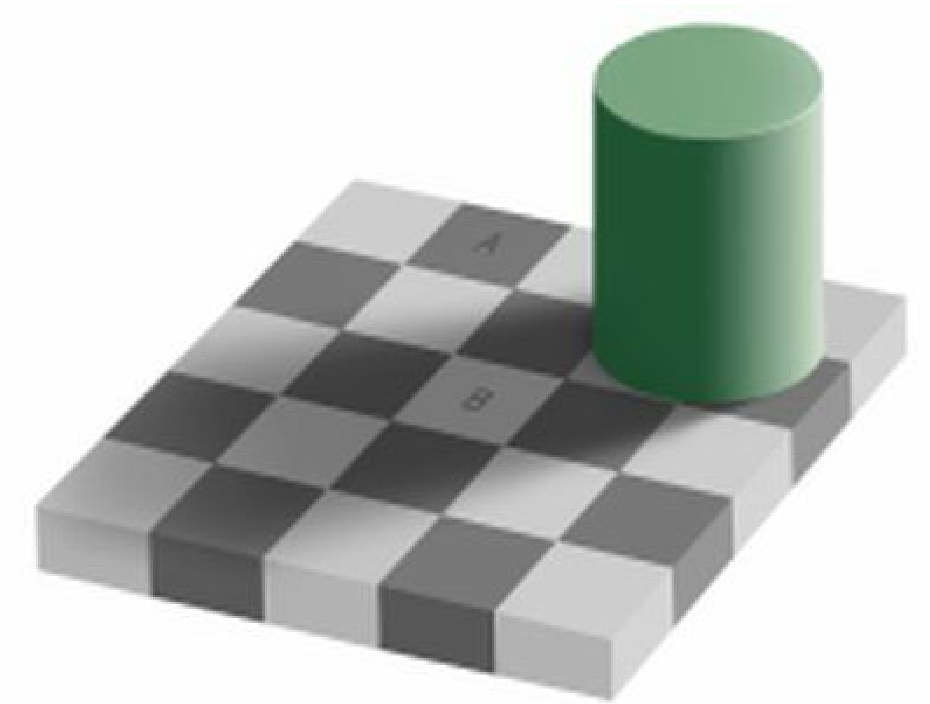
המשערך הבייזיאני ינחש שהמעוין הרזה נע באלכסון, ושהמעוין השמן נע אופקית.

וולאך- אנחנו מעדיפים מהירויות איטיות.

<u>שיעור 5, 27/04</u>

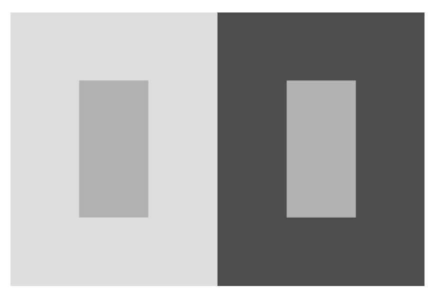
מתחילים את החלק השני, צבע. התחום מחולק לשני חלקים- שחור ולבן לעומת צבע.

יש הרבה אשליות במערכת הראייה, גם בגווני אפור, למשל הלוח המשובץ עם 2 משבצות שהן בדיוק באותו צבע אך אחת נראית בהירה והשנייה כהה ממנה.



גם כאן נרצה לבדוק מה הבעיה שמערכת הראיה מנסה לפתור, למה היא קשה, ואילו הנחות על העולם יכולות להקל אותה.

האשליה הבסיסית ביותר בהקשר רמות אפור נקראת simultaneous contrast,
הריבועים הפנימיים הם באותו צבע אבל זה שנמצא על רקע כהה נראה בהיר והשני נדמה כהה.
מכאן כי התפיסה של צבע האובייקט תלויה ברקע עליו הוא מופיע. זה נכון לא רק לגבי צבעים, אלא השפעה של הסביבה בשלל תחומים - גם מועמד בינוני ייתפס כחזק אם הוא מתראיין למשרה אחרי הרבה מאד מועמדים גרועים.



התופעה נקראת lateral inhibition, זהו הסבר שאינו נכון. זוהי תופעה במערכת הראייה, התאים שקולטים אור ברשתית הם הפוטורצפטורים, תאים שמגיבים תגובה חשמלית לפוטונים. הפלט שלהם מגיע לתאי center surround, וכל תא כזה מגיב להפרש בין כמות האור שיש במרכז לכמות האור שיש בסביבה.

ההסבר הוא שהתגובה החזקה של ה-center surround הוא זה שגורם לנו לראות את המרובעים בצבעים שונים.

עורמים מרובעים בגווני אפור שונים אחד על השני, נוצרת אשליה של צלב בהיר עליהם. את האשליה הזו אולי כן נתן להסביר על ידי התופעה הזו.
http://www.psy.ritsumei.ac.jp/~akitaoka/Vasarelyillusions.jpg
http://www.michaelbach.de/ot/lum-pyramid/index.html

לוקחים ריבוע שמתחיל בהיר והופך בקצב קבוע לכהה יותר. כששמים 2 כאלה אחד ליד השני הם נראים שונים, אחד כהה והשני בהיר. Craik-O'Brien-Cornsweet.
האשליה מתחזקת אם הופכים את הריבועים לקוביות תלת מימדיות, ונחלשת מעט אם הופכים את הריבועים לגלילים. מזה אפשר ללמוד שהתופעה מורכבת יותר מהקולטנים הרגישים להפרשים בתאורה.
http://www.math.tau.ac.il/~hezy/Vision%20Seminar/Lightness%20Perception%20and%20Lightness%20Illusions_files/Image4.jpg

הטבעת של Koffka- כשהטבעת שלמה היא נראית בצבע אחיד, כשמפרידים נראה שכל חצי בצבע שונה. מכאן ניתן ללמוד שלארגון המרחבי של התמונה יש השפעה מאד חזקה על התפיסה שלנו.
Gestalt (גשטאלט)- מצב שבו השלם הוא יותר מסכום חלקיו. אי אפשר להבין את הפרשנות של כל פיקסל בנפרד, אלא חייבים לבחון את המכלול של התמונה.
http://sc-optom.com/wp-content/uploads/2012/10/KoffkaRing-to-be-uploaded.jpg

corrugated plaid illusion- שני מלבנים באותה דרגת אפור, בין הרבה ריבועים אחרים.
כשמזיזים את צורת הרשת הצבעים נראים שונים, למרות שהצבעים שמקיפים אותם נשארים אותו דבר. זה קשור לתפיסה התלת מימדית, נראה שחלק אמור לקבל יותר תאורה או פחות.
http://www.math.tau.ac.il/~hezy/Vision%20Seminar/Lightness%20Perception%20and%20Lightness%20Illusions_files/Image10.gif

the haze illusion- הסביבה של מעויין יכולה לגרום לו להראות יותר מטושטש או חד.
http://i.ytimg.com/vi/3FebWnd2JWA/hqdefault.jpg

The snake illusion- המעויינים נראים בצבעים שונים על רקעים שונים. אולי תפיסה של אור וצל משפיעה עלינו.
http://media.log-in.ru/i/diamondsgrey.gif

מה הבעיה שמערכת הראייה מנסה לפתור?

מדוע הבעיה קשה?

אילו הנחות ניתן להניח על העולם הופכות את הבעיה לקלה יותר?

לא מעניין אותנו כבני אדם איזה פיקסל בהיר יותר מאחר, אלא איזה אובייקט בהיר יותר מאחר.
למשל אם מוטל צל על מדרכה, לא מעניין אותנו שהפיקסל שם כהה יותר, אלא שמדובר בצל ובסך הכל המדרכה כולה היא באותו צבע, רק בתנאי תאורה שונים.

אנחנו $I(x,y)$ מסתכלים על אובייקט.

חוק למברט- $I = \rho \cdot L \cdot cos(\theta)$

ρ - albedo, מקדם ההחזרה, בין 0-1 (בהיר וכהה).

אובייקט בהיר כמו דף לבן, מקדם ההחזרה שלו הוא כמעט 1. נניח 0.8, זה אומר ש-80% מהאור שמגיע אליו גם מוחזר. אובייקט כהה כמו חולצה שחורה, נניחה שמקדם ההחזרה שלו הוא 0.2, אז רק 20% מהאור שמגיע אליו מוחזר.

L- כמות האור שמגיעה לאובייקט.

θ - כיוון האור, הזווית בין הנורמל למשטח לכיוון האור.

גופים למבטריאנים- האור מוחזר בצורה שווה לכל הכיוונים. מראה אינה גוף למברטיאני, כמות האור שתוחזר תלויה בזווית.

האשליה של השחמט-

$0.42 = \rho_A L_A$

$0.42 = \rho_B L_B$

אנחנו יודעים ש- $L_A > L_B$ ולכן מניחים ש- $\rho_A < \rho_B$

<u>שיעור 6, 04/05</u>

איך העין יודעת מתי שינוי בצבע הוא כתוצאה משינוי בתאורה ומתי הוא כתוצאה משינוי במקדם ההחזרה?

אלגוריתמים intrinsic images.

$I(x,y) = \rho(x,y) \cdot L(x,y) \cdot cos\theta(x,y)$

$I(x,y) = R(x,y) \cdot L(x,y)$

בכל פיקסל עוצמת האור שמגיעה לעין היא מכפלה של מקדם ההחזרה (R) וכמות האור האפקטיבית המגיעה (L, illumination), שינוי בתאורה כתוצאה מהצללה וגם כתוצאה משינוי בזווית).

<u>קלט</u>: $I(x,y)$ N משוואות

<u>פלט</u>: $R(x,y),\ L(x,y)$ 2N נעלמים

<u>פתרון טריוויאלי 1:</u>

$L(x,y)=1$ **לכל** x,y

$I(x,y)=R(x,y)$

דוגמאות: ציור שמן, ציור על הלוח.

<u>פתרון טריוואלי 2:</u>

$R(x,y)=1$ **לכל** x,y

$I(x,y)=L(x,y)$

דוגמאות: הקרנה של סרט על מסך אחיד.

pragnanz- פשטות בגרמנית.

איזה הנחות ניתן להניח כך שהבעיה תהיה יותר קלה?

<u>אלגוריתם Retinex</u>

שנות ה-70, Land & McCann.

לנד היה הממציא של מצלמת הפולארויד. פתרון הבעיה הזו הוא חשוב לפיתוח תמונות. יש הפרשים עצומים בכמות האור שמגיעה, צריך לכווץ את הטווח הדינאמי הגדול הזה לטווח דינאמי קטן שמאפשר הדפסה של התמונה. הפיתוח מנסה לחלק את התמונה ל-R ול-L ולהפעיל כיווצים שונים על כל אחד מהם.

R נע בין 0.2 ל-0.8, השינויים המהותיים יותר הם ב-L.

הנחות:

1. כל שינוי בתמונה הוא או שינוי בתאורה או שינוי במקדם ההחזרה, אבל לא שניהם.
2. שינוי הדרגתי הוא שינוי בתאורה, שינוי חד ולא הדרגתי הוא שינוי במקדם ההחזרה.

במימד אחד:

יש את הפונקציה $I(x)$.

נקח נגזרת של הפונקציה של $I(x)$, נבחר איזשהו סף T.

נפצל לערכים גדולים מהסף (אלה יהיו ה-R) וקטנים מהסף (יהיו ה-L).

נעשה ל-2 הפונקציות שקיבלנו אינטגרציה.

בדו-מימד:

אותו דבר רק עם נגזרת לפי x ולפי y ואז העלתם בריבוע וחיבור.

אם הנורמה קטנה מ-T זה תאורה, אחרת מקדם החזרה.

ניצור את הפונקציה של R,L עבור x,y בנפרד (כלומר מאחדים את ערכי ה-x וה-y לצורך החישוב וההחלטה אם הם עוברים סף, אבל אחר כך משתמשים בהם בנפרד, ומכניסים אותה לפונקציה של R או של L בהתאם).

יש R x, R y, L x, L y.

בעיית האינטגרציה בדו-מימד-

קלט- $\frac{dz}{dx}, \frac{dz}{dy}$

פלט- z

פתרון אפשרי- נחפש פונקציה $\hat{z}$ כך ש- $(\frac{d\hat{z}}{dx} - \frac{dz}{dx})^2 + (\frac{d\hat{z}}{dy} + \frac{dz}{dy})^2$ מינימלי (סכום על כל התמונה).

<u>שיפורים</u>:

color retinex

התפיסה היא ששינוי תאורה לא גורמים לשינוי בגוון אלא רק בעוצמה. זה לא מדויק אבל נכון חלקית.

החלטות פחות לוקאליות.

שימוש בלמידה חישובית.

אלה לא שיפורים משמעותיים.

באיזה מידה מערכת הראיה האנושית משתמשת ב-retinex?

אם רוצים לצייר צל שיראה אמין צריך שיהיה מעבר הדרגתי, כך אנשים נוטים להאמין שזה צל אמיתי.

<u>שיעור 7, 11/05</u>

אפקט גלב (Gelb effect)- עם אור חזק של מקרן, ריבוע שחור היה נראה כלבן. כשהצבנו לידו ריבוע בהיר יותר, הריבוע השחור נדמה היה מעט כהה יותר אך עדיין מאד בהיר. הוספנו עד ריבוע ועוד אחד, רק אז נדמה היה שהריבועים הראשונים כהים יותר. כשכיבינו את אור המקרן הסתבר שהריבוע הראשון, שהיה נדמה כלבן בהתחלה, הוא ממש שחור.

היה ברור לנו שהשינוי בצבעים בין הדפים הוא כתוצאה ממקדם ההחזרה ולא מהתאורה, כלומר היה L גלובלי.

$I = \rho \cdot L \cdot cos\theta$.

Retinex מחשב את $log(x,y)$ עד כדי קבוע אינטגרציה, $L(x,y)$ עד כדי קבוע כפלי.

הבעיה שמערכת הראייה מנסה לפתור היא $I(x,y) = R(x,y) \cdot L$, למצוא לכל פיקסל את מקדם ההחזרה ואת L הגלובלי.

הבעיה קשה כי אם יש n פיקסלים, יש n משוואות ו-n+1 נעלמים.

זו בעיה שקיימת גם בצילום- כשמפתחים תמונות ורוצים להדפיס אותן, ניתן לפתח בכל מיני צורות על פי מה שאנחנו חושבים שהוא מקדם ההחזרה. זוהי בעיית העיגון (anchoring/white balance).

נניח יש תמונה עם חמישה פיקסלים, $I(x,y) = (100, 125, 125, 125, 150)$.

ניתן לפתור את המשוואה על ידי $= 255 \cdot (\frac{100}{255}, \frac{125}{255}, ..., \frac{150}{255})$

וגם על ידי $= 150 \cdot (\frac{100}{150}, \frac{125}{150}, ..., \frac{150}{250})$

<u>הנחה 1</u> (Gray World Assumption): ה-reflectance הממוצע בתמונה הוא אפור.

$$\sum \frac{\frac{I(x,y)}{L}}{N} = \frac{\sum I(x,y)}{L \cdot N} = 0.5$$

$$\Rightarrow L = 2\frac{\sum I(x,y)}{N}$$

תחת ההנחה הזו, בדוגמה שלנו $L = 250$ (מחשבים את הממוצע של I ומכפילים ב-2).

<u>הנחה 2</u> (Brightest White): הפיקסל הבהיר ביותר בתמונה הוא לבן ולכן מקדם ההחזרה 1.

$$max\frac{I(x,y)}{L} = 1$$

$$\Rightarrow L = max\ I(x,y)$$

תחת ההנחה הזו, בדוגמה שלנו $L = 150$.

באיזה הנחה מערכת הראייה משתמשת?

נדמה שבהנחה 2, ממה שרואים קודם ב-Gelb effect.

Articulation- הכלל של "הבהיר הוא לבן" עובד יותר טוב ככל שיש יותר פיקסלים שונים בתמונה.

נתחיל לדבר על צבע ממש.

נדבר על מודל מקורב ומופשט של פיזיקה של צבע ותפיסת צבע.

יש לנו 3 ערוצים R,G,B. זה לא מדוייק אבל מקורב.

חוק למברט- בכל נקודה יש 3 ערוצים $I_R(x,y), I_G(x,y), I_B(x,y)$.

$$I_R(x,y) = \rho_R(x,y) \cdot L_R(x,y) \cdot cos\theta$$

$$I_G(x,y) = \rho_G(x,y) \cdot L_G(x,y) \cdot cos\theta$$

$$I_B(x,y) = \rho_B(x,y) \cdot L_B(x,y) \cdot cos\theta$$

בעיית העיגון בצבע

$$I_R(x,y) = \rho_R(x,y) \cdot L_R$$

$$I_G(x,y) = \rho_G(x,y) \cdot L_G$$

$$I_B(x,y) = \rho_B(x,y) \cdot L_B$$

יש לנו n פיקסלים, 3n משוואות ו-3n+3 נעלמים.

$\rho = (1, \varepsilon, \varepsilon)$ זה "אדום"

$\rho = (\varepsilon, 1, \varepsilon)$ זה "ירוק"

$\rho = (\varepsilon, \varepsilon, 1)$ זה "כחול"

$\rho = (\varepsilon, \varepsilon, \varepsilon)$ זה "שחור"

$\rho = (1, 1, 1)$ זה "לבן"

$\rho = (1, 1, \varepsilon)$ זה "צהוב"

נניח יש פיקסל עם (10,10,255)

אז או שהאור הוא לבן (1,1,1) ומקדמי ההחזרה הם (10/255,1/255,1)

או שהאור הוא (10,10,255) ומקדמי ההחזרה הם (1,1,1)

ויש אינסוף פתרונות אפשריים כאלו לכל נקודה.

אפשרות לפתרון היא להפעיל על כל ערוץ בנפרד את ההנחות שראינו מקודם.

המסך צהוב, $I = (255, 255, 10) = (255, 255, 255).*(1, 1, \varepsilon) = (255, 255, 10).*(1, 1, 1)$

המקרן מאיר אור צהוב, אם נאיר על דף לבן נוכל לטעות ולחשוב שהוא צהוב.

אם נאיר על דף מקושקש (קופסת סירייל) נטעה פחות, נחשוב שהלבן על הדף הוא לבן ולא צהוב.

מכנסיים שחורים תחת אור צהבצהב $(255, 255, 100) \cdot (0.5, 0.5, 0.5) = (128, 128, 50)$

חולצה כחולה תחת אור צהבהב $(255, 255, 10) \cdot (0.5, 0.5, 1) = (128, 128, 100)$

העין רואה (128,128,50) עבור המכנסיים ו-(128,128,100) עבור החולצה.

את התוצאות ניתן לפרש בשתי דרכים-

$$(128,128,50) = (128,128,128) \cdot (1,1,50/128) = (255,255,100) \cdot (0.5,0.5,0.5)$$

מכנס צהוב תחת אור לבן/ מכנס שחור תחת אור צהוב

$$(128,128,100) = (128,128,128) \cdot (1,1,100/128) = (255,255,100) \cdot (0.5,0.5,1)$$

חולצה לבנה תחת אור לבן / חולצה כחולה תחת אור צהוב

ראינו במטלב- מכפילים תמונה של פסים בשחור וכחול בתמונה צהובה (כמו להוסיף תאורה צהובה), ואז התמונה נראית בפסים של לבן וזהב.

תמונה של הדוגמנית בשמלה כחולה, מוסיפים גוון צהוב. על אף שהפיקסל רחוק מלהחשב 'כחול' מבחינת ערכי ה-R,G,B, כולם רואים את זה ככחול כי הם יודעים שיש תאורה צהובה.

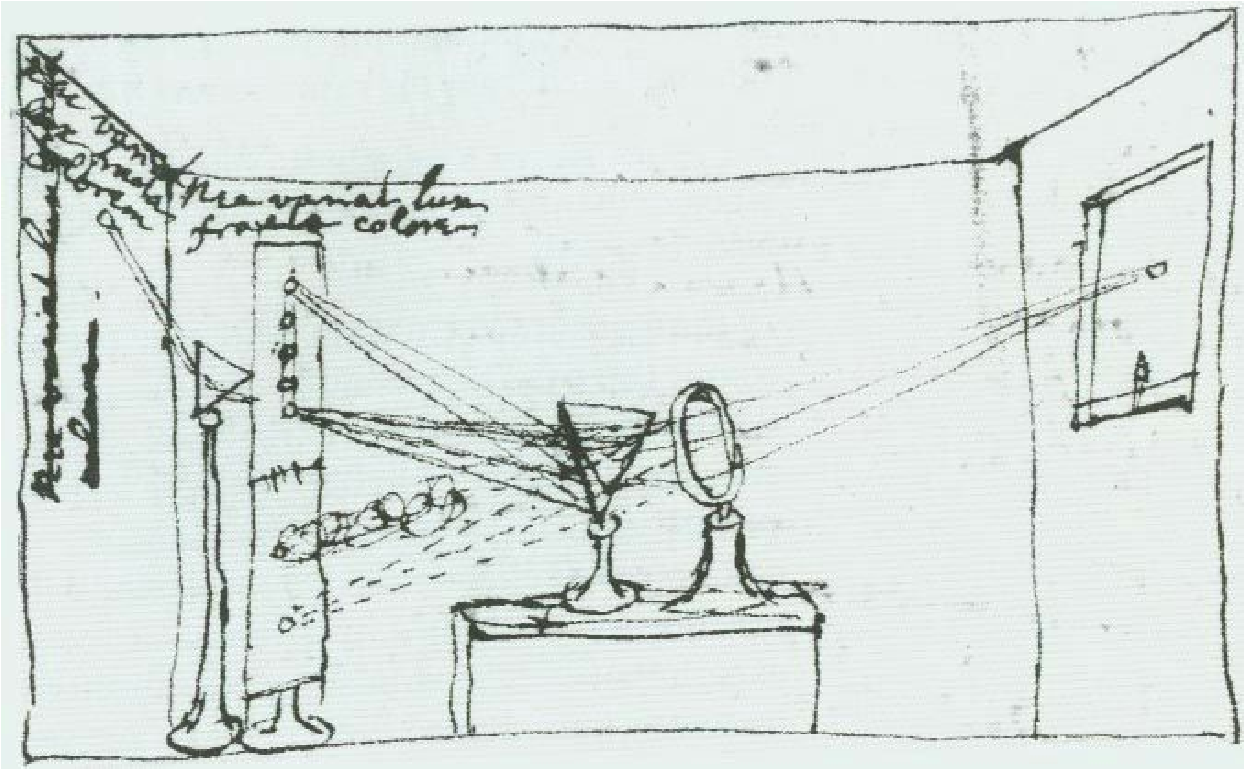
אם חותכים את התמונה ורואים רק את השמלה בלי הרקע, היא נראית בצבעים לבן וזהב. נמחק הקונטקסט אז אנחנו מפרשים את המשוואות בלי 'לדעת' שיש אור צהוב על התמונה.

<u>שיעור 8, 25/05</u>

תפיסת צבע ופיזיקה של צבע.

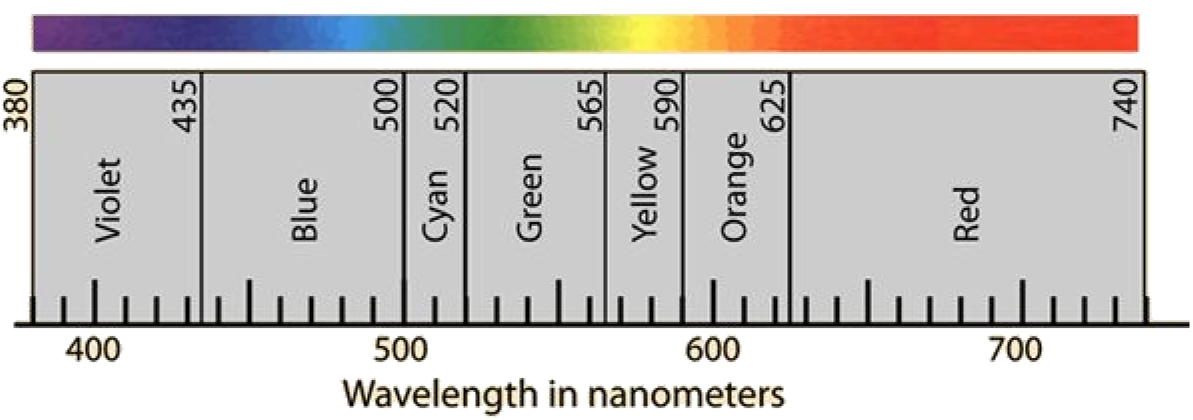
הניסוי המפורסם של ניוטון- אור שנראה בעין לבן מועבר דרך עדשה ואז דרך מנסרה, רואים על הקיר צבעים.

גם בכיוון ההפוך, מעבירים דרך חורים אור בשלל צבעי הקשת, מועבר דרך מנסרה ורואים אור לבן.



הדרך לתאר אור היא על ידי גרף שיש בו עוצמת אנרגיה כפונקציה של אורך גל.

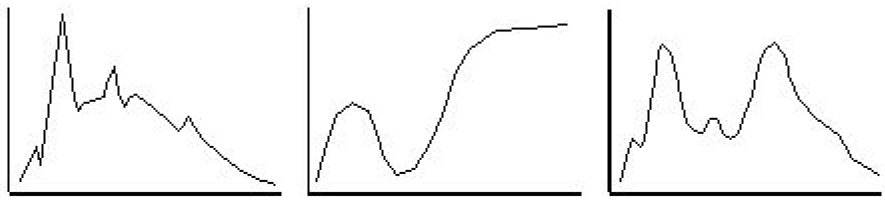
הספקטרום הנראה על ידי בני אדם הוא בערך 400-700 ננומטר. יש אינסוף צבעים שאפשר לתפוס בספקטרום הזה, זו פונקציה רציפה, אך אפשר לחלק לקבוצות של צבעים למשל אורכי הגל הארוכים הם אדומים והקצרים סגולים.



הדרך הנכונה, אם כך, לתאר את חוק למברט-

$I(\lambda) = \rho(\lambda)L(\lambda)cos\theta$ ובעצם אין 3 משוואות אלא אינסוף, לכל אורך גל.

זוהי בעצם מכפלה של פונקציות- פונקציה של התאורה שמתארת כמה תאורה יש בכל אורך גל ופונקציה של מקדם החזרה שמתאר כמה מגיע מכל אורך גל. המכפלה יוצרת פונקציה שהיא הסיגנל שמגיע אלינו לעין.



אפשר למדוד את עוצמת האור שמגיעה לעין עם מכשיר. העין לא עושה את זה.

שילוב של גלים שונים לא מניב את התוצאה שהיינו מצפים לפי מה שלמדנו בגן.

<u>תאורית ה-trichomacy של Young–Helmholtz</u>

אפשר לתאר את הסיגנל שמגיע לעין בתור וקטור באורך 400 (לכל אורך גל בין 300-700). אמנם זוהי קוונטיזציה אבל למעשה בני האדם וגם חיות קולטים מספר נמוך בהרבה של סוגי צבעים.

התאוריה היא שניתן לתאר את תפיסת הצבע של בני האדם על ידי הטלה לינארית בשלושה מימדים.

$I(\lambda)$ הוא הגירוי הפיזיקלי. לכל אורגניזם יש מספר סופי של קולטני צבע, אפשר לתאר על ידי אינטגרל.

$$I(\lambda) = \int I(\lambda)S_k(\lambda)d\lambda \approx \sum_{i=300}^{700} I(\lambda)S_k(\lambda)$$

ניסוי-

מראים לנבדק אור בצד שמאל של המסך, ובצד ימין יש 3 פנסים. הנבדק מתבקש לכוון את צבעי הפנסים עד ששני הצבעים בשני הצדדים יראו בדיוק אותו דבר. הנבדקים הצליחו לתאר כל צבע אור על ידי שילוב של 3 צבעים שונים בעוצמות כלשהן.

<u>חוקי גרסמן</u>

סיכום של הרבה ניסויים מסוג זה. הסימן '=' מציין שוויון תפיסתי, כלומר שנדמה לנו שהם זהים גם אם זה לא כך.

סימטריה- $U = V \Leftrightarrow V = U$, אור U זהה מבחינת התפיסה לאור V.

כנ"ל לגבי טרנזטיביות, פרופורציוניות, אדטיביות.

<u>השערה 1</u>- כל קולטן צבע מבצע מכפלה פנימית של האור l (וקטור בגודל 400,1) עם וקטור רגישות S_k (מטריצה בגודל 3,400). כלומר את התגובה של הקולטן ה-k ניתן לתאר על ידי $r_k = S_k^T l$, כאשר r הוא וקטור בגודל 3,1.

<u>השערה 2</u>- לבני אדם יש 3 קולטנים שונים $r_1 = S_1^T l, r_2 = S_2^T l, r_3 = S_3^T l$.

<u>השערה 3</u>- l_1 זהה תפיסתית ל-l_2 אם $Sl_1 = Sl_2$.

<u>טענה</u>- אם l_1, l_2 זהים תפיסתית אם ורק אם $Sl_1 = Sl_2$ אז חוקי גרסמן יתקיימו.

<u>הוכחה</u>-

$$Sl_2 = Sl_1 \Leftarrow Sl_1 = Sl_2$$

$$Sl_1 = Sl_3 \Leftarrow Sl_1 = Sl_2,\ Sl_2 = Sl_3$$

$$Sl_1 + Sl_3 = Sl_2 + Sl_4 \Leftarrow Sl_1 = Sl_2,\ Sl_3 = Sl_4$$

<u>הגדרה</u>- אור l_1 אור l_2 נקרים מטמרים (metamers) אם $l_1 \neq l_2$ אבל $Sl_1 = Sl_2$.

<u>טענה</u>- נניח ש-S היא הטלה לשלושה מימדים, אז כמעט לכל שלושה פנסים t_1, t_2, t_3 יהיה ניתן להתאים בדיוק צירוף של שלושתם לכל מקור אור l.

$$l = \alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3$$

$$l_{400*1} = P_{400*3}\alpha_{3*1}$$

$$S_{3*400} l_{400*1} = (S_{3*400} P_{400*3})\alpha_{3*1}$$

$$\alpha = (SP)^{-1} Sl$$

<u>הוכחה</u>- אם המטריצה SP הפיכה אז $\alpha = (SP)^{-1}Sl$ יקיים $Sl = SP\alpha$ ואז הצירוף של שלשת הפנסים יהיה שקול תפיסתית ל-l.

<u>הערה</u>- אם מטריצת ההטלה היא S או מטריצת ההטלה היא QS כאשר Q מטריצה בגודל 3,3 הפיכה, אז המטמרים של מערכת ראייה עם S שקולים למטמרים של מערכת ראייה עם QS.

<u>הוכחה</u>- $Sl_1 = Sl_2$ אז $(QS)l_1 = (QS)l_2$.

$$\alpha = (SP)^{-1} Sl = Cl$$

$$\alpha_{3*1} = C_{3*400} l_{400*1}$$

<u>הערה</u>- ניתן לחשב אקספרמנטלית את c על ידי רישום ה-α שנותנות שקילות ל-$l = (00..1..00)^T$ (וקטורי יחידה במרחב ה-400 ממדי).

$L = QS$ כאשר $Q = (SP)^{-1}$

<u>שיעור 9, 08/06</u>

מתחילים לדבר על צורה.

איך תופסים תלת-מימד מדו-מימד. ברור לנו שאנחנו חיים בעולם תלת-מימדי, אך כשאנחנו מסתכלים על העולם אנחנו מקבלים אינפוט דו-מימדי, אבל אנחנו מפרשים אותו כתלת-מימד.

structure from X- רמזים בדו-מימד על תלת-מימד.

היום נדבר על כיצד מחשבים תלת-מימד מרצף תמונות דו-מימדיות.

נקודות שחורות על רקע לבן, כתמונה בודדת הן נראות דו-מימדיות, אבל כשיש רצף של תמונות משתנות (אנימציה) נוצרת אשלית תלת-מימד. נראה שהנקודות יושבות על גליל שמסתובב, ואפשר לראות אותן כמסתובבות לשני הכיוונים.

http://cogsci.uci.edu/~ddhoff/cylinderapplet.html

מה הבעיה שמערכת הראיה מנסה לפתור?

תהי נקודה במרחב תלת-מימדי (x,y,z). אם נסובב אותה סביב ראשית הצירים, זה כמו לכפול במטריצה. (x',y',z') היא הנקודה החדשה אחרי הכפל. המטריצה היא $M = (cos\theta\ 0\ -sin\theta; 0\ 1\ 0; sin\theta\ 0\ cos\theta)$. זוהי מטריצה אורתוגונלית, מתקיים $M^T M = I$.

אפשר לייצר סרט של הנקודה הבודדת הזו כשכל פעם נשנה את ה θ במטריצה במעט.

$(x(t), y(t))$ שהם x בזמן t ו-y בזמן t הן הטלה (prog) של $(x(t), y(t), z(t))$ המתקבלות מהכפלת הנקודה במטריצה עם $\theta = t$.

הטלה אורתוגרפית- $prog(x\ y\ z) = (x\ y)$, פשוט מזניחים את הקואורדינטה השלישית.

קלט- $\{x_i(t), y_i(t)\}^T{}_{t=1}\}\quad i \in \{1..p\}$

פלט- $x_i(t),\ y_i(t),\ z_i(t)$

למה הבעיה קשה- אין לנו את z, חסרה לנו אינפורמציה. לכל נקודה יש 2 משוואות עם 3 נעלמים.

פתרון אפשרי- נניח שהנקודות נמצאות על גוף קשיח, אם קיים פתרון של גוף קשיח. Incremental Rigidity Assumption.

הנחה- $(x_i(t),\ y_i(t)) = prog(M(t)\ (x_i(o),\ y_i(o),\ z_i(o))$

כמה משוואות עכשיו יש? יש P נקודות, לכל אחת 3 קואורדינטות תלת-מימדיות, 3P נעלמים למבנה. בנוסף, יש T פריימים, לכל פריים יש מטריצה 3*3 סימטרית המוגדרת על ידי 2 זויות- 2T נעלמים לתנועה. סה"כ 2T+3P נעלמים, 2PT משוואות.

עבור P,T גדולים נקבל כי מספר המשוואות גדול בהרבה ממספר הנעלמים, כלומר אם קיים פתרון ריגידי אז ניתן למצוא אותו בקלות, למשל עבור P=100, T=60 נקבל כי 2T+3P=420, 2PT=12,000.

Tomasi-Kanade-

אלגוריתם המקבל כקלט נקודות בדו-מימד, מניח את הנחת הריגידיות ומחשב לכל נקודה קואורדינטות תלת-מימדיות ואת המטריצה M(t), בכמה סובב האובייקט בזמן t.

הטלה אורתוגרפית, זורקים את הקואורדינטה האחרונה אז אפשר להקטין במקום זה את M.

$$(x_i(t)\ \ y_i(t)) = M_{2*3}(t'\,(x_i(o)\ \ y_i(o)\ \ z_i(o)))$$

$$(x_1(t)\ y_1(t));\ x_2(t)\ y_2(t);\ ...;x_p(t)\ y_p(t)) = M_{2*3}(t'(x_1(o)\ x_2(o)\ ...\ x_p(o);\ y_1(o)\ y_2(o)\ ...y_p(o);z_1(o)\ z_2(o)...z_p(o)))$$

הדרך לייצר את הקואורודינטות הדו-מימדיות של כל הנקודת ב-2 פריימים:

$$(x_1(1)...x_p(1);\ y_1(1)...y_p(1);x_1(2)...x_p(2);\ y_1(2)...y_p(2)) = (M_{2*3}(1);M_{2*3}(2))\cdot(x_1(o)...x_p(o);y_1(o)...y_p(o);z_1(o)...z_p(o)$$

באותו אופן אפשר להמשיך עד ל-T פריימים ולא רק 2, נסמן את המטריצות $W_{2T*P} = M_{2T*3}S_{3*P}$.

S היא מטריצה המבנה, נותנת את המבנה התלת-מימדי של כל הנקודות.

M היא מטריצת התנועה, נותנת 2 שורות ממטריצת הסיבוב בכל פריים.

אנחנו יודעים את W, צריכים לחשב את S,M.

הערה- המטריצה W היא מדרגה 3 לכל היותר.

הוכחה- W היא מכפלת של מטריצה בגודל 2T*3 ומטריצה בגודל 3*P. הדרגה של מכפלה של מטריצות היא לכל היותר הדרגה של הרכיבים שלה, ולכן הדרגה היא לכל היותר 3.

הערה- SVD- singular value decomposition: כל מטריצה m*n מדרגה k ניתנת לכתיבה $W_{m*n} = U_{m*k}\Sigma_{k*k}V_{k*n}$ כאשר Σ מטריצה אלכסונית, U,V מטריצות אורתוגונליות.

ניתן לחשב את פירוק ה-SVD בזמן פולינומיאלי. במטלב- $[U,S,V] = SVD(W)$.

W שהזכרנו קודם היא מדרגה 3 ומכאן כי ניתן למצוא U,Σ,V כך ש- $W_{2T*P} = U_{2T*3}\Sigma_{3*3}V_{3*P}$.

פתרון אפשרי הוא ש- $U\Sigma = M, V = S$, פתרון נוסף $U = M, \Sigma V = S$.

$W = MS$ ולכן גם $W = (MA)(A^{-1}S)$ לכל מטריצה A_{3*3} הפיכה, כלומר אפשר לפרק את Σ בשלל דרכים ו'לחלק' אותה בין U,V ליצירת M,S.

כל שורה ב-M בעלת נורמה 1. כל 2 שורות עוקבות בעלות מכפלה פנימית 0.

האלגוריתם:

בהנתן W מחשבים את U,Σ,V . כותבים את W בתור $W=\tilde{M}\cdot\tilde{S}$, למשל $\tilde{M}=U,\tilde{S}=\Sigma U$.

מחפשים מטריצה A בגודל 3*3 הפיכה כך ש- $M=\tilde{M}\cdot A,\ S=A^{-1}\cdot\tilde{S}$ וכך ששורות ב-M הן בעלות נורמה 1 ושורות עוקבות הן עם מכפלה פנימית 0.

נסמן ב- M_1 את השורה הראשונה של M. נרצה ש- $M_1\cdot M_1{}^T=1$. מתקיים כי $M_1=\tilde{M}_1\cdot A$, $M_1{}^T=A^T(\tilde{M}_1{}^T)$.

נסמן $AA^T=Q$.

$\tilde{M}_1 Q M_1{}^{\tilde{T}}=1$.

$\tilde{M}_1 Q \tilde{M_2}=0$.

כלומר ניתן לבטא את כל האילוצים האלו על ידי משוואות. $\tilde{M}_1 Q \tilde{M}_1{}^{T}=\sum_{i,j}Q(i,j)\tilde{M}_1(i)\tilde{M}_i(j)=1$.

<u>שיעור 10, 15/06</u>

נדבר על חישוב תלת-מימד מתמונה בודדת.

קלט- x,y

פלט- { Xi Yi Zi} R_{3*3} .

הנחה- $(x_i\, y_i)=proj(R(X_i\ Y_i\, Z_i))$.

אנחנו רואים סטריאו אבל זה כנראה לא מה שעוזר לנו לחשב תלת-מימד כי גם אנשים שלא רואים סטריאו (בגלל עין עצלה למשל) רואים תלת-מימד.

גם כשאנחנו רואים ציור או תצלום אנחנו מפענחים דו-מימד כתלת-מימד.

זו בעיה קשה לחשב 3 קואורדינטות לנקודה בעלת 2 קואורדינטות.

Ames Room- נדמה כאילו האיש הולך וגדל. החדר הוא טרפזי, האשלייה גורמת לנו לחשוב שהחדר 'רגיל' והאיש גדל, כשלמעשה החדר גדול בפינה אחת וקטן באחרת.

בסביבות המאה ה-15-16 ציורים נהיו מאד אמינים, מצליחים לתאר היטב סצנה תלת-מימדית. הכלים לצייר תמונה אמינה כזו הם-

1. קווים אורתוגונלים- הולכים ומתרחקים מהצופה, נפגשים בנקודת המגוז.
2. נקודת המגוז (vanishing point)- הנקודה בה נפגשים הקווים, על האופק.
3. אופק- הקו בו השמים והארץ נפגשים. חשוב לצייר אותו בתמונה.

למה החוקים האלה נובעים ממתמטיקה ובאיזה תנאים הם באמת נכונים?

ההנחה היא שההטלה האורתוגרפית עליה דיברנו בשבוע שעבר היא לא נכונה.
הנחת מנהטן- קווים מקבילים בעולם הם או בכיוון X או בכיוון Y או בכיוון Z. תחת הנחה זו יהיו 3 נקודות מגוז לכל היותר וזה תלוי בכיוון המצלמה.

שיעור 11, 22/06

בהמשך לחישוב תלת-מימד מתמונה אחת-
הקשר בין מיקום נקודת המגוז למיקום המצלמת ביחס למערכת הקואורדינטות.

מישור בעל Y קבוע- למשל מישור התקרה בחדר. נניח $Y = 2$, 2 מטר, התקרה. $Y = -1$ מישור הרצפה.
ההטלה של מישור כזה בלתי תלויה ב-Y. כל המישורים בעולם שיש להם Y קבוע הולכים לאותו קו בתמונה ויפגשו שם.
קו האופק- הקו בתמונה בו נפגשים כל המישורים שעבורם Y קבוע.
כאשר המצלמה מסתובבת סביב ציר x בזווית θ, קו האופק הוא הקו $y = -tan\theta$, ונקודת המגוז תמיד נמצאת על האופק.

הנחת מישור האדמה- בסצנה קיים מישור שמקיים כי Y הוא קבוע, זה מישור האדמה.
בחדר המעוות יש אשליה שהרצפה היא מישור אדמה עם Y קבוע, אך למעשה זה לא נכון.

מר- ניתן להסביר תופעות ב-3 רמות: הרמה החישוביות, רמת האלגוריתם ורמת המימוש.
ברמה החישובית השאלות הרלוונטיות הן מה הבעיה שמערכת הראיה מנסה לפתור, מדוע היא קשה ואילו הנחות על העולם עוזרות לפתור את הבעיה.

אשליות מהאתר illusion of the year.

השמלה.
הבעיה- מה הצבע.
למה זה קשה- המון משוואות ומעט נעלמים.
הנחות- יש תאורה אחת על כל התמונה.

גגון של מכונית.
הבעיה- מבנה תלת מימדי מתמונה בודדת.
למה זה קשה- 2 משוואות ב-3 נעלמים לכל נקודה.

הנחות- מנהטן. מניחים שיש מישור של קצה הגג.

גשם.

הבעיה- חישוב תנועה.

למה זה קשה- aperture problem.

הנחות- הנחת שימור הבהירות.

הצבעים.

הבעיה- חישוב תנועה, חישוב צבע.

למה זה קשה- חישוב תאורה לעומת מקדם החזרה, aperture problem, מחזוריות.

הנחות- תמונה היא תחת תאורה אחת.

מדרונות.

הבעיה- חישוב תלת-מימד מדו-מימד.

למה זה קשה- דיברנו.

הנחות- הנחת מנהטן, תנועה צפידה. רואים את מישור הרצפה ואת העמודים