

פברואר 2007

פרויקט בקורס: מבוא לראייה חישובית וביולוגית

ז י ה ו י ת ו ו י ם מוסיקליים



מגיש: אייל רונאל, 21655550

המרצה: ד"ר אוהד בן-שחר

תוכן העניינים

3.....	הרעיון
4.....	השיטות ותוצאותיהן
4.....	זיהוי מוסיקה
4.....	ניפוי רעשים
4.....	זיהוי ראשי תווים
5.....	זיהוי גבעולי תווים
6.....	זיהוי זנבות
6.....	יצירת MIDI
6.....	סוגיות הנדסיות בפיתוח
7.....	סוגיות להתעמקות נוספת
7.....	אישוש מוסיקלי
7.....	יישור דפים עקומים
7.....	ייצוג אוניברסלי של סימנים מוסיקליים
8.....	מסקנות

הרעיון

ראייה חיושית עוסקת בניסיונות לגשר על הפער העמוק שבין היכולת החיושית הפנטסטית של מערכת הראייה שלנו לבין היכולות של המחשב, אשר הן אינן מוגבלות ביסודן, אבל רחוקות עדיין במידה ניכרת מאלה שלנו. במהלך הקורס ראינו שיטות רבות ומגוונות לחקוק את הראייה שלנו באמצעים מכניים ובאמצעות היריסטיקות חיושיות, והתוצאות היו מניחות את הדעת במקרים הטובים והנוחים, ולא קשורות למציאות במקרים הפחות נוחים.

כשהידע הזה באמתחתי, ניגשתי ליישם אותו על תחום כלשהו לבחירת, שיהיו בו אלמנטים מהחומר שלמדנו וחדשנות מסוימת. מצד אחד לא רציתי נושא תיאורטי ענף, שבו אתיימר להתחרות במדענים ידועי שם ולעמת תיאוריה משלי מול הממצאים שלהם. לכן רציתי לעסוק בתחום שנוגע ליישום של זיהוי ויזואלי מצד אחד, ושנעשה בעולם צר של מושגים כדי שיהיו לי הנחות בסיס מוצקות להישען עליהן. זיהוי תווים (OCR) אם כן היה רעיון טוב, אבל חברות מסחריות כבר פיתחו את הרעיון ברמה מתקדמת, ולכן לא היה כאן חידוש.

מבין עיסוקיי אני חוטא קצת במוסיקה, ולכן עלה בדעתי הרעיון של זיהוי תווים מוסיקליים. האלמנטים דומים לזיהוי תווים רגיל מבחינת הסכיבה שבה הם מצויים (סימנים שחורים על-גבי רקע לבן), ועם זאת שונים מספיק כדי להוות תחום חדש. אני מכיר מספר דרכים שבהם מתבצעת רכישה של מוסיקה לתוך המחשב. דרך אחת היא להשתמש בכלי מוסיקלי דוגמת אורגנית או גיטרה שמחוברת למחשב ומזינה תוכנה מתאימה, דרך אחרת היא באמצעות מיקרופון, ואפשר גם להקליד ידנית את המוסיקה באמצעות גרירה של סמלים מוסיקליים בעזרת העכבר, או באמצעות שפת תכנות ייעודית (על כך בהמשך). התוכנות שמייצרות מוסיקה ממוחשבת באמצעים אלה לעתים קרובות מפיקות דפי תווים. אבל עוד לא נתקלתי בתוכנה שמפענחת דפי תווים לכדי מוסיקה. ובתחום הזה רציתי ללכלך קצת את הידיים. חשבתי לחלק את הפרויקט לשני שלבים, שמתוכם בשלב הראשון אנסה לפענח דפי תווים נקיים שנוצרו בעזרת תוכנה ייעודית שכזו. בדרך זו אני מבטיח שהקווים ישרים, שהסימנים הם אחידים וידועים מראש. משימה זו התבררה כקשה מספיק וגולה את מירב הזמן מבלי שהתוצאות שהניבה יצדיקו מעבר לשלב הבא.

הפרויקט שלי מציג גישה בסיסית להתמודדות עם סוגיות של זיהוי תווים. השלד שיצרתי מותאם להרחבה ולפיתוח ע"י הצגה של קונספטים חדשים ומימוש שיטות קיימות באמצעים חדשים לפי רעיונות של תכנות מונחה עצמים כמו ירושה ומימוש מנשקים.

השיטות ותוצאותיהן

זיהוי מוסיקה

את המסע לפענוח התווים התחלתי בתקיפת הסביבה שלהם. ידעתי שבכל דרך שלא אבחר לזהות את הצורות המעוגלות שמרכיבות את הראשים שלהם, הדרך בה אוכל לעמוד על גובהם של העיגולים יהיה כרוך בקווים שמקיפים אותם. תו נמדד לפי הצורה שלו (הראש מלא או ריק, עם תורן או בלי תורן, מספר זנבות, מספר נקודות, נסקים ופחתים) לא פחות משהוא נמדד לפי המיקום שלו על גבי החמשה. קווים ישרים נראו כמו משימה פשוטה יותר ולכן התחלתי ממנה. הזיהוי היה מוצלח למדי, ולכן יכולתי להשתמש בו מאוחר יותר לניקוי קווי החמשה מתוך התמונה ובכך לשפר את הזיהוי של האלמנטים שמצויים על גביהם.

ניפוי רעשים

התחלתי מהדיריסיטיקה שהוכיחה את עצמה כיעילה: בשורת נקודות (pixels) בתמונה שמכילה קו חמשה, יותר מ-80% מהנקודות הן כהות. כדי לזהות כהות של נקודה, השתמשתי בייצוג ה-HSB שלה והשוויתי את רכיב הבהירות (Brightness) ל-0.8. זה נראה כמו ערך בהיר למדי, אבל הסתבר שלעתים קרובות, בייחוד עבור דפים מוקטנים, שורות של נקודות כהות הפכו לשורות כפולות של נקודות בהירות יותר. כך גם הגעתי למושג העובי של קו. נקודת ההתחלה של הקו היתה בקו העליון שלו, אולם לעתים קרובות הקו נמשך מעבר לשורה אחת, ואסור היה לי כמובן להחשיב זאת כשני קווי חמשה, שאז הייתי מקבל יותר מדי קווים. לאחר קיבוץ של עוביים לכדי קווים וקיבוץ של קווים לכדי חמשות, יכולתי לחלק את העניין שלי לפי חמשות במקום בצורה גלובאלית על פני כל התמונה. השלב הבא היה לנפות את קווי החמשה, ברגע שהמידע אודותיהם (מיקום, עובי) היה מצוי ברשותי. השתמשתי בהדיריסיטיקה הבאה: אם אוסף נקודות אנכי נמצא לאורך העובי של קו חמשה, ומעל ומתחת לאוסף זה ישנן נקודות שאינן כהות – אזי ניתן לנפות את הנקודות הללו מתוך התמונה. הרעיון נבע מכך שאם אלמנט היה בדיוק בעובי של הקו עצמו, גם מוסיקאי אנושי לא היה יכול להבחין בו, ולכן הוא לא יכול להתקיים. הניפוי היה יעיל למדי. כמעט ולא היו מקרים שבהם הוסרה נקודה בשוגג. יוצאים מכלל זה היו התעללויות שעבר מפתח סול בחלקים הדקים יותר שלו, שחלק חפפו קווי חמשה, והחלק העליון של הסימן C שמסמל ארבעה רבעים בקצב מהיר. עם זאת, הניפוי לא נפטר ממספיק נקודות ששייכות בעליל לקווי חמשה. כך נותרתי עם תרני תווים שמהם מבצצות נקודות של קווי חמשה שיכולות להתפרש בטעות כזנבות, וקווי תיבה שמהם מבצצים נקודות של קווי חמשה שיכולות להתפרש כהתחלה של ראש-תו. עדיין הניפוי היה טוב יחסית, והשתמשתי בו כדי להמשיך במלאכת הפענוח.

בהמשך לזיהוי האופקי של קווים, חשבתי שיהיה נוח לעשות את אותו הדבר בכיוון האנכי. קל לזהות קווי תיבה משום שהם באורך של חמשה. בהסתמך על עובדה זו בלבד איכות הזיהוי לא היתה טובה במיוחד, כי אקורדים למשל נוטים להיות תווים ארוכים למדי, וכך גם תווים מסוימים זוהו בתור קווי תיבה. הוספתי עוד תנאי שלפיו אם קו הוא באורך של חמשה והטורים שמשמאלו ומימינו הן ריקות, אזי זהו קו תיבה. באופן זה נפטרתי מהזיהוי השגוי של תווים כקווי תיבה, אבל היו קווי תיבה שלא זוהו. היו לכך כמה סיבות: (1) בכתוב צפוף מדי תווים, במיוחד עם זנבות, קרובים מאוד לקו התיבה. (2) קשתות שמחברות בין תווים בתיבות שונות חוצות את קו התיבה ומתפרשות כטורים לא ריקים. (3) ניפוי לא מספק של קווי החמשה מותירים שאריות של נקודות משמאל ומימין לקו התיבה. כאשר מציאים את המקרים האלה, הזיהוי היה טוב למדי. בדיעבד לא נזקקתי לזיהוי כזה של קווי תיבה כאשר עברתי לשיטה מונחית ראשי-תווים לזיהוי תווים, אבל ניתן עדיין להשתמש בה ביישום עתידי של חישוב מספר הפעמות בתיבה.

זיהוי ראשי תווים

בהתחלה ניסיתי לנקוט בגישה כללית עד כמה שניתן. למצוא מאפיינים כלליים ביותר שמבחינים בין תווים לבין אלמנטים אחרים, ואז למצוא את המאפיינים הכלליים ביותר שמבחינים בין סוגים שונים של תווים. לפי גישה זו, כל אלמנט הוא גוף כהה עד שזהותו נתבררה. ברגע שקווי החמשה גפו ניתן לאסוף בקלות יחסית גופים כהים כאלה בדומה לאיסוף קווי תיבה. לכל גוף כהה יש רוחב, גובה ומיקום. בשיטת האיסוף הזאת השתמשתי בכל הגישות שניסיתי.

בגישה הראשונה ניסיתי לסרוק בכל פעם שורה אחת של החמשה ולמצוא על גביה צורות עגולות. לקביעת הגובה של שורה כזאת מדדתי את המרחק בין כל זוג קווי חמשה סמוכים והשתמשתי בממוצע ביניהם. שמרתי עבור הגוף הכהה מידע האם הוא מורכב מבהירות אחידה (ראש תו מלא) או לא אחידה (ראש תו ריק), וסרקתי לאורך השורה כדי לזהות את הגופים, ואז מדדתי את האחידות. השיטה לא היתה יעילה בעיקר משום שהשורות חופפות. אם שורה 1 נמצאת על גבי קו החמשה התחתון ושורה 2 נמצאת בין קו החמשה התחתון לסמוך לו למעלה, התווים "מי" ו"פה" שנמצאים בשורות האלה בהתאמה יופיעו באופן ניכר בשתי השורות הנסרקות. גם מזהה צורות עגולות מתוחכם יחסית יתקשה כנראה לזהות שצורה אחת חתוכה והשנייה לא. גם הגישה הנאיבית לזיהוי אחידות לא עבדה, ובכל המקרים הוכרזה אחידות לא זהה. לכן הסקתי שסריקה לפי שורות וזיהוי ראשי תווים בתוכן היא לא גישה טובה, וחיפשתי מוצא אחר.

הגישה השנייה היתה יותר מוצלחת, לדעתי. לאחר ניפוי של קווי חמשה אופקיים וניפוי חלקי של קווי תיבה, סרקתי עצמים כהים לאורך כל החמשה (לא רק לאורך שורה), וביצעתי סיווג ראשוני פשוט: אם העצם גבוה מגובה חמשה, הוא כנראה מפתח. בהבחנה בין סוגי מפתחות החלטתי לא לטפל. אחרת ניגשתי לטפל בתווים.

חשבתי על גישה גנרית נוספת לטיפול בתווים, אבל היא נפלה בשלב הרעיון. כאשר תווים אינם מחוברים ביניהם בקו או בקשת אלא עומדים בפני עצמם, ניתן לקבוע שהשורה הרחבה ביותר של נקודות כהות בעצם הכהה שייכת לראש התו, אם למעלה ואם למטה. הבעיה החמורה היא איך לזהות שהשורה הרחבה ביותר לא נועדה לקשר בין תווים שונים. היה עלי לזהות את מספר הקווים האנכיים שקיימים באובייקט או את הקווים האופקיים, בעקמומיות ובזוויות שונות. במקרה הזה הקווים מחוברים בוודאות לקווים אחרים ויהיה קשה לבדוד אותם. לפני שניגשתי לחשוב על היישום המעשי של זה, החלטתי לנסות גישה פחות מורכבת.

הגופן שבו כותבים תווים דומה למדי ברוב הדוגמאות שראיתי. יוצאים ברורים מכלל זה הם תווים שנכתבים ביד חופשית, אבל מראש בחרתי לדחות את העיסוק עם דפים כאלה לגרסה עתידית של המערכת. לכן, אילו היה לי מילון שלפיו יכולתי להשוות ראשי תווים, הזיהוי היה הרבה יותר ישיר ופחות מועד לשגיאות.

יצרתי מילון כזה, שמורכב מתמונות קטנות שנקודותיהן בשלושה צבעים: שחור, לבן ואדום. עבור נקודה שחורה בודקים האם בדף התווים הנקודה כהה, עבור נקודה לבנה מוודאים שהנקודה אינה כהה, ועבור נקודה אדומה לא בודקים, כי היא משולה לרקע. כעת הזיהוי התבצע באופן הבא: ראשית סרקתי כמקודם את החמשה לזיהוי עצמים כהים. עבור כל עצם כהה השויתי את התמונות במילון כנגד כל אחת מהנקודות של העצם הכהה. אם היה זיהוי, הוחזרה הקואורדינטה של הנקודה שהשוותה. ביצעתי איסוף של קואורדינטות כאלה, למקרה שהעצם הכהה מורכב מכמה תווים המחברים ביניהם.

הגרסה להגשה מבוססת על הגישה הזאת. הדיוק שלה לא מרשים, אבל אפשר לשפר אותו עם החלפה של התמונות במילון. השיטה הזאת טובה גם משום שהיא מאפשרת לזהות גם עצמים מצורות שונות שאינן עגולות, כמו סימני הרמה והנחתה של חצי טון וסימני הפסקה.

זיהוי הגובה של התו נעשה בעזרת קווי החמשה שנאספו מראש. הקואורדינטה שהוחזרה נמדדת כנגד קווי החמשה, ומאחר שהקואורדינטה מייצגת את הקצה העליון של התו, אם היא עצמה נמצאת בקרבת קו חמשה, התו כולו נמצא בין הקווים, ואם היא נמצאת בין קווי חמשה, הרי שהתו עצמו נח על גבי קו החמשה שמתחתיה. עבור ראשי תו שנמדדו מתחת או מעל לחמשה, מתבצעת הערכה. מחושב המרחק הממוצע בין זוג קווי חמשה סמוכים בחמשה זו, ולפי מספר זה מדלגים מקו החמשה האחרון ועד שמגיעים לקואורדינטה המבוקשת. דרך יותר אינטואיטיבית ומעט מורכבת יותר תהיה לזהות את הקווים האופקיים שמעל או מתחת לתו ולספור אותם. כנראה שעבור תווים בכתב יד לא עתיד להיות מנוס מלהשתמש בשיטה זו, שכן שם אין שום התחייבות לשמירה על מרחקים קבועים מהחמשה.

זיהוי גבעולי תווים

זה התברר כמשימה פשוטה יחסית. מאחר שקווי החמשה נמחקו וראש התו זוהה בהצלחה (שהרי אין גבעול בלא ראש-תו), נותר לבצע סריקה פשוטה כלפי מעלה וכלפי מטה. גבעול יכול לבוע מימין לראש התו ואז הוא פונה מעלה או משמאל לראש התו, ואז הוא פונה מטה. נותר לעקוב אחרי הנקודות הכהות עד שהן נגמרות ולהכריז על הקצה העליון או התחתון של הגבעול, בהתאמה. הזיהוי הוכתר בהצלחה יחסית.

זיהוי זנבות

קבעתי את ההנחה הבאה: מתוך גבעול יכולים לנבוע שני סוגים של קווים: זנב וקו גובה, במקרה שהתו נמצא מתחת לקווי החמשה. אם זהו קו גובה, הוא חוצה את הגבעול משני צדדיו והוא דק. הקביעה לגבי העובי חשובה, משום שקו שמחבר תווים באורך שקצר מרבע פעמה גם הוא חוצה את הגבעול, אבל הוא עבה. לכן: אם קו נובע מתוך הגבעול מימין בלבד, או שהוא חוצה את הגבעול משני צדדיו וגם הוא עבה, מדובר בזנב. אחרת זה לא.

מסתבר שהיוריסטיקה הזאת לא תמיד עובדת: שאריות של נקודות שלא נוקו כמו שצריך מקווי החמשה זוהו בטעות כזנבות.

יצירת MIDI

יש מספר דרכים כיום שבאמצעותן ניתן לייצר קבצי MIDI. התוכנה החינמית LilyPond נועדה לשרטוט דפי תווים מתוך ייצוג טקסטואלי של הוראות. בנוסף לדפי תווים, יש לה אפשרות להפיק קובץ MIDI מתוך המידע המוסיקלי הזה. מהשיקולים האלה בחרתי להשתמש בה כדי להפיק את הפלט הרצוי מתוך התוכנה שלי. התווים המזוהים הם מימושים של הממשק Lilypondizable שמכיל שיטה אחת בלבד: החזר את הייצוג הטקסטואלי שלך בפורמט של LilyPond. לאחר שאוספים את כל החמשות, המפתחות והתווים, על כל אחד מהם לפי הסדר מבצעים את החילוף הטקסטואלי וכותבים את התוצאה לקובץ. התוצאה היא קובץ קלט עבור LilyPond שבהפעלתה עליו היא מפיקה קובץ MIDI וגם דף תווים, אם כבר. כתבתי קובץ אצווה עבור Windows שמריץ את התוכנה שלי ועל הפלט שלה את LilyPond. אם הכל פועל כשורה (התוכנה מסיימת בהצלחה, התוכנה LilyPond מותקנת ומזוהה אף היא) מתקבל קובץ MIDI. ניתן להשוות את דף התווים שמודפס בתוכנה לדף התווים שהיווה את המקור. על טיב התוצאות אעמוד בחלק המתאים.

סוגיות הנדסיות בפיתוח

כיוון שמדובר בתוכנה שסדר הגודל שלה לא ידוע מראש אבל לא גדול מדי (משימה למתכנת אחד), ההיקף שלה לא מוגדר בדייקנות אבל טיב התוצאה מוגדר היטב, ראיתי שדרך הפיתוח המומלצת במקרה שלי תהיה Extreme Programming. שיטה זו דוגלת בכתיבת מבחנים לקוד בטרם ייכתב הקוד עצמו, בבדיקה מתמדת, בהגדרה של יעדים קטנים מאוד, כתיבה ובדיקה שלהם ובמשוב מתמיד עם הלקוח. במקרה זה אני גם הלקוח וגם המפתח, אז זה היה קל. עוד כלל מפתח הוא להגדיר את רצף המשימות הבא ולבחור את הקלה ביותר קודם.

רשימת המשימות שהגדרתי היתה כדלקמן:

- לזהות חמשות
 - לזהות את המרחק המינימלי בין קווים (להלן: רווח)
 - לקבץ חמישיות של קווים לפי מרחקים מינימליים (אם יש יותר משלושה קווים אז זה כנראה זה)
 - לזהות קווי תיבה (חוצים את החמשות שזיהינו)
 - לזהות צורות עגולות (להלן: עיגולים) בגובה שקרוב לרווח.
 - לזהות עיגולים מלאים וריקים.
 - לזהות זנבות שמתחברים לעיגולים.
 - לזהות נקודות מימין לעיגולים.
 - לזהות סימנים נוספים: דיאז, במול, הפסקה: חצי, רבע, שמינית, שש-עשרית.
- הבעיה המרכזית עם שיטת הפיתוח של Extreme Programming היא שהתמונה הכללית לא ברורה לעין בגלל המזעריות של היעדים. לכן למפתח אין תחושה מדויקת אודות מידת ההתקדמות הגלובלית, וכאשר מועד ההגשה מתקרב, האמונה בצדקת הדרך מתחילה להתערער. את הזיהוי הראשוני של קווי חמשה וקווי תיבה (זיהוי קווי התיבה התברר מאוחר יותר כלא חיוני, והיה קשה להסב אותו למטרה אחרת) ביצעתי לפי הכללים, אבל בשלבים מאוחרים יותר, כאשר נדרשה כבר תוצאה להצגה כלפי חוץ ולא רק יעדי ביניים קטנים, העדפתי לזנוח את השיטה ולהסתער על כתיבת הקוד בצורה פחות מרוסנת. ניפוי השגיאות בקוד היה מסובך יותר, אבל בסופו יכולתי להיות בטוח שהתוכנה מפיקה פלט לפי הדרישות החיצוניות של המערכת.

סוגיות להתעמקות נוספת

אישור מוסיקלי

ברגע שהפענוח עצמו יהיה מניח את הדעת, יהיה טעם לנסות ולוודא את הנכונות של הפענוח לפי שמירה של מספר הפעמות בתיבה. כל עוד לא מוסיפים פענוח של תווים רגילים (OCR) לצד פענוח התווים המוסיקליים, לא ניתן לפענח את המשקל שלפיו התיבות בנויות, דבר שניתן לבקש מהמשתמש לספק כקלט (בהנחה שהמשקל קבוע לאורך היצירה). מלבד זאת, לא יהיה טיפול נכון בשלוש ומקבצים אחרים של תווים שתופסים ביחד פחות מסכומם – הם מסומנים על-ידי המספר של הקיבוץ (לדוגמא, שלשה מסומנת בעזרת הספרה 3 מעל שלושת תוויה), שגם אותו יש לפענח. זה כבר דורש ממשוק עם תוכנה חיצונית שמטפלת בתחום OCR, או להוסיף להיקף המערכת את יישומו של מנגנון כזה.

יישור דפים עקומים

יש לי הרגשה שההיריסטיקה כגון זו תעבוד, אם כי לא היתה לי ההזדמנות לבדוק אותה: לספור את מספר השורות הכהות עבור כל אוריינטציה אפשרית, ולבחור את האוריינטציה שמניבה את המספר המקסימלי של שורות כהות.

ייצוג אוניברסלי של סימנים מוסיקליים

על מנת לעבור לשלב הבא בפיתוח המערכת, יש לאפיין את הסימנים שהמערכת צריכה לזהות בצורה כללית מאוד, אשר תהיה נכונה גם עבור מרבית הגופנים המשמשים להדפסת תווים (במחשב ובדפוס), וגם תתאים לגישה שנוקטים כותבי תווים ביד כאשר הם מחקים את צורת הדפוס באמצעות עט מודרני (ולא בעזרת ציפורן וקולמוס דיו כמו לפני כמה מאות). זה דורש גישה חדשה לעיבוד התמונה המוסיקלית שתתמוך בזיהוי של צורות גיאומטריות כגון מעגל, קשת ועיקול. אף על פי שהמשימה נשמעת במבט ראשון כמו זיהוי גבולות, היא השלב הבא לאחריו, שכן התמונה שברשותנו מורכבת כולה מגבולות אשר יש כעת לפענח אותם – לא מבחינת "סיבת" הגבול אלא מבחינה גיאומטרית. האם זה קו ישר, קו עקום, ועבור קו עקום – האם זוהי קשת, מפתח סול, סלסול של הפסקה, או מה. מנקודה זו יהיה אפשר להתאים את הצורות שאובחנו לעצמים שבמילון צורות, לפי טרנספורמציות Hough או בכלים בסיסיים יותר.

מסקנות

בתור התנסות ראשונית, הפרויקט מציע כמה דרכים בסיסיות לגשת לעולם הרחב של זיהוי צורות בהקשר המוסיקלי. הגישה צנועה ומעט נאיבית, אבל היא הצליחה להשיג כמה יעדי ביניים ולהגיש תוצר סופי שמצליח לתת מידה מסוימת של דיוק, עבור מקרים שעליהם ההנחות המקדימות מתקיימות. עם שיפור של כל אחד מהכלים שביססתי יכולת הפענוח לבטח תשתפר, ובינתיים היא מהווה טעימה קטנה מהתחום הרחב של ראייה חישובית.