# **Summary in Hebrew**

לממשק בין אדם-מחשב יש השפעה רבה על הפרודוקטיביות וקלות השימוש בכלים ובמחשבים של ימינו. בשנים האחרונות נעשו מספר ניסיונות לפתח טכנולוגיות חדשות גם בתחומים הדורשים דיוק רב במיוחד במהלך , כמו במציאות מדומה, או שימוש במכונות כבדות. ציוד לכך כיום הוא לעתים מגושם ומלווה בחומרה נוספת שהופכת אותו ליקר. מטרת הפרויקט היא ליצור פתרון עבודה לאינטראקציה בין אנשים למחשבים. פרויקטנו נועד לספק אלטרנטיבה אינטואיטיבית וזולה יחסית המאפשרת תנועה מדויקת בזמן אמת. הוא רותם ארכיטקטורה של למידת מכונה כדי לחלץ תנוחות ידיים ומפרקים מקלט חי של תמונות RGB ואז ממפה אותן לפלט תיאום ג'ויסטיק. המערכת מאפשרת למשתמש להכניס קלט למערכת הדומה לג'ויסטיק על ידי הטייה של הידיים מול מצלמה.

שתי חלופות מערכת נשקלו:

1. שימוש בתמונות גולמיות לאימון מודל רגרסיה באמצעות תמונת RGBD. הבעיות ? של שיטה זו הן הרגישות לשינויים בתאורה, בהגדרות, ושינויים נוספים במערך.
2. שימוש במודל גרפי hand-graph-cnn לחילוץ מיקומי היד והמפרקים מתמונת RGB. שיטה זו כביכול מדויקת, אך אינה מיועדת לזיהויי אינטראקציות של אובייקטים וידיים, דבר שיקשה רבות על אימון המודל השני שלנו, כמו כן היא אינה מייצרת תוצאות בזמן אמת.

לכן עבור המוצר הסופי שלנו החלטנו להשתמש במודל הפועל על המעבד ומספק תוצר למגוון רחב יותר של מכשירים.

המוצר מורכב משתי רשתות - הראשונה היא רשת נקודות ציון ידיים המוציאה את המיקומים התלת מימדיים של מפרקי הידיים מזרם של תמונות RGB בעזרת מצלמה. הרשת השנייה היא רשת וקטור-קואורדינטות המתרגמת את המיקום התלת-מימדי למיקום הדו מימדי של הג'ויסטיק הרצוי באמצעות רשת למידה עמוקה.

ביצועי המערכת טובים מספיק כדי לפעול בזמן אמת, והיא מצליחה לתפקד אפילו בתנאי סביבה שונים - כגון תאורה משתנה, מיקום פיזי, רקע משתמשים שונים שמפעילים אותה.   
  
הנתונים לשלב האימון והבדיקה נאספו באמצעות ג'ויסטיק פיזי המשמש כבסיס לשלבי האימון והבדיקה שלנו. חישבנו את ה- MSE (שגיאת ריבוע ממוצעת) בין מיקומי הג'ויסטיק האמיתיים לבין לתוצאה החזויה על ידי המודל.

המסקנה אליה הגענו בסוף הפרויקט היא ששיטה זו אמינה מספיק לשימוש ביישומים רבים, היא עובדת עם מגוון ידיים ורקעים.

# **Executive Summary**

Human-Computer interface has a large influence on productivity and ease of use of today's tools and computers, and in recent years there were several attempts to expand new technologies to fields that require different types of handling, such as Virtual Reality gear, or heavy machinery. Equipment for this today is often clumsy and accompanied by surrounding hardware which makes it more expensive than it could be. The purpose of this project is to create a working solution for interaction between people and computers. Our project aims to provide an intuitive and relatively cheap alternative that allows precise movement in real time. It harnesses Machine-learning architecture to extract hand and joint poses from a live input stream of RGB images then map them to a joystick coordinate output. This allows the user to emulate joystick input to the system by tilting his/her hand in front of the camera.

Two system alternatives were considered:

1. Using raw images for training a regression model using an RGBD Image. The caveats of this method is the sensitivity for changes in the lighting, settings and variations in setup.
2. Using an alternative model hand-graph-cnn for obtaining the hand and joints coordinates from an RGB Image. This method is supposedly accurate, however it is not robust for hand-object interactions, and due to lack of access to the college gpu and work environment, we’ve decided to use a model which runs on the cpu, and provides a working prediction to a larger variety of devices.

The project consists of two networks.

1. A hand landmark network that extracts the 3d coordinates of the joints of the hands from a stream of RGB images from a camera.
2. A Coordinates-to-vector network which translates the 3d coordinates to the joystick coordinates required.

The system is fast enough to work in real time, and is functional in various conditions. The data for the training and testing phase was gathered using a physical joystick that functions as the baseline for our training phase and results. We calculated the MSE (mean squared error) between the joystick and the produced result.

This method is reliable and accurate enough for usage in many applications. The tracking works in various environments and types of hands, and requires only minor adjustments to be comfortably used by different users.