קידוד יעיל של רצפי תמונות
של סצינות מישוריות

מאת

אודי שרוט

עלון יוצא
מאגר הספר
0458

c.1

לא להשאלה

קידוד יעיל של רצפי תמונות של סצינות מישוריות

חיבור על עבודת גמר

לשם מילוי חלקי של הדרישות לקבלת התואר

מגיסטר למדעים

בהנדסת חשמל

אודי שרוט

הוגש לסנט הטכניון - מכון טכנולוגי לישראל
אב תש"ס חיפה אוגוסט 2000

2230458

הספריה המרכזית ע"ש אלישר
מס'
מערכת _____

עבודת הגמר נעשתה בהנחיית פרופ' דוד מלאך וד"ר מאיר בר-זוהר בפקולטה להנדסת חשמל.

עבודת הגמר נתמכה חלקית ע"י רפא"ל וע"י מוסד שמואל נאמן בטכניון.

תודתי העמוקה נתונה לפרופ' דוד מלאך ולד"ר מאיר בר-זוהר על הנחייתם המסורה והצמודה.

תודה לצוות המעבדה לעיבוד אותות ותמונות , נמרוד פלג, זיוה אבני וליתר אנשי המעבדה על עזרתם הרבה ותרומתם ליצירת תנאי עבודה מצוינים.

תוכן עניינים

תקציר 1

רשימת סמלים וקיצורים 4

1 מבוא 5
1.1 קידוד רצפי תמונות 5
1.2 מטרת העבודה 6
1.3 מבנה העבודה 7

2 התאמת תמונות 9
2.1 הקדמה 9
2.2 הגדרת התאמת תמונות 10
2.3 שערוך התמרה מרחבית 10
2.3.1 באמצעות הנחת מודל תמונה 10
2.3.2 באמצעות Optical Flow 11
2.3.3 מבוססת קורלצייה 12
2.3.4 באמצעות התמרת פורייה 12
2.3.5 באמצעות התאמת נקודות התאמה 13
2.4 שערוך התמרה מרחבית מקומית 16
2.5 סיכום 17

3 מציאת נקודות התאמה 18
3.1 סקר ספרות - מציאת נקודות התאמה 18
3.2 אלגוריתמים למיצוי נקודות התאמה 20
3.2.1 אלגוריתם I ("שיטת גראדיינטים") 20
3.2.2 אלגוריתם II ("שיטת הערכים העצמיים") 21
3.3 סימולצייה והשוואת ביצועים 25

4 קידוד רצף תמונות המבוסס על שערוך התמרה אפינית גלובלית 28
4.1 הקדמה 28
4.2 תנועת מצלמה 30
4.3 שיטת ריבועים פחותים (Least Squares) לשערוך פרמטרי תנועה גלובלית 35
4.4 תוצאות שערוך תנועה גלובלית באמצעות פתרון בעיית Least Squares 38
4.5 קידוד רצף תמונות באמצעות תמונת ייחוס ומידע התנועה 39

5 החלקת פרמטרי התמרה אפינית באמצעות מסנן קלמן לינארי 44
5.1 מסנן קלמן לינארי 44
5.2 אלגוריתם להחלקת נקודות התאמה 45
5.2.1 תוצאות סימולצייה 46

5.3 החלקת פרמטרי טרנספורמציה אפינית באמצעות מסנן קלמן 48

6 שערוך פרמטרי התמרה מישורית 50
6.1 התמרה מישורית 50
6.2 סקר ספרות - שערוך פרמטרי התמרה מישורית 52
6.3 שיטות Least Squares לשערוך פרמטרי התמרה המישורית 60
6.4 פתרון מקורב לשערוך פרמטרי התמרה מישורית 62
6.5 אלגוריתם מוצע לשערוך פרמטרי התמרה מישורית ברצף תמונות 66

7 שערוך פרמטרי תנועת מצלמה בסצינה מישורית 74
7.1 סקר ספרות - שערוך פרמטרי תנועת מצלמה 76
7.1.1 שערוך פרמטרי תנועת מצלמה בהתבסס על שערוך optical flow 76
7.1.2 שערוך פרמטרי תנועת מצלמה המבוסס על נקודות התאמה 77
7.1.3 שערוך פרמטרי תנועת מצלמה בתסריטים מישוריים 79
7.1.4 שימוש במספר מישורים לצורך שערוך פרמטרי תנועת המצלמה 80
7.2 שערוך פרמטרי תנועת מצלמה באמצעות פרמטרי ההתמרה המישורית 81

8 סיכום ומסקנות 86

נספח א: ראיית סטריאו 88

נספח ב: התמרה מסדר שני 92

מקורות 95

תקציר באנגלית I

רשימת איורים

איור	עמוד
3.1: אלגוריתם I ("שיטת גראדיינטים") למציאת נקודות התאמה	20
3.2: מוצא אלגוריתם I למיצוי נקודות התאמה - "שיטת הגראדיינטים"	21
3.3: תמונת "Salesman" נעה במהירות זוויתית של 2.5° למסגרת	26
3.4: בחירת נקודות ההתאמה	26
3.5: אנרגיית תמונת ההפרשים – שערוך תנועה באמצעות נקודות התאמה	27
4.1: מצלמת חריר – pin-hole	28
4.2: הקשר בין מערכת הצירים של המצלמה לבין מערכת צירי הייחוס	30
4.3: גיאומטריה של שיטת LS	36
4.4: השוואה בין שיטות שונות לשערוך תנועה	39
4.5: "שיטת הפרשים" – (Incremental Method)	40
4.6: שיטת שערוך תנועה מבוססת חיזוי	41
4.7: השוואה כמותית בין שיטות שערוך תנועה ברצף תמונות	42
4.8: השוואה סובייקטיבית בין שיטות שערוך תנועה ברצף תמונות	43
5.1: מסנן קלמן לינארי	45
5.2: החלקת נקודות עקיבה באמצעות מסנן קלמן	47
5.3: החלקת נקודות עקיבה באמצעות מסנן קלמן – שתי איטרציות	47
5.4: החלקת פרמטרי התמרה אפינית באמצעות מסנן קלמן	49
6.1: סצינה מישורית	50
6.2: התמרה אפינית	64
6.3: התמרה בי-לינארית	65
6.4: מודל תנועה מסדר שני	65
6.5: אלגוריתם לשערוך התמרה מישורית	67
6.6: שימוש במודל מסדר שני כאמצעי לשיפור שערוך התמרה מישורית	68
6.7: רצף תמונות סינתטיות שנוצרו באמצעות התמרה מישורית	70
6.8: ביצועי האלגוריתמים לשערוך התמרה מישורית על רצף סינתטי	70
6.9: מיצוע תמונות הרצף לאחר קיזוז תנועה (20 תמונות) – אלגוריתם I (רצף סינתטי)	71
6.10: מיצוע תמונות הרצף לאחר קיזוז תנועה (20 תמונות) – אלגוריתם II (רצף סינתטי)	71
6.11 : רצף תמונות שצולמו ממצלמה נעה	72
6.12 : ביצועי האלגוריתמים לשערוך התמרה מישורית על רצף אמיתי	72
6.13 : מיצוע תמונות הרצף לאחר קיזוז תנועה (30 תמונות) – אלגוריתם I (רצף אמיתי)	73
6.14: מיצוע תמונות הרצף לאחר קיזוז תנועה (30 תמונות) – אלגוריתם II (רצף אמיתי)	73
7.2: ביצועי אלגוריתם לשערוך פרמטרי תנועת מצלמה בתסריטים מישוריים	85
A.1: מערכת צילום סטריאו	89

תקציר

ההתפתחויות האחרונות בעולם הטכנולוגיה בכלל ובעולם התקשורת בפרט מעלות אתגרים קשים ודרישות גבוהות לתקשורת ויזואלית באיכות טובה, מחד גיסא, תוך חסכון במשאבי זכרון ובקצב שידור, מאידך גיסא. דרישות אלה מתנגשות ביניהן ולכן יש צורך בשיטות המבצעות אופטימיזציה בין הדרישות, תוך עמידה בתנאי העבודה הנדרשים. שיטות אלה צריכות לאפשר מצד אחד את החיסכון הנדרש במשאבי זכרון ובקצב שידור, אך עליהם להקפיד על כך שאיכות התמונה המתקבלת תהיה טובה ותאפשר צפייה ברורה ונוחה בתמונה.

השיטות הנפוצות לקידוד תמונות רצף מתבססות על ניצול היתירות המרחבית והיתירות הזמנית הקיימות ברצף. היתירות הזמנית מתבטאת בכך שתמונות עוקבות ברצף דומות זו לזו וניתן לנצל אותה ע"י קידוד ההפרש בין התמונות במקום קידוד התמונות עצמן. מכיוון שאנרגיית ההפרש קטנה בד"כ מאנרגיית האות עצמו, מושג החסכון ע"י קידוד ההפרש. היתירות המרחבית מתבטאת בכך שיש קורלציה חזקה בין פיקסלים קרובים בתוך התמונה והיא מנוצלת ע"י קידוד התמרה. ההתמרה מבוצעת על בלוקים והיא מביאה לריכוז של רוב מידע הבלוק במספר מקדמים קטן. קוונטיזציית מקדמי התמרה נעשית בתחום התדר בצורה מותאמת לתכונות מערכת הראייה האנושית והיא קובעת את מידת החיסכון בקצב הקידוד ואת מידת הפגיעה באיכות התמונה. ע"י בחירה נבונה ניתן לחסוך בקצב השידור הנדרש מבלי שהמתבונן יוכל להרגיש בירידת האיכות המתקבלת. חסכון נוסף בקצב השידור מתקבל ע"י קידוד מקדמי ההתמרה ומידע צד נוסף קוד באורך משתנה (קוד VLC).

לשם שיפור החיזוי של התמונה שאותה רוצים לקודד משתמשים גם בשערוך וקיזוז תנועה. חלק גדול מהשינויים בין תמונות עוקבות ברצף תמונות נובע מתנועות של עצמים בתוך הסצינה ולפיכך ניתן להפחית את הקצב ע"י קידוד תמונת השגיאה בין התמונה המקודדת לחזוי שלה, המתקבל בהסתמך על התמונה הקודמת ושערוך התנועה בין התמונות. במקרה זה נדרש, בד"כ, לקודד גם את המידע על התנועה אולם השיפור בחיזוי מצדיק את הגישה. צורת קיזוז התנועה המקובלת במקודדים סטנדרטים של רצף-תמונות היא קיזוז תנועה מקומי המניח שתנועת הבלוקים בתמונה נובעת מהזזה אחידה. הנחה זו באה לאפשר תיאור התנועה ע"י כמות מידע צד סבירה ונובעת גם משיקולי חישוביות ופשטות אלגוריתמית, המאפשרת שימוש נוח בחומרה. ברצפים של סצינות תלת-מימדיות

הנחה זו אינה מתאימה, אפילו כאשר יש רק תנועת הזזה של המצלמה, מכיוון שמידת התנועה של פיקסלים סמוכים משתנה בהדרגתיות בהתאמה למרחקם מהמצלמה. כאשר יש Zoom או סיבוב של המצלמה ההנחה אינה נכונה אפילו לסצינות דו-מימדיות. במקרים אלו כדאי להפעיל טכניקות לשערוך תנועה גלובלית המבוססות על מודל תנועה מתאים יותר, במטרה לשפר את קיזוז התנועה. ככל שקיזוז התנועה טוב יותר, כך תתקבל תמונת חיזוי טובה יותר לתמונה שיש לקודד, וכתוצאה מכך מתקבל שיפור בקצב השידור הנדרש או באיכות התמונה.

התנועה הגלובלית בסצינות מסוימות כגון סצינות רחוקות מהמצלמה או סצינות על מישור יחיד ניתנת לתיאור ע"י סט פרמטרים יחיד המתאר את התנועה במישור התמונה. בנוסף ליתרון של צמצום מידע צד הנדרש לתיאור תנועה זו ישנו גם היתרון של דיוק גבוה יותר בתיאור התנועה. מטרת עבודה זו להציע מקודד יעיל לקידוד רצף התמונות באמצעות שידור התמונה הראשונה והמידע על התנועה הגלובלית. עבודה זו עוסקת בשערוך פרמטרי השינויים הגלובליים בסצינות אלו ומושם דגש, באלגוריתמי השערוך המוצעים, על שערוך פרמטרים בדיוק גבוה, על רובסטיות השערוך ועל כמות חישובים נמוכה.

כאשר מישור הסצינה מקביל בקירוב למישור המצלמה ניתן לתאר את התנועה הגלובלית באמצעות ההתמרה האפינית בעלת שישה פרמטרים. כאשר הסצינה נמצאת על מישור אחד ואיננה מקבילה בקירוב למישור המצלמה יש לתאר את תנועת הסצינה באמצעות ההתמרה המישורית בעל שמונה פרמטרים. שיטות שערוך רגילות לשערוך ההתמרה המישורית סובלות מחוסר יציבות נומרי בתנאי רעש עקב אי - הלינאריות של ההתמרה המישורית. עבודה זו מציעה שיטה לשערוך הטרנספורמציה המישורית בצורה רובוסטית יותר המניבה תיאור סצינה מדוייק יותר.

כאשר התנועה בין התמונות קטנה ניתן להשתמש במודל מקורב של ההתמרה המישורית. שערוך המודל המקורב רובסטי יותר בתנאי רעש משערוך המודל המישורי המדוייק. השיטה המוצעת מורכבת משילוב של המודל המקורב והמודל המישורי המדוייק לצורך ייצוג רצפי תמונות. זאת באמצעות שידור התמונה הראשונה והמידע על התנועה הגלובלית, למרות שהתנועה הגלובלית בין תמונת הייחוס לתמונות הרצף גדולה.

אלגוריתמי השערוך המתוארים בעבודה זו מומשו בתוכנה על גבי מחשב ונוסו על תמונות רצף אמיתיות ומלאכותיות. התוצאות שהתקבלו מובאות בגוף העבודה. התוצאות

מצביעות על כך שאלגוריתם השערוך המוצע נותן תוצאות מדוייקות יותר מהשיטות המקובלות.

רשימת סמלים וקיצורים

BM	Block Matching
F_1	Camera focal length
$I(X,Y)$	Image intensity at pixel (X,Y)
LS	Least Squares
M_{ext}	Camera's extrinsic parameters matrix
M_{int}	Camera's intrinsic parameters matrix
R	Rotation matrix
Rank[]	The rank of a matrix
RMSE	Root Mean Squared Error
SVD	Singular Value Decomposition
$\vec{T}$	Translation vector
TLS	Total Least Squares
(u,v)	2-Dimensional motion vector
(X,Y)	A point on the image plane (in pixels)
(x,y)	A point in the camera's coordinate system
(x',y')	A point in the world reference coordinate system

פרק 1 מבוא

כיום, הדרישה לקידוד יעיל וחסכוני של רצפי תמונות גדלה והולכת .הדרישות הגדלות לשידור אות חוזי לצורכי מולטימדיה ו – Video Conferencing בערוצים הקיימים שהם בעלי קצב השידור נמוך יחסית מחייבות שימוש בשיטות לדחיסת המידע. הדרישה הגוברת לתקשורת המשלבת מידע חזותי כגון תמונות וידאו ודרישה לאחסנתם באיכות גבוהה ובאופן מהיר ונוח לגישה, הביאו למחקר רב בנושא של קידוד רצפי תמונות.

1.1 קידוד רצפי תמונות

אות טלוויזיוני אנלוגי (CCIR-601) שנדגם בפורמט CIF (בגודל תמונה של 180x360 פיקסלים) מחייב קצב תמסורת של 37Mbps. לכן המשאבים הדרושים לאחסון או לשידור סרט ספרתי הם עצומים ומחירם גבוה למדי. עקב בעיה זו והדרישה הגוברת לשדר ו\או לאחסן סרטים ספרתיים, הוצעו בשנים האחרונות מספר שיטות לדחיסת סרטים במטרה להשיג הורדת הקצב או נפח האחסון תוך שמירה על איכות טובה של הסרטים המשוחזרים. הטכנולוגיה הנוכחית מאפשרת מימוש של רכיבי דחיסה יעודיים הפועלים בזמן אמת, ללא השהיות.

אמצעי אחסון כגון CD –ROM (Compact Disk – Read Only Memory) מסוגלים לפעול בקצב של עד 4 Mbps ולכן הסטנדרטים שגובשו ע"י קבוצת מומחים הקרויה MPEG (Motion Pictures Expert Group) ונקבעו ע"י ארגון ISO (International Standard Organization). מתאימים לקצבי שידור אלה. תקן MPEG 1 – מיועד לשידור רצף תמונות בקצב 1.15 Mbps (התקן מאפשר לשדר מידע בקצב של עד 1.5 Mbps כך שההפרש מנוצל לשידור מידע קול ואינפורמציית צד – נוספת). תקן MPEG – 2 מיועד לדחיסת רצפי תמונות באיכות גבוהה . הקצבים של תקן זה הינם 4-9 Mbps.

הבסיס של השיטות לקידוד רצף תמונות הינו ניצול היתירות שבמידע המייצג את התמונות. בנוסף משתמשים בתכונות מערכת הראייה האנושית HVS (Human Visual System) לדחיסה עמוקה יותר של המידע שהעין איננה רגישה לו ושימור המידע שהעין כן רגישה לו. בנוסף לקורלצייה המרחבית הקיימת בתמונה הבודדת (still image), שניתן לנצלה לשיפור הקידוד של תמונה זו, הרי שברצף תמונות קיימת גם קורלצייה זמנית הנובעת מהעובדה שתמונות קרובות בזמן דומות זו לזו. לפיכך, רוב השיטות לדחיסת רצף

תמונות מבוסס על רעיון החיזוי של תמונה נוכחית בהסתמך על תמונות קודמות הנמצאות כבר במקלט, כלומר התמונות ששודרו ושוחזרו. בתקני ה – MPEG חיזוי תמונה ברצף נעשה על סמך תמונה קודמת או התמונה הבאה או שילוב של שניהם. בשלב השני, נבנות ומקודדות תמונת ההפרש בין התמונה החזוייה לתמונה הרצויה.

אלגוריתם החיזוי שבין התמונות מבוסס בד"כ על שערוך התנועה שבין התמונות. התקנים המקובלים כיום H.261, H.263, H.263+, MPEG – 1, MPEG – 2 [MP90] מבוססים על שערוך תנועה מקומית של קבוצות פיקסלים (בלוקים).קיימות גם עבודות המשלבות שערוך תנועה גלובלית בין התמונות וקיזוז של תנועה זו [WuK90]. שינויי תנועה גלובלית נובעים מתנועת מצלמה או שינוי פרמטרי מצלמה כגון בפעולת zoom . סצינות מסויימות, כגון סצינות רחוקות מהמצלמה או סצינות על מישור יחיד ניתנת לתיאור ע"י סט פרמטרים יחיד. בסצינות מורכבות יותר התנועה הגלובלית קשה יותר לתיאור ותיאורה אפשרי אז באמצעות סגמנטציה של התמונה למספר איזורים ומציאת פרמטרי התנועה לכל אחד מאיזורים אלו.

קווי תקשורת רבים כגון קו ISDN (המאפשר שידור בקצבים של 64-128 Kbps) או קווי טלפוניה רגילים (קצב 0.3-56 Kbps) מאפשרים שידור בקצבים נמוכים יחסית. עבודה זו מציעה לקודד רצפי תמונות הניתנות לתיאור באמצעות תנועה גלובלית באמצעות שידור של התמונה הראשונה (כתמונת ייחוס), בקבוצת תמונות ברצף, ופרמטרי התנועה בין התמונה הראשונה לתמונות הרצף. צורת קידוד זה יכולה להניב יחסי דחיסה גבוהים.

1.2 מטרת העבודה

תנועת מצלמה כגון הזזה תלת – מימדית שלה במרחב, שינוי מרחק המוקד או סיבובה סביב ציר כלשהוא גורמים לשינויים גלובליים ברצף התמונות. פעולות אלו משפיעות על כל התמונה אך ניתנות לתיאור ע"י מספר קטן יחסית של פרמטרים.

התנועה הגלובלית בסצינות מסוימות כגון סצינות רחוקות מהמצלמה או סצינות על מישור יחיד ניתנת לתיאור ע"י סט פרמטרים יחיד המתאר את התנועה במישור התמונה. מטרת עבודה זו לבחון גישה לקידוד יעיל של קבוצת תמונות ברצף באמצעות שידור התמונה הראשונה והמידע על התנועה הגלובלית. עבודה זו עוסקת בעיקר בשערוך פרמטרי השינויים הגלובליים בסצינות אלו, תוך השמת דגש באלגוריתמי השערוך המוצעים

על שערוך מדוייק ככל האפשר של הפרמטרים, על רובסטיות השערוך ועל כמות חישובים נמוכה. ככל ששערוך פרמטרי התנועה מדוייק יותר, ייצוג קבוצת התמונות מדוייק יותר. כמו כן, אם קיימת אפשרות לשידור מידע צד לשיפור קידוד הרצף, אזי שערוך מדוייק יותר של פרמטרי התנועה מאפשר לשדר מידע צד מצומצם יותר לקבלת אותה איכות ייצוג.

כאשר מישור הסצינה מקביל בקירוב למישור המצלמה ניתן לתאר את התמונה הגלובלית באמצעות טרנספורמציה אפינית שהיא בעלת שישה פרמטרים. כאשר הסצינה נמצאת על מישור אחד ואיננה מקבילה למישור המצלמה יש לתאר את תנועת הסצינה באמצעות טרנספורמציה מישורית בעלת שמונה פרמטרים. עבודה זו מתמקדת במיוחד בשיטה לשערוך מדוייק יותר של פרמטרי הטרנספורמציה המישורית.

איכות שערוך התנועה בעבודה זו נמדדת בעבודה זו בשתי דרכים שונות: אנרגיית תמונת ההפרש בין התמונה החזוייה לבין התמונה הרצוייה והאיכות הויזואלית של תמונות הרצף לאחר קיזוז התנועה (עד כמה תמונות אלו דומות לתמונה הראשונה ברצף).

1.3 מבנה העבודה

על מנת לייצג קבוצת תמונות רצף באמצעות התמונה הראשונה (תמונת ייחוס) יש למצוא בכל שלב את ההתאמה בין שתי תמונות. כלומר לבצע רגיסטרציה בין שתי התמונות. פרק 2 מכיל סקירה של שיטות להתאמה בין תמונות.

שערוך התנועה הגלובלית מבוצע באמצעות מציאת סט של נקודות בולטות בשתי תמונות ושערוך פרמטרי התנועה מתוכן. פרק 3 מכיל סקירה של שיטות למיצוי נקודות בולטות.

פרק 4 מכיל תיאור של גישת הקידוד המוצעת, כאשר הסצינה מקבילה בקירוב למישור התמונה (שימוש בטרנספורמציה אפינית).

פרק 5 מציג מספר גישות לשיפור שערוך פרמטרי הטרנספורמציה האפינית בהתבסס על מסנן קלמן.

פרק 6 מכיל סקירה של השיטות הקיימות לשערוך פרמטרי טרנספורמציה מישורית ואת השיטה הרובוסטית יותר המוצעת בעבודה זו.

פרק 7 מכיל סקירה של שיטות לשערוך פרמטרי תנועת מצלמה בתסריטים מישוריים וכן כיצד שערוך מדוייק יותר של פרמטרי הטרנספורמציה המישורית מאפשר לשפר את שערוך פרמטרי תנועת המצלמה בתסריטים אלה.

פרק 8 מכיל סיכום של עבודה זו, מסקנות וכיווני המשך.

פרק 2 התאמת תמונות

2.1 הקדמה

בהתאמת תמונות הכוונה היא מציאת העתקה של נקודות בתמונה אחת למיקום חדש בתמונה השנייה תוך שינוי אפשרי בערכי הפיקסלים.

התאמת תמונות (רגיסטרצייה) הינה אחת המשימות הבסיסיות בעיבוד תמונה. מטרת הרגיסטרציה להתאים שתי תמונות שבמקרה הכללי צולמו מזוויות שונות, באמצעות חיישנים שונים, בזמנים שונים או בתנאי תאורה שונים. לאחר שהתמונות הותאמו ניתן להבחין בשינויים "האמיתיים" בין התמונות – כלומר שינויים שלא ניתן לייחסם לתנאי הצילום (כגון תנועות מקומיות בין שתי התמונות).

להתאמת תמונות שימושים מעשיים רבים. בין השימושים ניתן לציין:

א. שילוב אינפורמציה המתקבלת מחיישנים שונים על – מנת לשפר את הבנת התמונה. לדוגמא,שילוב תמונות רפואיות המתקבלות ממחושים שונים :תמונת רפואיות "אמיתיות" כגון MRI (Magnetic Resonance Imaging) ו – CT (computed tomography) ואולטרהסאונד המשולבות עם תמונות תפקודיות של הגוף כגון SPECT (single photon emission computed tomography) וכו' מאפשר מצד אחד לזהות את הבעיה (באמצעות התמונות "התפקודיות") ולמקמה במדויק באמצעות התמונות "האמיתיות". התאמת תמונות רפואיות משפרת את הבחנת הבעיה הרפואית, מאפשרת תכנון ניתוח טוב יותר וטיפול מדוייק יותר[Br92].

ב. שינויים שחלו בתמונות שצולמו בזמנים או בתנאים שונים - כגון התאמת תמונות לווין לצורך ניטור [ToJ89].

ג. מיצוי מידע תלת-מימדי של הסצינה כאשר המצלמה נעה או לחלופין שערוך פרמטרי המצלמה מתוך רצף תמונות [WHA93].

ד. זיהוי עצם ייחוס בתוך התמונה - שימוש צבאי בהתאמת תמונת מטרה לחלק מתמונה המתקבלת מהמצלמה בזמן – אמת [Br92].

עקב השימושים והאפשרויות הרבות של התאמת תמונות קיים מגוון רחב של שיטות לביצוע ההתאמה. רב שיטות ההתאמה מבוססות על חיפוש במרחב ההתמרות מסוג מסויים את ההתמרה המניבה את ההתאמה הטובה ביותר. סוג השינויים שבין התמונות מכתיב את אופן שערוך ההתמרה, בחירת מרחב ההתמרות האפשרי לביצוע ההתאמה וקביעת המדד לאיכות ההתאמה.

יש להבדיל בין התמרה גלובלית בה סט מקדמים אחד מתאר את שינוי התמונה כולה לבין התמרה מקומית, בה לכל אזור בתמונה סט פרמטרים שונה.

קיימים שלוש תרחישים עיקריים להתאמת התמונה, בהתאם לאופי השינויים שבין התמונות.

1. שערוך התמרות (טרנספורמציות) מרחביות שנועדו לתקן שינויים עקב צילום ממיקום שונה או מזווית שונה. בחירת הטרנספורמציה המרחבית מבוססת על הנחות לגבי מאפייני השינוי.
2. מציאת התאמה עבור מקרים בהם התמונות צולמו מזוויות שונות, אולם לא ניתן להתאימם בצורה פשוטה עקב שינויי תאורה והשפעות אטמוספריות. כלומר התיקון איננו מרחבי בלבד, אלא כולל גם שינוי של עוצמת התמונה. תיקון זה קשה לביצוע משום שאופי השינוי תלוי בתכונות הפיזיות של הסצינה.
3. מציאת התאמה כאשר חלו שינויים בסצינה כגון, תנועה מקומית של אלמנטים בתוך אחת התמונות. לא ניתן לפצות על שינויים אלו ובד"כ שינויים אלו רצויים משום המידע הקיים בהם לגבי תנועות עצמים בתוך הסצינה.

2.2 הגדרת התאמת תמונות

התאמת תמונה הינה מציאת המיפוי בין שתי תמונות הן מרחבית והן מבחינת העוצמה:

$$I_2(x,y) = g(I_1(f(x,y))) \qquad (2.1)$$

f הינה טרנספורמציה מרחבית דו מימדית $(x',y') = f(x,y)$ ו - g הינה טרנספורמצית עוצמה חד – מימדית.

מטרת ההתאמה הינה מציאת טרנספורמצייה כוללת כך שימוזערו ההפרשים בין התמונות, לצורך קבלת פונקציות f ו - g מתאימות, או לצורך גילויי השינויים הרצויים בין שתי התמונות.

טרנספורמציית העוצמה g לא תמיד חשובה ובד"כ טבלת lookup מספיקה. הטבלה נקבעת באמצעות ידיעת מאפייני המצלמה או תנאי התאורה בהן נלקחו התמונות.

2.3 שערוך התמרה מרחבית

התאמת תמונה מושגת בד"כ באמצעות שערוך ההתמרה המרחבית האופטימלית (כך ששתי התמונות תתאמנה ככל האפשר). על הטרנספורמציה המרחבית לפצות רק על העיוותים הנגרמים עקב השינוי במיקום ובזווית הראייה.

2.3.1 שערוך התמרה המרחבית באמצעות הנחת מודל תמונה.

בשיטות אלו פרמטרי התנועה הגלובלית מחושבים באמצעות הנחת מודל של ערכי רמת האפור בתמונה. ב– [WuK93], [WuK90] מוצע למצוא את פרמטרי ההתמרה המרחבית בעזרת איטרציות של התאמת הגרדיאנטים בתמונה לגראדינטים הצפויים להתקבל עקב התנועה הגלובלית. ב – [Di91], [HoT88] מבוצעת מינמיזציה איטרטיבית של פונקצית

השגיאה בין התמונה הנוכחית לחיזוייה. תמונת החיזוי מתקבלת באמצעות קיזוז תנועה של התמונה הקודמת תוך הנחת מודל מסויים של פונקציית רמות האפור של התמונה.

2.3.2 שערוך התמרה מרחבית באמצעות Optical Flow

בגישה זו מבצעים שערוך של ה- Optical Flow, כלומר שערוך וקטור תנועה דו מימדי (u,v) לכל פיקסל בתמונה בין שתי התמונות. ישנן מספר שיטות לשערוך Optical Flow, כגון [HoS81], [KTB87]. שיטות אלו מבוססות על ההנחה הבסיסית שעוצמת כל אלמנט בתמונה קבועה בין תמונות עוקבות, בין אם האלמנט סטטי או דינמי. ניתן להראות באמצעות פיתוח קירוב טור טיילור מסדר ראשון של הקשר

$$I(x,y,t) = I(x+dx, y+dy, t+dt) \tag{2.2}$$

שעבור אלמנט (x,y) הנע (dx,dy) בזמן dt מתקבל

$$\frac{\partial x}{\partial t}\bullet\frac{\partial I(x,y,t)}{\partial x} + \frac{\partial y}{\partial t}\bullet\frac{\partial I(x,y,t)}{\partial y} + \frac{\partial I(x,y,t)}{\partial t} = 0 \tag{2.3}$$

הידוע כ - "Optical flow Constraint Equation"

מתוך משוואה יש לחלץ את וקטור התנועה הדו-מימדי שרכיביו הם:

$$u = \frac{\partial I(x,y,t)}{\partial x}, v = \frac{\partial I(x,y,t)}{\partial y}$$

מכיוון שלמשוואה (2.3) אין פתרון יחיד הוצעו מספר אילוצי חלקות נוספים על וקטורי התנועה על מנת לפתור את בעיית תת – אילוצים זו, לדוגמא [HoS81], [KTB87].

השיטות לשערוך Optical Flow תקפות רק בתנאים מסויימים, שהחשובים בהם:

1. על רמות הבהירות בין התמונות להשתנות באיטיות רבה, כך שניתן להניח שבהירות האלמנטים בין שתי תמונות קבועה.
2. קיימת תנועה קטנה בין שתי תמונות סמוכות (כך שקירוב טור טיילור של (2.2) יהיה מדוייק).

לשיטות אלו מספר חסרונות:

1. ההנחות לגבי קביעות בהירות התמונה ותנועות קטנות בין מסגרות סמוכות אינן מדוייקות בהכרח.
2. על מנת לפתור את בעיית תת-האילוצים של (2.3) יש להשתמש בהנחות על חלקות ה - Optical Flow . הנחות אלו אינן תקפות כאשר קיימים אי רציפויות, לדוגמא סביב גבולות עצם נע.

מתוך שדה התנועה הוקטורי הצפוף ניתן לחשב בתסריטים מסויימים פרמטרי תנועה גלובלים.

2.3.3 שערוך התמרה המרחבית מבוססת קורלציה[Br92]

קורלציה משמשת בעיקר כמדד לחישוב הצלחת ההתאמה בין שתי תמונות. כאשר ההתמרה המרחבית הינה הזזה בלבד ניתן לשערך את ההתמרה המרחבית באמצעות מציאת ההזזה שתניב קורלצייה מקסימלית בין שתי התמונות $T(X,Y)$ ו- $I(X,Y)$.

$$C(u,v)=\frac{\sum_x\sum_y(T(x,y)-\mu_T)(I(x-u,y-v)-\mu_I)}{\sqrt{\sum_x\sum_y(T(x,y)-\mu_T))^2\sum_x\sum_y(I(x-u,y-v)-\mu_I)^2}} \qquad (2.4)$$

$\mu_T = mean(T(X,Y))$ ו- $\mu_I = mean(I(X,Y))$.

באמצעות אי-שוויון קושי-שוורץ ניתן להראות בקלות ש- $C(u,v)$ מקבל את ערכו המקסימלי כאשר התבנית מתאימה בדיוק לתמונה (עד כדי קבוע נרמול). באמצעות חיפוש אחר כל ההזזות האפשריות ניתן למצוא את ההזזה שתביא למקסימום את $C(u,v)$. החיסרון של שיטות אלו נובע מרגישותם לשינויים "איטיים" שבין התמונות לדוגמא, שינויי תאורה (יש לציין חסינות גבוהה יחסית לרעש לבן אדיטיבי) וסיבוכיות חישובית גבוהה. שיטה זו הינה הבסיס לשערוך תנועה בשיטת Block – Matching, בה לכל בלוק בתמונה הראשונה מחושב וקטור תנועה הטוב ביותר (לדוגמא, במובן של קורלציה מקסימלית בין הבלוק המוזז בתמונה הראשונה לבלוק בתמונה השנייה).

תחום החיפוש תלוי רק בכח החישוב , אולם מכיוון ששיטת Block – Matching נפוצה במקודדים של רצפי תמונות קימים רכיבי חומרה יעודיים למציאת וקטורי התאמת הבלוקים בזמן – אמת.

2.3.4 שערוך התמרה מרחבית באמצעות התמרת פורייה

שיטות מבוססות התמרת פורייה מתאימות למקרים בהם התמונה לאחר הטרנספורמציה קשורה בתחום התדר בדרך מתמטית פשוטה לתמונה לפני התנועה.

השיטות המוצעות במישור התדר מבוססות על מציאת Phase Correlation [KuH75] (מכיוון שתנועה במישור המקום מתבטאת בשינוי פאזה במישור התדר) ולכן מניבות חסינות גבוהה לרעשים בתדרים מסויימים (כגון שינוי תאורה, האיטיים בד"כ, ומרוכזים בפסי בתדר הנמוך) אולם לא עמידות לרעש לבן (על פני כל הספקטרום).

בהנתן תמונה $f_1(x,y)$ המועתקת לפי $f_2(x,y)=f_1(x-d_x,y-d_y)$ אזי הקשר בין שתי התמרות הפורייה מוגדר כ:

$$F_2(\omega_x,\omega_y)=e^{-j(\omega_x d_x+\omega_y d_y)}F_1(\omega_x,\omega_y) \qquad (2.5)$$

כאשר התמרת פורייה של התמונה $f_1(x,y)$ הנה $F_1(\omega_x,\omega_y)$ והתמרת פורייה של התמונה $f_2(x,y)$ הנה $F_2(\omega_x,\omega_y)$.

באמצעות (2.5) ניתן לקבל

$$\frac{F_1(\omega_x,\omega_y)\bar{F}_2(\omega_x,\omega_y)}{|F_2(\omega_x,\omega_y)\bar{F}_2(\omega_x,\omega_y)|}=e^{j(\omega_x d_x+\omega_y d_y)} \tag{2.6}$$

כאשר $\bar{F}_2(\omega_x,\omega_y)$ הנו הצמוד הקומפלקסי של $F_2(\omega_x,\omega_y)$.

באמצעות התמרת פורייה ההפכית:

$$F^{-1}\{e^{j(\omega_x d_x+\omega_y d_y)}\}=\delta(x+d_x,y+d_y) \tag{2.7}$$

ניתן למצוא בקלות את וקטור התנועה באמצעות מיקום ההלם.

מכיוון שהפרש הפאזה בין שתי התמונות קבוע בכל הספקטרום אזי כאשר בין שתי התמונות נוסף רעש צבעוני(המרוכז בפס תדר צר) ניתן למצוא בדיוק טוב את וקטור ההזזה. כאשר בין התמונות נוסף רעש לבן (בכל הספקטרום), אזי שיטה מבוססת תדר זו לא תצליח לשערך את וקטור ההזזה במדויק. במקרים אלו שיטת הקורלצייה (2.3.3) תניב תוצאות טובות יותר משום שמציאת הקורלציה הנה סינון עם מסנן (מסננת מתואמת) באותו תחום תדרים של האות הרצוי – תוך הקטנת השפעת הרעש בשאר התדרים.

שיטה זו מאפשר מציאת וקטור הזזה בלבד. עבור התאמת תמונות הכוללת גם סיבוב, מוצע ב – [DeM87] להשתמש בתכונת האינוארינטיות לסיבוב של התמרת פורייה.

השיטה המוצעת כוללת חישוב התמרת פורייה של שתי התמונות בקואורדינטות כדוריות (r,θ) ומציאת הזווית ϕ שתניב:

$$F^{-1}\{G(r,\theta,\phi)\}=F^{-1}\{\frac{F_1(r,\theta)\bar{F}_2(r,\theta-\phi)}{|F_2(r,\theta)\bar{F}_2(r,\theta-\phi)|}\}\approx\delta(x+d_x,y+d_y) \tag{2.8}$$

מציאת הזווית ϕ מבוצעת באמצעות חישוב $F^{-1}\{G(r,\theta,\phi)\}$ עבור מספר ערכי ϕ (עד לקבלת תוצאה קרובה ככל האפשר להלם בודד $F^{-1}\{G(r,\theta,\phi)\}\approx\delta(x+d_x,y+d_y)$) לאחר שנמצאה זווית זו ההזזה הינה מיקום ההלם המתקבל.

2.3.5 שערוך התמרה מרחבית באמצעות התאמת נקודות עניין

כל השיטות שהוצגו עד כה מאפשרות לבצע התאמת תמונות עבור מספר מוגבל בלבד של התמרות מרחביות (סיבוב והזזה בלבד – שלוש דרגות חופש בלבד). כאשר איזור קטן יחסית בתמונה משתנה משמעותית, השיטות הגלובליות (שיטות 2.3.3 ו – 2.3.4) שהוצגו

לעיל, עלולות לא לפעול, אפילו בסיטואציות של תנועות קטנות. אילוצי השיטה לשערוך optical flow (2.3.2) כוללים הזזה קטנה בין תמונות עוקבות ותאורה קבועה. באמצעות מציאת מספר נקודות עניין (נקודות בולטות) בין שתי התמונות ניתן לשערך התמרות מרחביות בעלות דרגות חופש רבות יותר ולהזניח שינויים מקומיים בין שתי התמונות.

כאשר בסצינה חלו שינויים מקומיים (לדוגמא, כביש שצולם בתמונת אינפרא-אדום בשעות שונות של היום הינו בעל טמפרטורה שונה ולכן נראה בצורה אחרת בין שתי התמונות). ניתן לבצע התאמה גלובלית ואחריה ביצוע התאמה מקומית עבור האיזורים שההתאמה הגלובלית לא הצליחה בהם. מכיוון שנקודות העניין נמצאות בכל התמונה כולה, אזי שיטה זו רגישה פחות לשינויים מקומיים בתוך התמונה.

כמו – כן חשוב להדגיש שהעומס החישובי של התאמה גלובלית בין שתי התמונות עצמן (כלומר עבור כל פיקסלי התמונה) גדול בהרבה מהעומס החישובי הנובע מחיפוש התאמה בין אוסף קטן יחסית של נקודות עניין.

מציאת ההתאמה מחולקת לשלושה חלקים עיקריים:

1. מציאת נקודות בולטות בתמונה הראשונה (נקודות עניין).
2. מציאת התאמה בין נקודות העניין בתמונה הראשונה לנקודות בתמונה השנייה.
3. שימוש בנקודות המותאמות בשתי התמונות לחישוב ההתמרה מרחבית (פונקציית ההתאמה).

נקודות עניין יכולות להיות אמיתיות או מלאכותיות. נקודות ענין מלאכותיות בתמונה הינם סמנים בולטים שהוצבו בסצינה לצורך התאמת תמונה נוחה ומדוייקת יותר. בתמונות רפואיות סמנים מלאכותיים מושמים על החולה במקומות ידועים מראש כנקודות ייחוס. בתמונות MRI (Magnetic Resonance Imaging) מושמות בנקודות ידועות מראש בגוף החולה צינוריות בצורת האות "N" המכילים $CuSo_4$ [Ev88].

נקודות עניין אמיתיות יכולות להבחר בצורה אוטומטית או ידנית. לבחירת נקודות ידנית באמצעות צופה אנושי מספר יתרונות עקב היכולת לבחור נקודות עניין בולטות וברורות הקיימות בשתי התמונות. כאשר יש לבחור מספר רב של נקודות העניין או שקיימות תמונות רבות, בחירת נקודות עניין בצורה ידנית אינה מעשית. לפיכך, בחירת נקודות עניין בצורה אוטומטית נפוצה בשימושים רבים. נקודות עניין נבחרות כך שיוכלו להתאימן בצורה ייחודית בין שתי התמונות. מכיוון שבאמצעות נקודות התאמה מחושבת הטרנספורמציה המרחבית אזי יש לבחור די נקודות עניין, אם כי בחירת נקודות עניין רבות מדי עלולה לגרום קשיים בביצוע ההתאמה בין הנקודות (לדוגמא, מכיוון שנבחר מספר רב של נקודות, נבחרו נקודות לא בולטות מספיק).

על נקודות העקיבה הנבחרות להיות חסינות יחסית לשינויים בין התמונות. רגישות לשינויים כאלה לא תאפשר מציאת התאמה נכונה.

ניתן ורצוי לבצע עיבוד מקדים (pre-processing) על התמונות כך שייקל למצוא נקודות עניין ולבצע התאמה טובה. קיימות מספר של שיטות לשיפור מיצוי נקודות ההתאמה, ביניהן:

1. סינון רעש תוך שמירה על השפות בתמונה [MaS88].
2. שיטות למתיחת תמונה (הגדלת הטווח הדינמי של התמונה).
3. שיטת לחידוד תמונה (שיטות להדגשת שפות) [MoY88].

שיטת העיבוד המקדים נבחרת עפ"י סוג נקודת העניין שיש למצוא.

ניתן לבחור נקודות עניין מסוגים שונים, לדוגמא:

1. התאמה פשוטה בין עוצמות התמונה – השוואה בין רמות האפור של איזורים כלשהם בתמונה.
2. קווים בולטים בתמונה כגון שפות [ErG85] ועקומים [MeN84].
3. נקודות בולטות כגון פינות או מרכזי איזורים סגורים [Go86].

פרק 3 מכיל סקירה קצרה של מספר קריטריונים לבחירת נקודות עניין.

לאחר שנבחרו נקודות עניין, מבוצעת ההתאמה בין נקודות העניין בשתי התמונות. ההתאמה מבוצעת בד"כ באמצעות מציאת ההזזה שתתאים בצורה הטובה ביותר (לדוגמא, באמצעות מציאת ההזזה שתניב קורלצייה מקסימלית) כל אחת מנקודות העניין בתמונה הראשונה לתמונה השנייה. ההתאמה בין נקודות העניין מאפשרת שערוך ההתמרה המרחבית בין התמונות.

להלן מופיעים מספר סוגי התמרות מרחביות שניתן לשערך באמצעות התאמת נקודות עניין.

א. התמרה פרספקטיבית

כאשר שתי תמונות נלקחות מזוויות שונות, אזי ההתמרה המרחבית המקשרת ביניהן הינה ההטלה הפרספקטיבית. ההטלה הפרספקטיבית הינה השלכת סצינה תלת-מימדית קשיחה כלשהיא על מישור המצלמה (תחת מודל מצלמה pin-hole). מכיוון שלא ידועים עומקי התמונה (דהיינו מרחקי כל נקודה בסצינה מהמצלמה) לא ניתן לשערך התמרה זו בצורה מדוייקת ויש להשתמש בקירובים גלובלים שונים (כגון התמרה מישורית או התמרה אפינית בעלות שמונה או שש דרגות חופש בהתאמה). בפרק 4 יוסברו בפרוטרוט ההתמרות המרחביות השונות הנובעות מצילום התמונות משתי נקודות ראייה וכיצד ניתן לשערך את פרמטרי ההתמרות המקורבות.

ב. התמרה פולינומיאלית

כאשר חוסר-ההתאמה בין התמונות לא נובע אך ורק מההבדלים בין זוויות הצילום, אלא גם מעיוותים נוספים כגון עיוותי המצלמה ניתן להשתמש בהתמרה מרחבית פולינומיאלית [Go88]:

$$X' = \sum_{i=0}^{m}\sum_{j=0}^{i} a_{ij} X^i Y^{i-j}$$

(2.9)

$$Y' = \sum_{i=0}^{m}\sum_{j=0}^{i} b_{ij} X^i Y^{i-j}$$

(X,Y) הינן הקואורדינטות בתמונה הראשונה המועתקות לקואורדינטות (X',Y') בתמונה השנייה באמצעות ההתמרה הפולינומיאלית המוגדרת באמצעות $(a_{ij}, b_{ij}),\ 0 \le i \le m, 0 \le j \le i$.

סדר הפולינום m הינו פשרה בין הדיוק הדרוש לייצוג ההתאמה לבין אילוצי סיבוכיות.

$2N$ מקדמי הפולינום משוערכים באמצעות מציאת לפחות $N = 0.5(m+1)(m+2)$ התאמות בין נקודות עניין בשתי התמונות.

קיימות שתי גישות לשערוך מקדמי הפולינום - שיטת הקירוב או אינטרפולציה.

- בשיטת הקירוב מחפשים את הטרנספורמצייה האופטימלית המקשרת בצורה האופטימלית (לדוגמא, במובן של מציאת מינימום של שגיאה ריבועית ממוצעת) את נקודות העניין בין שתי התמונות. שיטת הקירוב מבוססת על ההנחה שלהתאמה בין הנקודות נוסף רעש מקומי – שלא ניתן להסירו באמצעות שערוך הטרנספורמצייה. השערוך מבוצע באמצעות שיטות כגון LS (Least Squares Estimation) ומחייב מציאת מספר נקודות רב יותר מאשר N על מנת לאפשר שערוך אמין.
- שיטת האינטרפולצייה מבוססת על מציאת ההתמרה המדוייקת בין N נקודות ההתאמה (ואינטרפולצייה להתמרת שאר איזורי התמונה).

שיטת ההתמרה הפולינומיאלית טובה בד"כ לתיאור שינויים איטיים בין התמונות (כלומר, שימוש בפולינומים מסדר נמוך בלבד) עקב חוסר יציבות נומרית כאשר סדר הפולינום גבוה.

2.4 שערוך התאמה מרחבית מקומית

מכיוון שמציאת התאמה גלובלית בין שתי תמונות לא יכולה לפצות על עיוותים גיאומטריים מקומיים כגון אי – לינאריות של החיישן, תנאים אטמוספריים או תנועת עצמים שאינם קשיחים (deformable, כגון עננים) , מתעורר לעיתים הצורך לשערך סט שונה של פרמטרי התאמה מרחבית לכל איזור בתמונה - באמצעות קבוצות סמוכות של נקודות עניין.

השיטות המקובלות להתאמות מקומיות מבוססות על אינטרפולציה למקוטעין (piecewise interpolation). כלומר, בהנתן סט נקודות התאמה יש לשערך שתי פונקציות משטח S_x ו - S_y המקרבות

$$X'_i = S_x(X_i, Y_i)$$
$$Y'_i = S_y(X_i, Y_i) \qquad (2.10)$$

$(X_i, Y_i) \leftrightarrow (X'_i, Y'_i),\ i = 0,..,N-1$ סט נקודות ההתאמה.

שיטות כלליות לחישוב פונקציית אינטרפולציה מקומית (בהתבסס על מספר סופי של נקודות שכנות) אינן מבטיחות רציפות או קיום נגזרות. כאשר מעוניינים בקבלת פונקצייה חלקה (רציפה ובעלת נגזרות) ניתן לשערך את פונקציות המשטח בשיטת splines. כלומר מניחים שפונקציית המשטח נפרשת ע"י מספר בסיסים. מקדמי הבסיסים נבחרים לפי אילוצים המחייבים קיום תכונות "גלובליות" כגון רציפות וגזירות בין איזורים בעלי פונקציית משטח שונה [Fr79].

יש לשים לב למספר נקודות:
א. יש לבצע סגמנטציה של נקודות העניין בתמונה למספר קבוצות, כך שכל קבוצה תתואר באמצעות סט פרמטרים אחד.
א. המשקל החישובי הכבד של שערוך סט התאמות מקומיות ביחס לשערוך סט אחד של פרמטרי התאמה גלובלית.

ואריאציה של השיטה המקומית מתוארת ב [Fl92]. לפי שיטה זו מחפשים בשלב ראשון פונקציית התאמה גלובלית בין שתי התמונות. לכל איזור שלא הותאם היטב בהתמרה גלובלית זו מחושבת בנפרד פונקציית התאמה מקומית. התהליך של התאמת פרמטרי כל איזור חוזר על עצמו עד להתאמה טובה דייה של התמונה כולה.

2.5 סיכום

בפרק זה הוצג תיאור קצר של אלגוריתמים להתאמת תמונות באמצעות שערוך ההתמרה המרחבית המתאימה. קיימות מספר שיטות לשערוך ההתאמה והשיטות הכלליות ביותר הינן השיטות לשערוך ה - Optical Flow והתאמת נקודות עניין.

הפרק הבא (פרק 3) מציג בפרוטרוט באלו תנאים השיטה השנייה עדיפה. פרק זה מכיל סקר ספרות בנושא של מיצוי נקודות עניין ומציג את השיטות למציאת נקודות העניין ששימשו לצורך שערוך פרמטרי הטרנספורמציות המרחביות.

פרק 3 מציאת נקודות התאמה

אחת השיטות העיקריות לשערוך התמרה מרחבית הינה מציאת התאמה בין מספר מצומצם של מאפיינים בולטים בין שתי התמונות, ובאמצעות התאמה זו ניתן לשערך את ההתמרה המרחבית בין שתי התמונות. השיטה מסתמכת על התכונה שנקודות אלו בולטות דיים כך שניתן למצוא את תנועתן בדרגת אמינות גבוהה. שיטות אלו חסינות יותר לרעשים מאשר שיטות הגראדיינטים ויכולות להשיג מעקב אחר תנועות מהירות יותר. חסרונות שיטה זו נובעים מהקושי למצוא התאמה בין הנקודות המיוחדות ומהדרישה שימצאו מספיק נקודות מיוחדות כאלה.

על נקודות ההתאמה הנבחרות להיות חסינות יחסית לשינויים בין התמונות. רגישות לשינויים כאלה לא תאפשר מציאת התאמה נכונה. לדוגמא, לא ניתן למצוא התאמה בהתבסס על מציאת התאמה בין מומנטים הוואריאנטים לסיבוב בשתי התמונות כאשר השינוי בין התמונות מורכב מסיבוב (כאשר התנועה היחסית בין התמונות הינה הזזה שיטה זו תפעל).

פרק זה מכיל סקר ספרות של מספר שיטות למציאת נקודות התאמה. בנוסף מתוארת שיטה משופרת למציאת נקודות התאמה המניבה שיפור בשערוך ההתמרה המרחבית בין שתי תמונות.

3.1 סקר ספרות - מציאת נקודות התאמה

ניתן להשתמש בקווים ישרים כמאפיינים בולטים . ניתן להשתמש בהתמרת Hough לצורך מציאת קווים ישרים [Ja89] בתוך התמונה באמצעות מיקום נקודות שפה בתוך מרחב הפרמטרים של כל הישרים ומציאת ערכי הפרמטרים שלהם מתאים מספר רב של נקודות שפה. לאחר מציאת הקווים ניתן למצוא פינות – הצטלבויות של קווים.
אחת מחסרונות השיטה שקווים ארוכים ממסכים קווים קצרים ולכן מוצע ב - [CrG94] לחלק את התמונה לבלוקים (patches) ולמצוא קווים בכל בלוק בנפרד (ומתוך הישרים למצוא פינות).

התאמה של קווים ישרים הינה בעייתית מכיוון שקשה למצוא במדוייק את ההתאמה בין שפות. נקודות התאמה מתאימות יותר הינן פינות. סקירה מקיפה של אופרטורים לזיהוי מדוייק של פינות מופיע ב – [DeG90].

ב - [ErG85] מציעים להשתמש בחציות – אפס (נקודות בהן סימני ערכי האות מתחלפים) לאחר העברת מסנן "לפלסיאן של גאוסייאן" ($\nabla^2 G(x,y) = \left[\frac{x^2+y^2}{\sigma^2} - 2\right]\exp\left\{-\frac{(x^2+y^2)}{2\sigma^2}\right\}$) כנקודות התאמה. מסנן זה ניתן למימוש כהפרש של שני מסננים גאוסיאנים ומהווה מסנן מדגיש שפות. ההתאמה בין שתי התמונות מבוצעת בין שתי תמונות השפות (לאחר הפעלת המסנן) באיזורים של מקסימום מקומי.

ניתן להתאים עקומים כנקודות התאמה (כפי שמוצע ב – [Na92]). לפי המחבר ניתן לחבר את נקודות השפה בתמונה (לאחר מעבר של מסנן "לפלסיאן של גאוסייאן") לעקומים. ההתאמה מבוצעת באמצעות התאמת המאפיינים של כל עקום בשתי התמונות. המאפיינים המותאמים הינם אורך העקום ורדיוסו.

קיימות מספר שיטות למיצוי נקודות התאמה באמצעות תמונת הפרטים של התמרת ה – Wavelet. [MoC96] מציעים לקחת כנקודות עקיבה את המקדמים בעלי העוצמה הגדולה ביותר בפסי התדר HL ו- LH.

לפי [Hs96] ו - [Hs97] נקודות התאמה הנבחרות הינן נקודות של מקסימום מקומי בתמונת הפרטים הגדולות מסף. כמו כן יש לדרוש שמכפלת תמונות הפרטים בשתי רמות רזולוציה בנקודות אלו גדולות מתמונת הפרטים ברזולוציה הנמוכה ביותר על מנת להבטיח חסינות מפני רעש.

ב - [FDFS96] ו - [FDS98] מציעים למצוא מתוך תמונת הפרטים קוי שפה. עבור כל נקודת שפה ניתן לחשב את עוצמת התמרת ה – Wavelet כפונקציה של ה – scale.לטענת המחברים ניתן לקבוע מהו סוג השפה (כלומר האם הנקודה הינה שפה אמיתית, שיא בולט או קו בתמונה עצמה) לפי שיפוע פונקציה זו.

שיטות לפירוק מולטי-רזולוציה יכולות לשמש למציאת פרטים בגדלים מסויימים. [Wi97], [ZhC93] מציעים לזהות בתמונות הפרטים איזורים בעלי אנרגייה רבה. [ZhC93] מציע לסנן את התמונות לפני הפירוק במסנן המדגיש שפות, בעוד שב - [Wi97] מוצע קריטריון מבוסס אנטרופיה למציאת איזורים בעלי אנרגיה גבוהה.

ב – [Ch92a] מוצע להשתמש בנקודות השפה בתמונה (המתקבלות באחת השיטות הסטנדרטיות, לדוגמא מסיכת Sobel לגילוי שפות [Ja89]) כנקודת התחלה למציאת מאפיינים מוצלחים. לאחר קבלת תמונת שפות מבוצע תהליך סיווג של כל נקודות השפה לנקודות פינה, צומת, קווים ישרים וכו' בהתבסס על שכניהם. בתהליך של גידול איזורים (Region Growing) מאוחדות נקודות סמוכות מאותו סוג ולכל מקטע שכזה מבוצע קירוב באמצעות התאמה של conic curves (פולינומים מסדר שני). נקודות החיתוך של פולינומים אלו הנן נקודות התאמה הנבחרות.

עקב העומס החישובי הכרוך במציאת נקודות עניין בתמונה כולה, מוצע ב - [Ch92b] לחלק את התהליך לשני חלקים: א. חלוקת מפת השפות של התמונה לבלוקים ובאמצעות המאפיינים הסטטיסטיים של כל בלוק להגדיר האם הבלוק הינו בלוק אחיד (ללא שינויים חדים), בלוק המכיל בעיקר קווים או בלוק המכיל שפות (שפה בתוך בלוק הינה קו שרק חלקו נמצא בתוך הבלוק). ב. רק על הבלוקים המכילים שינויים מופעלים אלגוריתמים מורכבים יותר (בהתאם לסוג הבלוק) למציאת נקודות עניין.

3.2 אלגוריתמים למיצוי נקודות התאמה

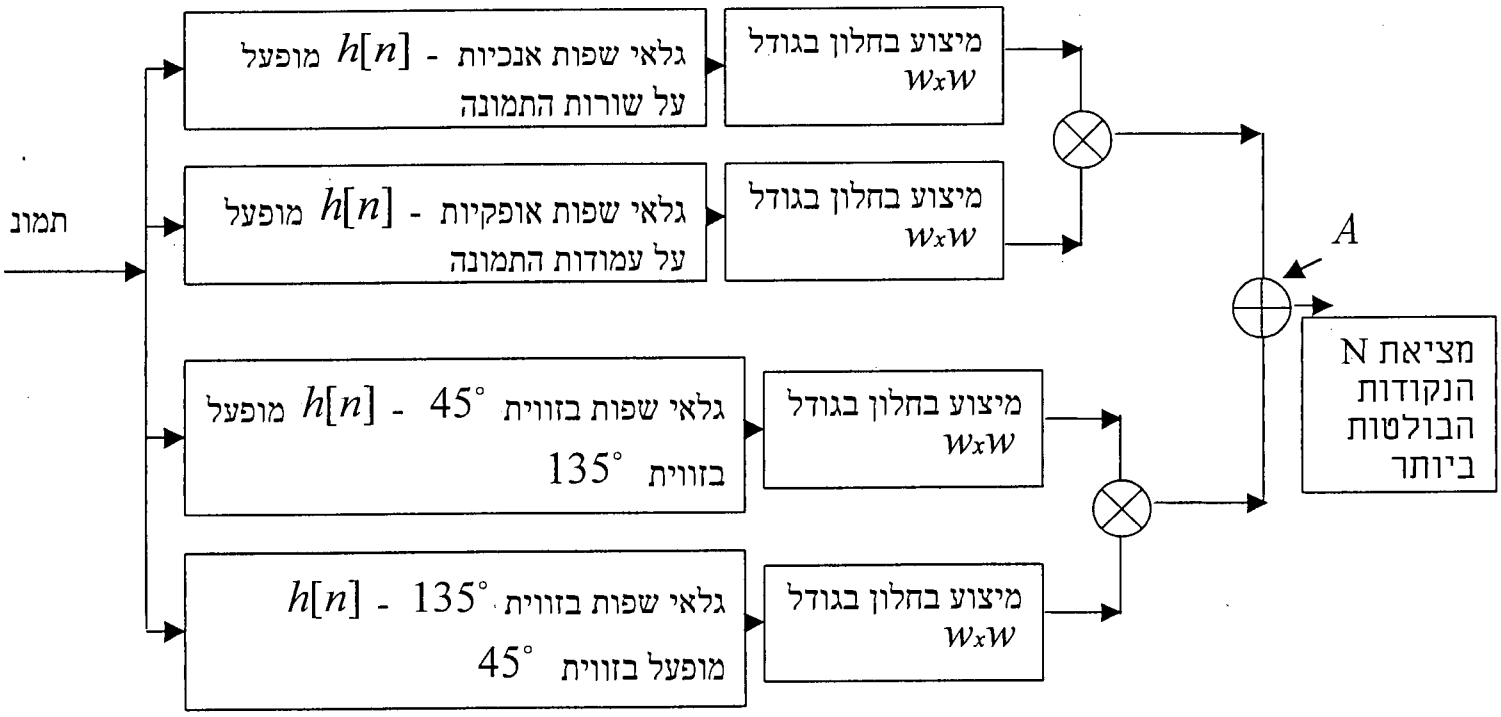
בעבודה זו נבדקו והושוו שני אלגוריתמים למיצוי נקודות בולטות בתמונה : אלגוריתם המבוסס על גראדיינטים ("שיטת גראדיינטים"). אלגוריתם זה פשוט יחסית ובעל תכונה שחלק מהנקודות הנבחרות על-ידו אינן טובות דיין לעקיבה (ולכן מחייב שימוש במספר רב של נקודות עניין על-מנת לשערך את התנועה) . האלגוריתם השני ("שיטת ערכים עצמיים") מורכב יותר ומניב נקודות התאמה טובות יותר.

3.2.1 אלגוריתם I ("שיטת גראדיינטים")

מטרת המערכת המתוארת כאן הינה למצוא את N הנקודות המתאימות ביותר למציאת התאמה בין שתי תמונות. אלגוריתם זה מבוסס על מציאת מכפלת נגזרות מקסימליות ופועל כדלקמן:

על התמונה מופעל גלאי שפות חד-מימדי מהצורה:

$$h[n] = n \cdot \exp(-n^2)\,,\; n = -3,..,3 \qquad (3.1)$$



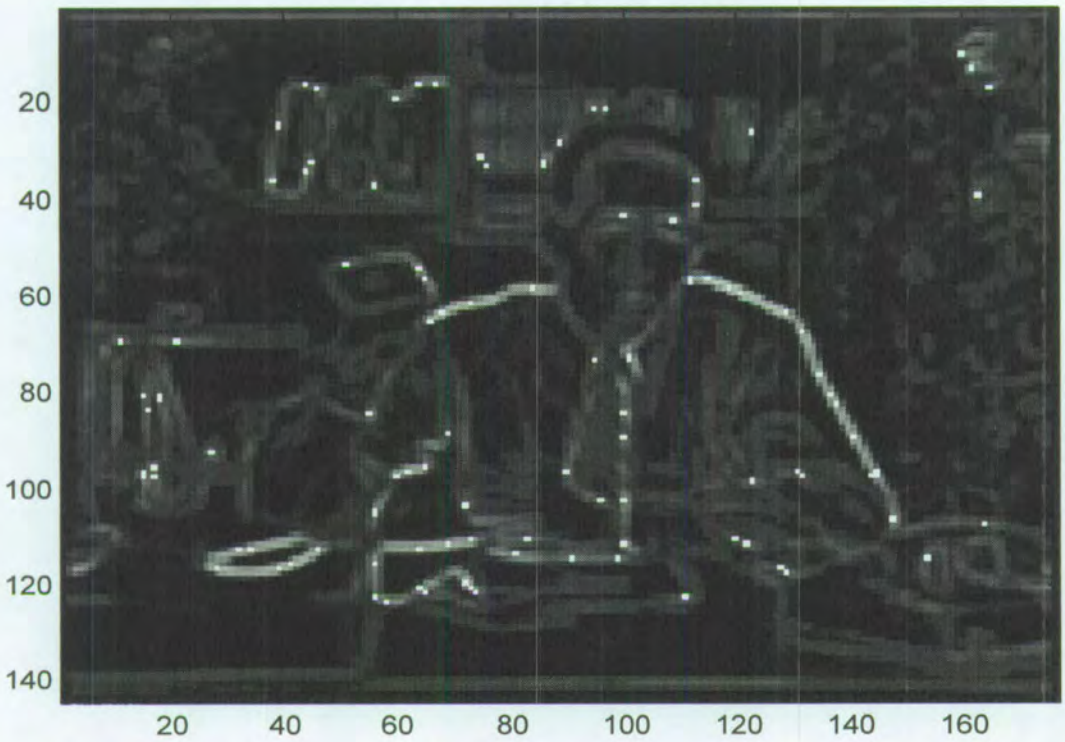
איור 3.1 – אלגוריתם I ("שיטת גראדיינטים") למציאת נקודות התאמה

Fig. 3.1: Algorithm I ("Gradient Algorithm") for finding matching points

הקריטריון לבחירת הנקודות איננו טוב מספיק מכיוון שלדוגמה, יכולה להתקבל פינה שרק אחת מצלעותיה בעלת נגזרות חזקות בעוד הצלע השנייה בעלת נגזרות חלשות. במקרה זה מכפלת הנגזרות עלולה להניב, לכאורה, נקודה עניין טובה בעוד שהנקודה אינה מתאימה להתאמה מכיוון שהצלע "החלשה" לא מאפשרת להנעל במדוייק על מיקום הפינה.

להלן מופיעה דוגמא למוצא האלגוריתם בנקודה A והנקודות הבולטות:

The edge detector's output and the chosen best points



איור 3.2 - מוצא אלגוריתם I למיצוי נקודות התאמה - "שיטת גראדיאנטים"
Fig. 3.2: The output of Algorithm I ("Gradient Algorithm")

איור 3.2 מציג את תפוקת אלגוריתם I כאשר מספר נקודות העקיבה הנדרשות גדול ($N = 60$). כאשר N קטן לא נבחרות נקודות עקיבה טובות דיין לצורכי שערוך ההתאמה.

3.2.2 אלגוריתם II ("שיטת הערכים העצמיים")

אלגוריתם זה מבוסס על מציאת קריטריון מתימטי לאיכות נקודת ההתאמה כתלות בנגזרות המרחביות בסביבת הנקודה. ניתן לקבל קריטריון זה באמצעות מציאת התנאים לשערוך מדוייק של תנועת הנקודה בין שתי תמונות. כלומר, נקודה נחשבת לנקודת התאמה טובה אם ניתן לחשב את תנועתה במדוייק ככל האפשר. להלן מופיעים שני אילוצים שעל נקודות העקיבה לקיים המתבטאים בקריטריון טיב זהה.

א. אילוץ המבוסס על עקרונות שערוך [HoS81] Optical Flow

אובייקט הנע במהירות $\vec{v} = (v_x, v_y)$ מקיים

$$I(\vec{x},t) = I(\vec{x} - \vec{v}t, 0) \tag{3.2}$$

כאשר I הינה עוצמת התמונה ו - $\vec{x} = (x,y)^T$.

מתוך פיתוח טיילור של (3.2) או מתוך הקירוב הקובע שבזמן $t \to 0$ השינוי באנרגיית התמונה זניח ולכן גזירה של (3.2) לפי הזמן t נותנת

$$\dot{I} = \frac{\partial I(\vec{x},t)}{\partial t} = \frac{\partial I(\vec{x} - vt, 0)}{\partial \vec{x}} \cdot \frac{\partial \vec{x}}{\partial t} + \frac{\partial I(\vec{x} - vt, 0)}{\partial t} = 0 \tag{3.3}$$

הידוע כמשוואת אילוץ התנועה:

$$\nabla I(\vec{x},t) \cdot \vec{v} + I_t(\vec{x},t) = 0 \tag{3.4}$$

כאשר $\nabla I(\vec{x},t) = (I_x(\vec{x},t), I_y(\vec{x},t))$, $\nabla I \cdot \vec{v}$ הינה מכפלה סקלרית, ו - I_t הינה נגזרת חלקית של $I(\vec{x},t)$ לפי הזמן.

כאשר מנסים לשערך את מהירות העצם תוך שימוש במשוואת אילוץ התנועה עבור בלוק X הנע במהירות $\vec{v}$, יש למזער את

$$\sum_X W^2(\vec{x})[\nabla I(\vec{x},t) \cdot \vec{v} + I_t(\vec{x},t)]^2 \to \min \tag{3.5}$$

$W(\vec{x})$ הינה פונקציית חלון (פונקצית משקל) כך שיתאפשר למזער את (3.5) בעיקר באיזור מרכז הנקודות של X .

הפתרון המקבל באמצעות גזירה של (3.5) לפי $\vec{v}$ והשוואה לאפס זהה לפתרון LS עבור אוסף הנקודות X :

$$A^T W^2 A \vec{v} = A^T W^2 b \tag{3.6}$$

$$A = [\nabla I(\vec{x}_1), \ldots, \nabla I(\vec{x}_N)]^T$$
$$W = diag(W(\vec{x}_1), \ldots, W(\vec{x}_N))$$
$$b = -[I_t(\vec{x}_1), \ldots, I_t(\vec{x}_N)]^T$$
$$\vec{x} = (x,y)^T$$

הפתרון ל – (3.6) הינו

$$\vec{v} = (A^T W^2 A)^{-1} A^T W^2 b \tag{3.7}$$

על מנת לקבל פתרון יציב מבחינה נומרית, על המטריצה

$$Q = A^T W^2 A = \begin{bmatrix} \sum W(\mathrm{x}) I_x^2(\mathrm{x}) & \sum W(\mathrm{x}) I_x(\mathrm{x}) I_y(\mathrm{x}) \\ \sum W(\mathrm{x}) I_x(\mathrm{x}) I_y(\mathrm{x}) & \sum W(\mathrm{x}) I_y^2(\mathrm{x}) \end{bmatrix} \tag{3.8}$$

להיות הפיכה עם condition number קטן ככל האפשר, כאשר רצוי

$$\text{condition_number} = \frac{\lambda_{\max}}{\lambda_{\min}} \to 1 \tag{3.9}$$

כמו כן, על מנת לשערך היטב את וקטור המהירות, על שני הערכים העצמיים להיות גדולים ככל האפשר (הערכים העצמיים מעידים על כמות השינויים ועוצמתם, כפי שהוכח ב – (3.13).

דרך נוספת לקבלת (3.8) נובעת מתוך מיזעור השגיאה הריבועית הממוצעת בין בלוקים עוקבים - ככל שנקודת העקיבה טובה יותר, הבלוק המכיל את הנקודה עשיר בפרטים ייחודיים המאפשרים מציאת התאמה מדוייקת יותר ובעלת שגיאה קטנה יותר.

ב. אילוץ המבוסס על עקרונות Block Matching [ShT94]

שגיאת השערוך של בלוק B בתמונה J כאשר נמצא בלוק תואם בתמונה I בהיסט $\vec{d} = (d_x, d_y)^T$ הינה:

$$\varepsilon^2 = \sum_B [(I(\vec{\mathrm{x}} - \vec{d}) - J(\vec{\mathrm{x}}))]^2 w^2(\vec{\mathrm{x}}) \tag{3.10}$$

$$\vec{\mathrm{x}} = (x, y)^T$$

פיתוח טיילור של $I(\vec{\mathrm{x}} - \vec{d})$ סביב ל- $\vec{\mathrm{x}}$ נותן

$$I(\vec{\mathrm{x}} - d) = I(\vec{\mathrm{x}}) - d_x \frac{\partial I(\vec{\mathrm{x}})}{\partial x} - d_y \frac{\partial I(\mathrm{x})}{\partial y} \tag{3.11}$$

הצבה לתוך (3.10) וגזירת השגיאה לפי $\vec{d}$ נותנת:

$$\frac{\partial \varepsilon^2}{\partial d} = -2 \sum_B (I(\vec{\mathrm{x}}) - J(\vec{\mathrm{x}}) - g(\vec{\mathrm{x}})^T \vec{d}) g(\vec{\mathrm{x}}) w^2(\vec{\mathrm{x}}) \tag{3.12}$$

$$g(\vec{\mathrm{x}}) = \begin{bmatrix} \frac{\partial I(\vec{\mathrm{x}})}{\partial x} & \frac{\partial I(\vec{\mathrm{x}})}{\partial y} \end{bmatrix}^T$$

השוואת (3.12) לאפס (מציאת $\vec{d}$ לקבלת שגיאה מינימלית) נותנת:

$$\sum_B (I(\vec{x}) - J(\vec{x}))g(\vec{x})w^2(\vec{x}) = \sum_B g(\vec{x})\vec{d}g(\vec{x})^T w^2(\vec{x}) = [\sum_B g(\vec{x})w^2(\vec{x})g(\vec{x})^T]\vec{d} \tag{3.13}$$

פתרון יציב עבור $\vec{d}$ יתקבל כאשר $\sum_B g(x)w^2(\vec{x})g(x)^T$ יציבה – דהיינו תנאים זהים ליציבות (3.8), כלומר קיום (3.9) ועוצמת הערכים העצמיים של (3.8).

- הבנה אינטואיטיבית של משמעות הערכים העצמיים של המטריצה Q (3.8):

(3.14)

$$\lambda_{1,2} = \frac{\sum w^2(\vec{x})(I_x^2(x) + I_y^2(x)) \pm \sqrt{(\sum w^2(\vec{x})(I_x^2(x) + I_y^2(x)))^2 - 4(\sum w^2(\vec{x})I_x^2(x)\sum w^2(\vec{x})I_y^2(x) - \sum w^2(\vec{x})I_x(x)I_y(x))}}{2}$$

$$= \frac{\sum w^2(\vec{x})(I_x^2(x) + I_y^2(x)) \pm \sqrt{(\sum w^2(\vec{x})(I_x^2(x) - I_y^2(x)))^2 + 4\sum w^2(\vec{x})I_x(x)I_y(x))}}{2}$$

כפי שניתן לראות, התנאי המחייב ערכים עצמיים גדולים ככל האפשר נובע מתוך הדרישה ש $\sum w^2(\vec{x})(I_x^2(x) + I_y^2(x))$ יהיה גדול ככל האפשר. כלומר, כמות הנגזרות המרחביות בחלון גדולה ככל האפשר.

הקטנת ה – condition number מתקבלת כאשר

$$(\sum w^2(\vec{x})(I_x^2(x) - I_y^2(x)))^2 + 4\sum w^2(\vec{x})I_x(x)I_y(x)) \approx 0$$

כלומר:

$\sum w^2(\vec{x})I_x^2(x) = \sum w^2(\vec{x})I_y^2(x)$ (סכום עוצמת הנגזרות בשתי הצירים זהה)

$\sum w^2(\vec{x})I_x(x)I_y(x) \approx 0$ (הנגזרות האופקיות והאנכיות אורתוגונליות זו לזו)

אלו הן שני תנאים מתאימים למציאת פינות.

- קריטריון כמותי לאיכות נקודות עקיבה

כזכור, התנאים ליציבות Q (וליכולת לשערך את וקטור ההזזה בצורה מדוייקת) הינם ערכים עצמיים גדולים ככל האפשר , כך שהיחס ביניהם שואף ליחידה. על מנת להשוות נקודות בצורה פשוטה ויעילה יש למפות את שני התנאים הללו לציון מספרי יחיד.

א. ערך עצמי מינימלי[SoE98]:

מיפוי הערך העצמי המינימלי לציון מתבצע כך שידגיש הפרשים קטנים כאשר הערך העצמי קטן ויזניח הפרשים קטנים (מבחינת הציון) כאשר הערך העצמי המינימלי הינו גדול דיו. המיפוי שנבחר הינו לתחום $[0-10]$ והינו לוגריתמי בתחום הביניים.

$$m_1 = \begin{bmatrix} 10 & \lambda > t_2 \\ 10\dfrac{\ln(\lambda)-\ln(t_1)}{\ln(t_2)-\ln(t_1)} & t_2 \geq \lambda \geq t_1 \\ 0 & \lambda < t_1 \end{bmatrix} \qquad (3.15)$$

$$\lambda = \min(\lambda_2, \lambda_1)$$

ב. Condition Number מינימלי[SoE98]:

ככל שה - $\text{cond} = \lambda_1 / \lambda_2$ קרוב יותר ליחידה, Q יציבה יותר ולכן ימופה ציון גבוה יותר ל - cond נמוך יותר. כאשר cond גדול מסף מסויים t_3, הציון m_2 עבור נקודה זו הינו אפס.

$$m_2 = \begin{bmatrix} 10\dfrac{t_3+1-cond}{t_3} & \text{cond} \leq t_3 \\ 0 & \text{cond} > t_3 \end{bmatrix} \qquad (3.16)$$

הציון הכולל של הנקודה הינו

$$m = \alpha \cdot m_1 + (1-\alpha)\cdot m_2 \qquad (3.17)$$

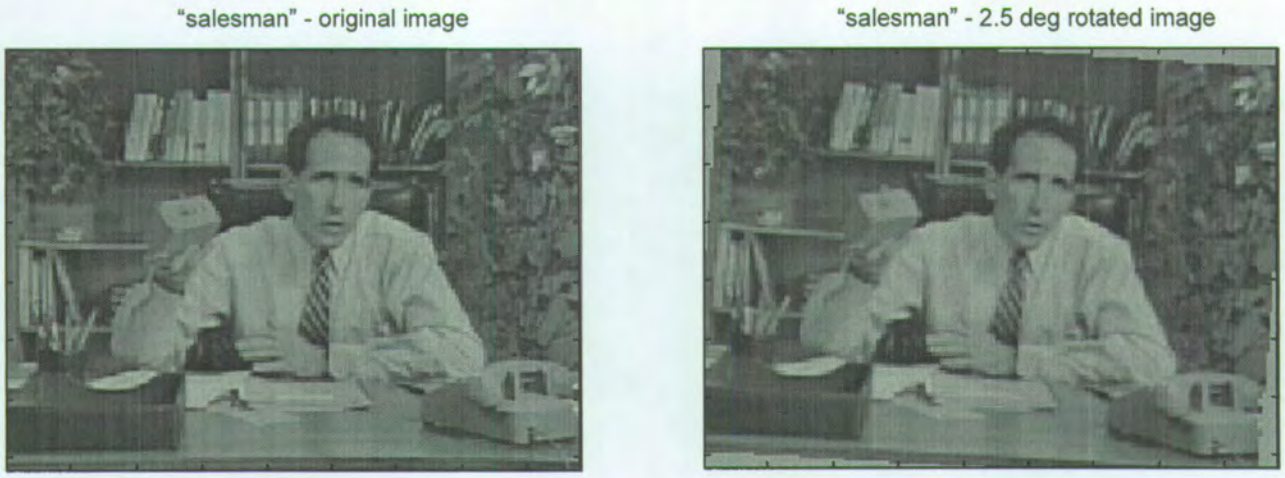
לסיכום, שלבי האלגוריתם הם:

א. חישוב הנגזרות הכיווניות בשני הצירים x ו- y, I_x, I_y.
ב. עבור כל נקודה (או נקודות פוטנציאליות בלבד) מחשבים את המטריצה Q סביב הנקודה (3.8).
ג. חישוב הערכים העצמיים של המטריצה Q וקביעת הציון m (3.17) בהתבסס על (3.16) ו – (3.15).
ד. בחירת הנקודות שבהן m מקסימלי כנקודות ההתאמה הנבחרות.

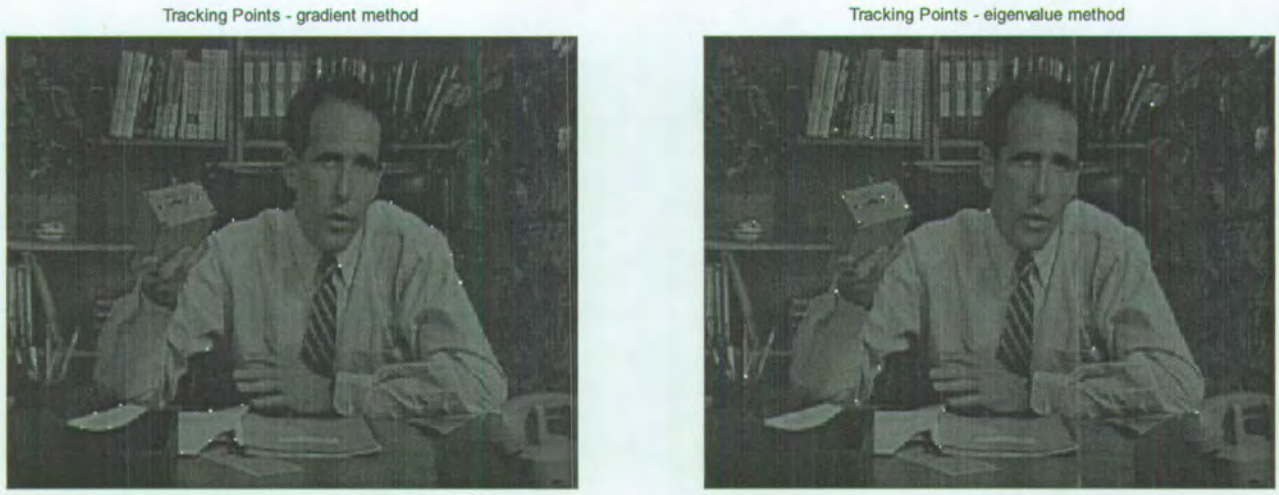
3.3 סימולצייה והשוואת ביצועים

עבור כל אחת מ - $N = 20$ נקודות התאמה המתאימות (ובלוק בגודל 8x8 סביב כל נקודה) מתבצע חיפוש בתמונה הבאה למציאת הבלוק בעל הקורלצייה הגבוהה ביותר. רק

פינות שעבורן קיים מקסימום בולט של ערך הקורלצייה מעל סף מסויים (דהיינו, לא קיימות שתי נקודות שיא קרובות של קורלצייה במסגרת הבאה מעל ערך הסף) נכללות בפתרון LS לשערוך פרמטרי ההתמרה הגלובלית (לפי השיטה המוצגת בפרק 4).
ביצועי שני האלגוריתמים הושוו על רצף סינטתי של תמונת "Salesman" הנעה במהירות זוויתית של 2.5° למסגרת (איור מס' 3.3)

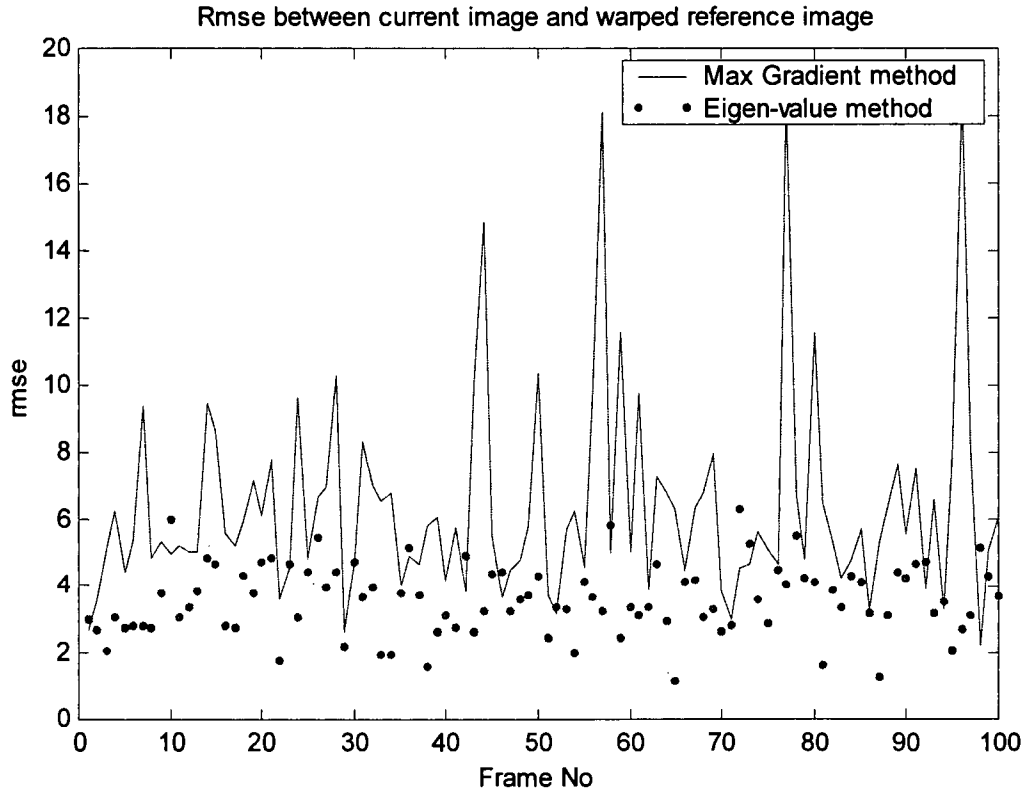


איור 3.3 – תמונת "Salesman" נעה במהירות זוויתית של 2.5° למסגרת
Fig. 3.3: The "Salesman" image rotating with angular velocity of 2.5° per frame



איור 3.4 – בחירת נקודות ההתאמה
Fig. 3.4: Choosing matching points

על מנת לבדוק את איכות נקודות ההתאמה יש לבדוק את ביצועי אלגוריתם שערוך התנועה ברצף התמונות. הפרק הבא יכלול הסבר מפורט של צורת שערוך התנועה ובשלב זה יובהר שנקודות העקיבה שימשו לשערוך פרמטרי הטרנספורמציה האפינית בעלת שישה פרמטרים באמצעות שיטת Least Squares. על מנת לבדוק את איכות השערוך מוצג להלן גרף (איור 3.5) המתאר את אנרגיית תמונת ההפרש בין כל תמונה ברצף לבין חיזוי התמונה באמצעות התמונה הראשונה ברצף ופרמטרי הטרנספורמציה האפינית ששוערכו באמצעות נקודות ההתאמה (פרק 4).



איור 3.5 – אנרגיית תמונת ההפרשים – שערוך תנועה באמצעות נקודות התאמה

Fig. 3.5: Energy of difference image – motion estimation based upon matching points

כפי שניתן לראות שערוך הטרנספורמציה האפינית באמצעות "שיטת הערכים העצמיים" מניבה תוצאות טובות יותר מ"שיטת הגראדיינטים".

פרק 4 קידוד רצף תמונות המבוסס על שערוך התמרה אפינית גלובלית

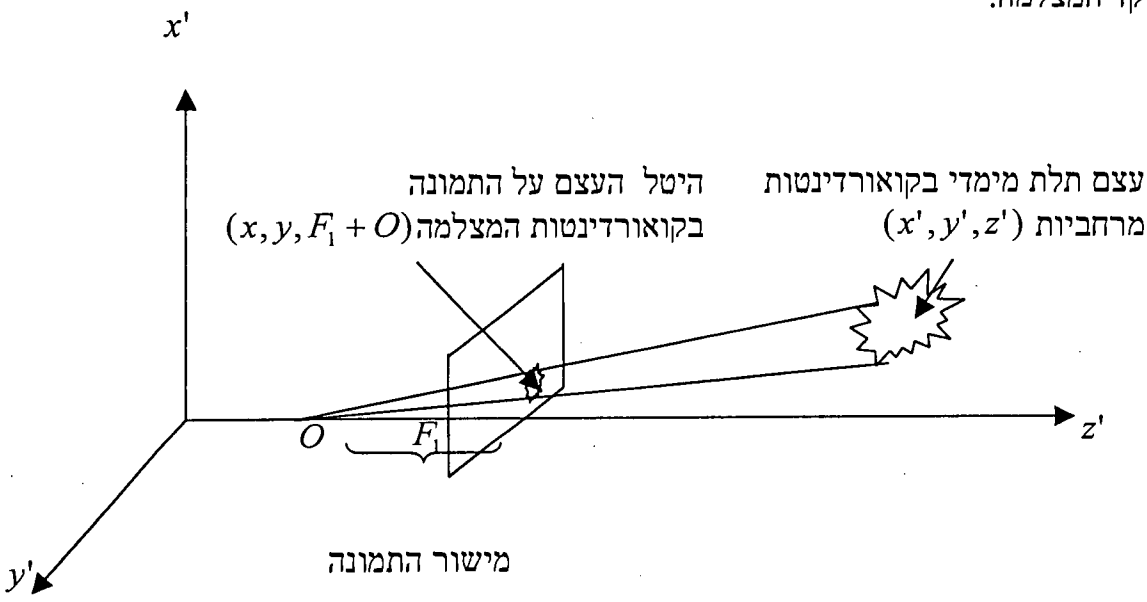
4.1 הקדמה

כאשר מצלמה נעה בתנועה תלת-מימדית במרחב, נקודות הסצינה משנות את מיקומן במישור המצלמה. לעיתים ניתן לייצג את השינוי בהיטל הנקודה על מישור התמונה באמצעות סט פרמטרים הקבוע לכל נקודות הסצינה. סט פרמטרים זה הינו ההתמרה המרחבית הנגרמת עקב התנועה.

בפרק זה יתוארו מספר מודלי תנועה גלובלית הנובעים מתנועה תלת-מימדית של המצלמה במרחב. לאחר מכן תוצע שיטה לקידוד יעיל של רצפי תמונות המתקבלים באמצעות מצלמה נעה. השיטה מבוססת על שידור של התמונה הראשונה ברצף (כתמונת ייחוס) ועדכון התנועה הגלובלית בין כל תמונה ברצף לבין תמונת הייחוס.

חשוב לציין שתנועת אובייקט (הנע בתנועה תלת-מימדית) על מישור מצלמה סטטית זהה עקרונית לתנועה המתקבלת באמצעות צילום של סצינה סטטית ע"י מצלמה נעה.

הטלה פרספקטיבית מתארת היווצרות תמונה באמצעות מצלמת חריר (pin-hole) אידיאלית לפי עקרונות האופטיקה הלינארית. כלומר, כל הקרניים מהעצם עוברות דרך מוקד המצלמה.



איור 4.1 - מצלמת חריר – pin-hole
Fig. 4.1: Pin-hole camera model

באמצעות דמיון משולשים ניתן להראות שכאשר הציר האופטי של המצלמה מתלכד עם ציר z' מתקיים:

$$\frac{x}{F_1} = \frac{x'}{z'-O}; \quad \frac{y}{F_1} = \frac{y'}{z'-O} \tag{4.1}$$

או:

$$x = \frac{x' F_1}{z'-O}; \quad y = \frac{y' F_1}{z'-O} \tag{4.2}$$

כאשר מוקד המצלמה נמצא בראשית מערכת הצירים מתקבל הביטוי הידוע עבור מודל מצלמה פרספקטיבי (pinhole):

$$x = F_1 \frac{x'}{z'} \quad y = F_1 \frac{y'}{z'}, \quad z = F_1 \tag{4.3}$$

מכיוון ש (4.3) אינו לינארי, ניתן להשתמש בקואורדינטות הומוגניות, על מנת למדל את המצלמה כמודל לינארי:

$$\begin{pmatrix} kx \\ ky \\ k \end{pmatrix} = \begin{pmatrix} F_1 & 0 & 0 & 0 \\ 0 & F_1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} lx' \\ ly' \\ lz' \\ l \end{pmatrix} \tag{4.4}$$

כאשר $(kx, ky, k)^T, (lx', ly', lz', l)^T$ מסמנים קואורדינטות הומוגניות במערכת צירי הייחוס ובמערכת הצירים של המצלמה, בהתאמה.

משוואה (4.4) מעתיקה קואורדינטה הומוגנית במערכת צירי הייחוס לקואורדינטה הומוגנית במערכת הצירים של המצלמה. אין משמעות פיזיקלית אמיתית לכל אחת מהקואורדינטות ההומוגניות בפני עצמן, אלא רק כמנה של כל אחת והקואורדינטה ההומוגנית התחתונה.

משוואה (4.3) איננה לינארית וקירוב ידוע [TrV98] שלה הינו מודל מצלמה weak-perspective. במודל זה המרחק היחסי לאורך הציר האופטי בין כל שתי נקודות בסצינה, dz, קטן מאוד ביחס למרחק הממוצע של נקודות הסצינה מהמצלמה.

במקרה זה (4.3) הופך ל –

$$\begin{aligned} x &= F_1 \frac{x'}{z'} \approx F_1 \frac{x'}{\bar{z}'} \\ y &= F_1 \frac{y'}{z'} \approx F_1 \frac{y'}{\bar{z}'} \end{aligned} \tag{4.5}$$

כאשר $\bar{z}$ הינו המרחק הממוצע של הסצינה מהמצלמה. לפי [Trv98] קירוב weak-perspective תקף כאשר $dz / \bar{z} < 0.05$.

קירוב של (4.5) הינו orthographic projection בו הקרניים האופטיות מקבילות לציר האופטי:

$$x = x'$$
$$y = y' \qquad (4.6)$$

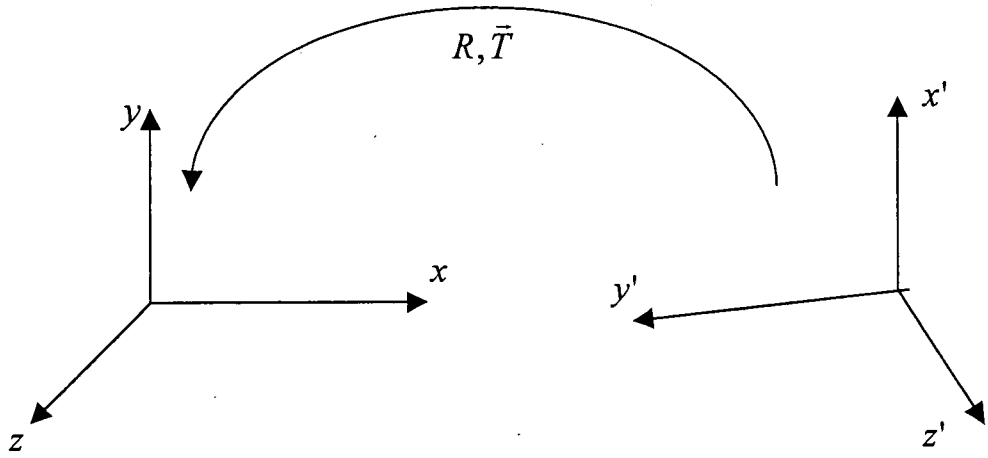
מבחינה מתימטית orthographic projection הינו קירוב של הטלה פרספקטיבית עבור $F_1 \to \infty$, אזי $z' \to \infty$ ו - $F_1 / z' \to 1$ [Trv98].

בפרק זה תוגדר נקודה (x', y', z') במערכת צירי הייחוס השקולה לקואורדינטה (x, y, F_1) במערכת הצירים של המצלמה. נקודה זו הינה נקודה (X, Y) בפיקסלים.

4.2 תנועת מצלמה

משוואה (4.3) מתארת את הקשר בין מערכת ייחוס כללית (x', y', z'), הידועה כ – world reference frame לבין מערכת הייחוס של המצלמה (x, y, z). באיור 1 ידוע במדוייק הקשר בין (x', y', z') ו - (x, y, z) - עקב התלכדות ציר z והציר האופטי של המצלמה.
במקרה הכללי ביותר קיימים מספר פרמטרים המקשרים בין מערכת הייחוס (world reference frame) לבין מערכת הצירים של המצלמה מצד אחד, ומצד שני הם מקשרים בין קואורדינטות מערכת הצירים של התמונה לבין הפיקסלים עצמם, שהינם הגדלים המדידים היחידים מהתמונות עצמן.

הפרמטרים החיצוניים (extrinsic) מתארים את ההתמרה בין מערכת הייחוס לבין מערכת הצירים של המצלמה ומורכבים מווקטור הזזה תלת-מימדי $\vec{T} = (dx, dy, dz)^T$ המתאר את ההזזה בין מרכזי שתי מערכות הצירים ומטריצת סיבוב תלת-מימדית R_{3x3} המתאימה את שתי מערכות הצירים:



איור 4.2 - הקשר בין מערכת הצירים של המצלמה לבין מערכת צירי הייחוס
Fig 4.2: Orientation between camera's coordinate system and world reference coordinate system

כפי שניתן לראות בקלות מאיור 4.2:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = R \cdot (\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} - \vec{T}) \tag{4.7}$$

הפרמטרים הפנימיים (intrinsic) מתארים את המאפיינים האופטיים והגיאומטריים של המצלמה. עבור מודל מצלמת חריר (pinhole) קיימים שלוש קבוצות של פרמטרים:

א. אורך המוקד F_1.

ב. ההתמרה בין קואורדינטות המצלמה לפיקסלי התמונה.

ג. העיוותים הגאומטריים הנובעים מהמכשור האופטי.

כאשר מניחים שהתמונה מתקבלת באמצעות מערך חיישנים דוגמים ריבועיים (כגון תמונת CCD), אזי

$$\begin{aligned} x &= -(X - o_x)s_x \\ y &= -(Y - o_y)s_y \end{aligned} \tag{4.8}$$

כאשר:

(X,Y) הינם קואורדינטות (x,y) בפיקסלים.

(o_x,o_y) הינו הקואורדינטה בפיקסלים של מרכז התמונה – נקודות החיתוך של הציר האופטי עם מישור התמונה.

(s_x,s_y) הינו הגודל האפקטיבי של פיקסל (ביחידות אורך).

היפוך הסימן נובע משום שמערכת הצירים של המצלמה (איור 4.2) הפוכה למערכת הצירים בתמונת הפיקסלים עצמה (בה הכיוון החיובי הינו ימינה ולמטה).

בהזנחת העיוותים האופטיים, סט הפרמטרים הפנימיים הינם (F_1,o_x,o_y,s_x,s_y).

לפי (4.8), (4.7) ו – (4.3) ניתן לראות ש:

$$\begin{aligned} x &= -(X - o_x)s_x = F_1 \frac{R_1^T(P_w - \vec{T})}{R_3^T(P_w - \vec{T})} \\ y &= -(Y - o_y)s_y = F_1 \frac{R_2^T(P_w - \vec{T})}{R_3^T(P_w - \vec{T})} \end{aligned} \tag{4.9}$$

כאשר

$P_w = (x',y',z')^T$ ו - R_i הינה השורה ה - i של R.

את (4.9) ניתן לרשום באמצעות מכפלת מטריצות M_{int} ו - M_{ext}

$$M_{\text{int}} = \begin{pmatrix} -F_1 / s_x & 0 & o_x \\ 0 & -F_1 / s_y & o_y \\ 0 & 0 & 1 \end{pmatrix}$$

(4.10)

$$M_{ext} = \begin{pmatrix} r_{11} r_{12} r_{13} & -R_1^T \cdot \vec{T} \\ r_{21} r_{22} r_{23} & -R_2^T \cdot \vec{T} \\ r_{31} r_{32} r_{33} & -R_3^T \cdot \vec{T} \end{pmatrix}$$

$r_{ij},\ 1 \le i, j \le 3$ הינם איברי המטריצה R.

ובאמצעות הוספת k כקואורדינטה הרביעית של P_w מקבלים את הקואורדינטה ההומוגנית של P_w המניבה התמרה לינארית

(4.11)
$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = M_{\text{int}} M_{ext} \begin{pmatrix} kx' \\ ky' \\ kz' \\ k \end{pmatrix} = M \begin{pmatrix} kx' \\ ky' \\ kz' \\ k \end{pmatrix}$$

השויון נכון עד כדי קבוע נרמול כלשהו k.

באמצעות (4.9) ו – (4.11) מקבלים ש -

(4.12)
$$X = \frac{x_1}{x_3}$$
$$Y = \frac{x_2}{x_3}$$

כאשר (X, Y) הינם הקואורדינטות (x, y) בפיקסלים.

לשם פשטות הפיתוח, נניח ש - $o_x = o_y = 0,\ s_x = s_y = 1$ ניתן אז לקבל את המודל הפרספקטיבי הכללי הבא של המצלמה:

$$M = M_{\text{int}} M_{ext} = \begin{pmatrix} -F_1 r_{11} & -F_1 r_{12} & -F_1 r_{13} & F_1 R_1^T \cdot \vec{T} \\ -F_1 r_{21} & -F_1 r_{22} & -F_1 r_{23} & F_1 R_2^T \cdot \vec{T} \\ r_{31} & r_{32} & r_{33} & -R_3^T \cdot \vec{T} \end{pmatrix} \tag{4.13}$$

התמרה פרספקטיבית

כפי שניתן לראות מתוך (4.9) ו – (4.12)

$$\begin{aligned} X &= \frac{x_1}{x_3} = \frac{-F_1 r_{11} x' - F_1 r_{12} y' - F_1 r_{13} z' + F_1 R_1^T \cdot \vec{T}}{r_{31} x' + r_{32} y' + r_{33} z' - R_3^T \cdot \vec{T}} \\ Y &= \frac{x_2}{x_3} = \frac{-F_1 r_{21} x' - F_1 r_{22} y' - F_1 r_{23} z' + F_1 R_2^T \cdot \vec{T}}{r_{31} x' + r_{32} y' + r_{33} z' - R_3^T \cdot \vec{T}} \end{aligned} \tag{4.14}$$

באמצעות נרמול ב - z', שימוש ב – (4.3) וההנחה ש - $o_x = o_y = 0, s_x = s_y = 1$ (כלומר קואורדינטות מערכת הצירים של המצלמה שוות לערכי הקואורדינטות בפיקסלים $(X', Y') = (x', y')$). מקבלים:

$$\begin{aligned} X &= F_1 \frac{-r_{11} X' - r_{12} Y' - F_1 r_{13} + R_1^T \cdot \vec{T} / z'}{r_{31} X' + r_{32} Y' + r_{33} - R_3^T \cdot \vec{T} / z'} \\ Y &= F_1 \frac{-r_{21} X' - r_{22} Y' - F_1 r_{23} + R_2^T \cdot \vec{T} / z'}{r_{31} X' + r_{32} Y' + r_{33} - R_3^T \cdot \vec{T} / z'} \end{aligned} \tag{4.15}$$

כפי שניתן לראות, ההעתקה בין כל זוג נקודות (X,Y) ו - (X',Y') תלויה במרחק z' ממוקד המצלמה.

התמרה של גוף קשיח

כאשר מעוניינים בקירוב מסוג weak – perspective, כלומר

$$z' \approx z = R_3^T((x' \ y' \ z')^T - \vec{T}) \equiv \bar{z} \tag{4.16}$$

ובאמצעות (4.3):

$$X' = F_1 \frac{x'}{z'} = F_1 \frac{x'}{\bar{z}}; \quad Y' = F_1 \frac{y'}{z'} = F_1 \frac{y'}{\bar{z}} \tag{4.17}$$

כזכור, $o_x = o_y = 0,\ s_x = s_y = 1$, ולכן קואורדינטות במישור התמונה בפיקסלים שוות לקואורדינטות האמיתיות של התמונה.

מקבלים:

$$
\begin{aligned}
X &= \frac{x_1}{x_3} = \frac{-F_1 R_1^T (P_w - \vec{T})}{R_3^T (P_w - \vec{T})} = \frac{-F_1 R_1^T (P_w - \vec{T})}{\bar{z}} \\
&= \frac{F_1 R_1^T \vec{T}}{\bar{z}} - F_1 R_1^T \left(\frac{x'}{\bar{z}}, \frac{y'}{\bar{z}}, 1\right)^T = \Delta_x - r_{11} X' - r_{12} Y' \\
Y &= \frac{x_2}{x_3} = \frac{-F_1 R_2^T (P_w - \vec{T})}{R_3^T (P_w - \vec{T})} = \frac{-F_1 R_2^T (P_w - \vec{T})}{\bar{z}} \\
&= \frac{F_1 R_2^T \vec{T}}{\bar{z}} - F_1 R_2^T \left(\frac{x'}{\bar{z}}, \frac{y'}{\bar{z}}, 1\right)^T = \Delta_y - r_{21} X' - r_{22} Y'
\end{aligned}
\tag{4.18}
$$

המניב את הביטוי המקובל להתמרה של גוף קשיח - העצמים בתמונה שומרים על צורתם וגודלם היחסי. התמרה זו מורכבת מסיבוב, הזזה ו – zoom זהה בשני הצירים:

$$
\begin{pmatrix} X' \\ Y' \end{pmatrix} = sR \begin{pmatrix} X \\ Y \end{pmatrix} + T' = s \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} \Delta_x \\ \Delta_y \end{pmatrix}
\tag{4.19}
$$

X ו - Y הינם הקואורדינטות (בפיקסלים) בתמונה הראשונה המועתקות לקואורדינטות X' ו - Y' בתמונה השנייה (בפיקסלים).
כפי שניתן לראות שהיחס הגיאומטרי בין הנקודות השונות, כלומר משולש בתמונה הראשונה ממופה למשולש דומה בתמונה השנייה - מכיוון שמטריצת הסיבוב R הינה אורתוגונלית אזי אורכי קווים וזוויות יחסיות בין קווים נשמרים.

במקרה כזה מודל המצלמה הינו (4.11):

$$
M = \begin{pmatrix} -F_1 r_{11} & -F_1 r_{12} & -F_1 r_{13} & F_1 R_1^T \cdot \vec{T} \\ -F_1 r_{21} & -F_1 r_{22} & -F_1 r_{23} & F_1 R_2^T \cdot \vec{T} \\ 0 & 0 & 0 & \bar{z} \end{pmatrix}
\tag{4.20}
$$

התמרה אפינית

בסצינות מסויימות קירוב מסוג weak – perspective איננו מתאים . על מנת לפצות על חוסר הדיוק, ניתן להוסיף עוד שתי דרגות חופש ל - (4.19). כלומר לא קיים אילוץ כלשהו על שתי השורות הראשונות של M . מודל תנועה זה הינו המודל האפיני [Te95].

פיתוח דומה ל - (4.19) מניב את איברי ההתמרה האפינית:

$$\begin{pmatrix} X' \\ Y' \end{pmatrix} = A\begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} dx \\ dy \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix} + \begin{pmatrix} \Delta_x \\ \Delta_y \end{pmatrix} \tag{4.21}$$

ההתמרה האפינית הינה מקרה כללי של התמרה של גוף קשיח ולכן מאפשרת תיאור תנועות מורכבות יותר. בהתמרה זו קווים מקבילים בתמונה הראשונה מועתקים לקווים מקבילים בתמונה השנייה.

ההתמרה האפינית הכללית ניתנת לתיאור "פיזיקלי" באמצעות פירוק SVD של A :

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{pmatrix}\begin{pmatrix} s_x & 0 \\ 0 & s_y \end{pmatrix}\begin{pmatrix} \cos\theta_2 & -\sin\theta_2 \\ \sin\theta_2 & \cos\theta_2 \end{pmatrix} \tag{4.22}$$

כלומר, התמרה אפינית מורכב מסיבוב בזווית θ_1, הפעלת zoom שונה בשני הצירים (s_x, s_y), סיבוב בזווית θ_2 והזזה דו-מימדית.

הן לפי מודל מצלמה אפיני או weak – perspective לא ניתן לשערך את מרחק העצם מהמצלמה מכיוון שזוהי הטלה מקבילה למישור התמונה.

4.3 שיטת ריבועים פחותים (Least Squares) לשערוך פרמטרי תנועה גלובלית

שיטת Least Squares היא שיטה המאפשרת התאמת פרמטרי מודל למדידות כך שהשגיאה הריבועית הממוצעת בין תחזית המודל למדידות הינו מינימלי. לשם ביצוע שיטה זו צריך לבחור אפריורית מודל למדידות, וערך פרמטרי המודל נמצאים בעזרת אלגוריתם המינימיזציה. באופן עקרוני אין הגבלות על המודל הפרמטרי, אך השיטה שתתואר בהמשך מתאימה לטיפול במודלים לינאריים בלבד.

במודל הלינארי סט המדידות $\vec{d}$ הינו פונקציה לינארית של סט הפרמטרים $\vec{a}$, כלומר

$$M\vec{a} = \vec{d} \tag{4.23}$$

כאשר M היא מטריצה ידועה מראש, ו - $\vec{a}, \vec{d}$ הם וקטורים. במערכת המשוואות (4.23) מספר המדידות (מספר המשוואות) יכול להיות (א) קטן (ב) שווה או (ג) גדול ממספר מקדמי המודל (מספר הנעלמים). שני המקרים הראשונים אינם מעניינים מכיוון שעבורם קימים אינסוף פתרונות אפשרים שנותנים אותה מידת שגיאה, או שקיים מודל דטרמיניסטי יחיד, בהתאמה. המקרה המעניין יותר הוא המקרה שבו מספר המדידות גדול ממספר מקדמי המודל. מערכת כזו של משוואות מכונה overdetermined system.

במערכת כזו כשאין התאמה מלאה של המדידות למודל ניתן להניח רעש אדיטיבי ומערכת המשוואות תירשם בצורה:

$$M\vec{a} + \vec{n} = \vec{d} \tag{4.24}$$

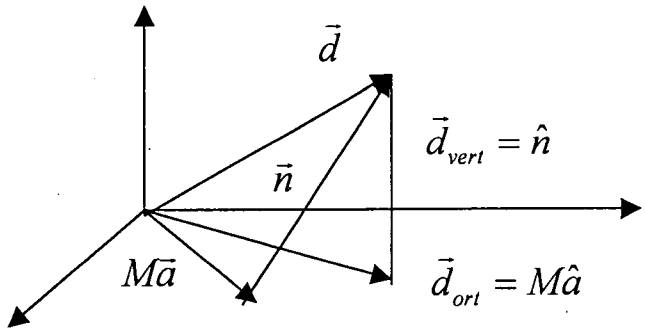
כאשר $\vec{n}$ הוא וקטור השגיאה במדידות. השגיאה הריבועית הינה:

$$e^2 = \vec{n}^T \cdot \vec{n} = (\vec{d} - M\vec{a}) \cdot (\vec{d} - M\vec{a})^T \tag{4.25}$$

והיא צריכה להיות קטנה ככל האפשר. הפתרון של בעיית ה – Least Squares המביא למינימיזציה של פונקציית המחיר הינו [GoV80]:

$$\hat{a} = (M^T M)^{-1} M^T \vec{d} \tag{4.26}$$

המשמעות הגיאומטרית מאחורי ביטוי זה היא שהרעש המשוערך $\hat{n}$ מתקבל מתוך רכיב המדידות $\vec{d}$ המאונך למרחב העמודות של המטריצה M (יסומן כ - $\vec{d}_{vert}$), והפרמטרים המשוערכים מתקבלים מתוך היטל וקטור המדידות אל מרחב העמודות של המטריצה M (יסומן כ - $\vec{d}_{ort}$), כך ש - $\hat{n} = \vec{d}_{vert}$, $M\hat{a} = \vec{d}_{ort}$



איור 4.3 – גיאומטריה של שיטת LS
Fig 4.3: Least squares solution geometry

עבור מודלי תנועה שונים יתקבלו ביטויים שונים של $\vec{a}, \vec{d}$ ו - M.

עבור מודל תנועה של גוף קשיח (weak-perspective) בעל ארבע פרמטרי תנועה תתקבל מערכת המשוואות (4.19):

$$\underbrace{\begin{pmatrix} X_1 & -Y_1 & 1 & 0 \\ Y_1 & X_1 & 0 & 1 \\ \vdots & & & \\ X_N & -Y_N & 1 & 0 \\ Y_N & X_N & 0 & 1 \end{pmatrix}}_{M} \underbrace{\begin{pmatrix} s\cdot\cos\theta \\ s\cdot\sin\theta \\ \Delta_x \\ \Delta_y \end{pmatrix}}_{\vec{a}} = \underbrace{\begin{pmatrix} X_1' \\ Y_1' \\ \vdots \\ X_N' \\ Y_N' \end{pmatrix}}_{\vec{d}} \tag{4.27}$$

כאשר (X,Y) הן קואורדינטות נקודת התאמה בתמונה הראשונה (בפיקסלים) להן נמצאה התאמה מתאימה בתמונה השנייה (X',Y'). עבור מערכת משוואות זו (4.27) מספיק למצוא שתי התאמות בלבד.

שערוך הפרמטרים הפיזיקלים מתוך וקטור הפרמטרים המשוערך הינו

$$\begin{aligned} \hat{\theta} &= tg^{-1}\left(\frac{\hat{a}_2}{\hat{a}_1}\right) \\ \hat{s} &= \sqrt{\hat{a}_1^2 + \hat{a}_2^2} \end{aligned} \tag{4.28}$$

עבור מודל תנועה אפיני (שישה פרמטרים) מערכת המשוואות לפתרון באמצעות Least Squares הינה (4.21):

$$\underbrace{\begin{pmatrix} X_1 & Y_1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & X_1 & Y_1 & 1 \\ \vdots & & & & & \\ X_N & Y_N & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & X_N & Y_N & 1 \end{pmatrix}}_{M} \underbrace{\begin{pmatrix} a \\ b \\ \Delta_x \\ c \\ d \\ \Delta_y \end{pmatrix}}_{\vec{a}} = \underbrace{\begin{pmatrix} X_1' \\ Y_1' \\ \vdots \\ X_N' \\ Y_N' \end{pmatrix}}_{\vec{d}} \tag{4.29}$$

ניתן לראות שבמקרה זה מציאת הנעלמים a,b,Δ_x ומציאת הנעלמים c,d,Δ_y הינן בעיות בלתי – תלויות וניתן לפתור אותן בנפרד מתוך

$$\underbrace{\begin{pmatrix} X_1 & Y_1 & 1 \\ \vdots & & \\ X_N & Y_N & 1 \end{pmatrix}}_{M} \underbrace{\begin{pmatrix} a \\ b \\ \Delta_x \end{pmatrix}}_{\vec{a}} = \underbrace{\begin{pmatrix} X_1' \\ \vdots \\ X_N' \end{pmatrix}}_{\vec{d}}$$

$$\underbrace{\begin{pmatrix} X_1 & Y_1 & 1 \\ \vdots & & \\ X_N & Y_N & 1 \end{pmatrix}}_{M} \underbrace{\begin{pmatrix} c \\ d \\ \Delta_y \end{pmatrix}}_{\vec{a}} = \underbrace{\begin{pmatrix} Y_1' \\ \vdots \\ Y_N' \end{pmatrix}}_{\vec{d}} \qquad (4.30)$$

צורת חישוב זו עדיפה מכיוון שהיא דורשת היפוך של שתי מטריצות בגודל 3x3 במקום מטריצה אחת בגודל 6x6.

בסצינות גלובליות בהן הנחות מודל weak-perspective אינן מדוייקות, רצוי להשתמש במודל האפיני המאפשר דרגות חופש רבות יותר.

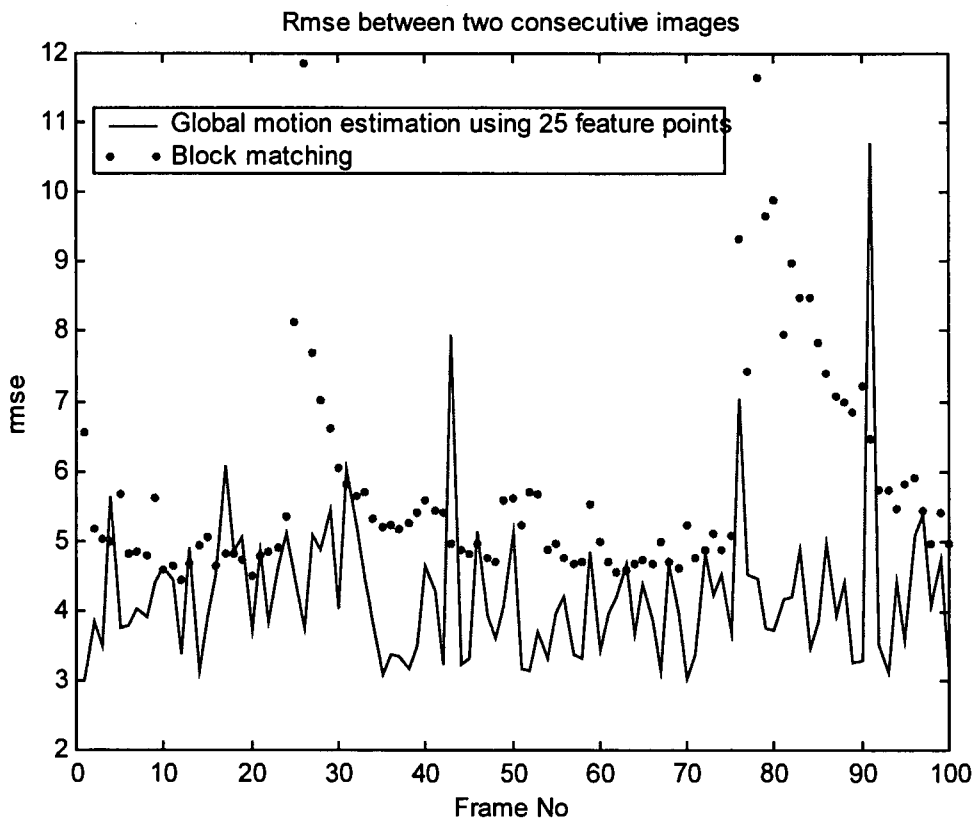
4.4 תוצאות שערוך תנועה גלובלית באמצעות פתרון בעיית Least Squares:

על מנת לבדוק את ביצועי אלגוריתם לשערוך תנועה גלובלית (בהסתמך על המודל האפיני) שוערכו פרמטרי התנועה הגלובלית באמצעות פתרון LS .

לכל תמונה ברצף "Salesman" המסובב במהירות של 2.5° למסגרת שוערכה התנועה בינה לבין התמונה הקודמת לה ברצף בשתי שיטות: האחת ע"י שערוך פרמטרי התמרה אפינית באמצעות שיטת LS והשנייה ע"י מציאת קיזוז תנועה בשיטת Block Matching (BM) עבור בלוקים בגודל 8x8. הגרף המצורף מציג את ה rmse בין כל תמונה לבין חיזוייה באמצעות שערוך התנועה לפי שתי השיטות. כאשר

$$rmse(k) = \sqrt{\frac{1}{\| A \|} \sum_{X,Y \in A} (I(X,Y)_k - \hat{I}(X,Y)_k)^2} \qquad (4.31)$$

וכאשר $\| A \|$ הינו שטח התמונה.



איור 4.4 - השוואה בין שיטות שונות לשערוך תנועה

Fig. 4.4: Comparison between different motion estimation methods

כפי שניתן לראות ביצועי שיטת BM גרועים יותר מהשיטה לשערוך פרמטרי מודל אפיניים. הדבר נובע משום ש – BM מסוגל למעשה למדל הזזה בלבד ובתמונות שנבחנו סיבוב מהיר בין התמונות.

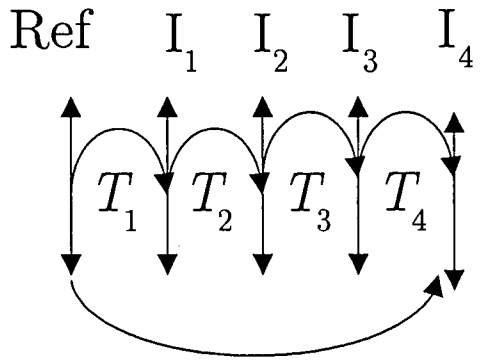
4.5 קידוד רצף תמונות באמצעות תמונת ייחוס ומידע התנועה

הדרישה הגוברת והולכת לקידוד רצפי תמונות מחייבות ייצוג יעיל של רצפי תמונות. לעיתים ניתן לייצג את רצף התמונות באמצעות שידור של תמונה ראשונה ברצף ואת מידע התנועה ליצירת כל תמונות הרצף. תקני ה – MPEG השונים, באופן עקרוני, משדרים עבור כל תמונה את העדכון בינה לבין התמונה הקודמת (או הבאה, ולכן יש לשנות את סדר שידור התמונות). בסצינות מסויימות כגון ברצף המתקבל ממצלמה נעה ניתן לשדר רק את התמונה הראשונה וסט קטן של פרמטרי תנועה גלובלית המתארים את השתנות הסצינה המצולמת בין התמונה הראשונה לכל אחת מתמונות הרצף.

השיטה הנפוצה לשערוך תנועה גלובלית מבוססת על מציאת נקודות התאמה בין שתי התמונות. מכיוון שקריטריון ההתאמה מבוסס על מציאת קורלציה מקסימלית בין נקודות ההתאמה, קשה לבצע התאמה ישירה בין שתי תמונות שביניהן הייתה תנועה רבה. יתכן שבין תמונת הייחוס לבין תמונות הרצף הייתה תנועה רבה ולכן מציאת התאמה ישירה בין תמונת הייחוס לתמונות הרצף אינה אפשרית. על מנת לשערך את התנועה קיימות שתי שיטות:

I. שיטת ההפרשים לשערוך פרמטרי תנועה גלובלית (Incremental Method)

בשיטה זו מחושבת ההתמרה הגלובלית בין כל שתיים מתמונות הרצף הסמוכות. התנועה בין תמונת הייחוס לבין כל אחת מתמונת הרצף מורכבת משרשור של ההתמרות שבין תמונות סמוכות.



$$T = T_1 \oplus T_2 \oplus T_3 \oplus T_4$$

איור 4.5 – "שיטת הפרשים" – (Incremental Method)
Fig. 4.5: Incremental motion estimation method

כאשר ההתמרה הגלובלית הינה אפינית אזי שרשור של התמרות אפיניות הינה לינארית ופשוטה לביצוע.

אם סט הפרמטרים

$$\begin{pmatrix} X_2 \\ Y_2 \end{pmatrix} = \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} \begin{pmatrix} X_1 \\ Y_1 \end{pmatrix} + \begin{pmatrix} dx_1 \\ dy_1 \end{pmatrix} \tag{4.32}$$

מתאר את הקשר בין (X_1, Y_1) לבין (X_2, Y_2)

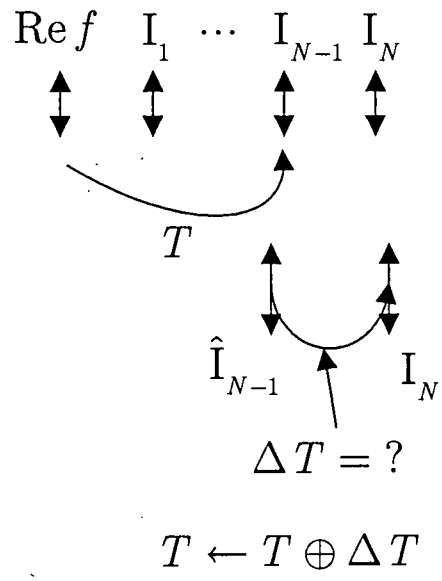
$$\begin{pmatrix} X_3 \\ Y_3 \end{pmatrix} = \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix} \begin{pmatrix} X_2 \\ Y_2 \end{pmatrix} + \begin{pmatrix} dx_2 \\ dy_2 \end{pmatrix} \tag{4.33}$$

מתאר את הקשר בין (X_3, Y_3) לבין (X_2, Y_2)
אִזי מהצבת (4.32) לתוך (4.33) מתקבל

$$\begin{pmatrix} X_3 \\ Y_3 \end{pmatrix} = \begin{pmatrix} a_2a_1 + b_2c_1 & a_2b_1 + b_2d_1 \\ c_2a_1 + d_2c_1 & c_2b_1 + d_2d_1 \end{pmatrix} \begin{pmatrix} X_1 \\ Y_1 \end{pmatrix} + \begin{pmatrix} a_2dx_1 + b_2dy_1 + dx_2 \\ c_2dx_1 + d_2dy_1 + dy_2 \end{pmatrix} \tag{4.34}$$

II. שיטה לשערוך תנועה מבוססת חיזוי

השיטה מבוססת על שערוך תנועה ΔT בין התמונה הנוכחית I_N לבין חיזוי התמונה הקודמת $\hat{I}_{N-1}$ בהתבסס על שערוך התנועה הקודם T . לאחר ששוערכה תנועה זו ΔT היא משורשרת לשערוך התנועה הקודם $T = T \oplus \Delta T$ כך שבשלב הבא תיבנה תמונת חיזוי של I_N , $\hat{I}_N$ בהסתמך על הטרנספורמציה T. החיזוי נבנה באמצעות הפעלת T על תמונת הייחוס.



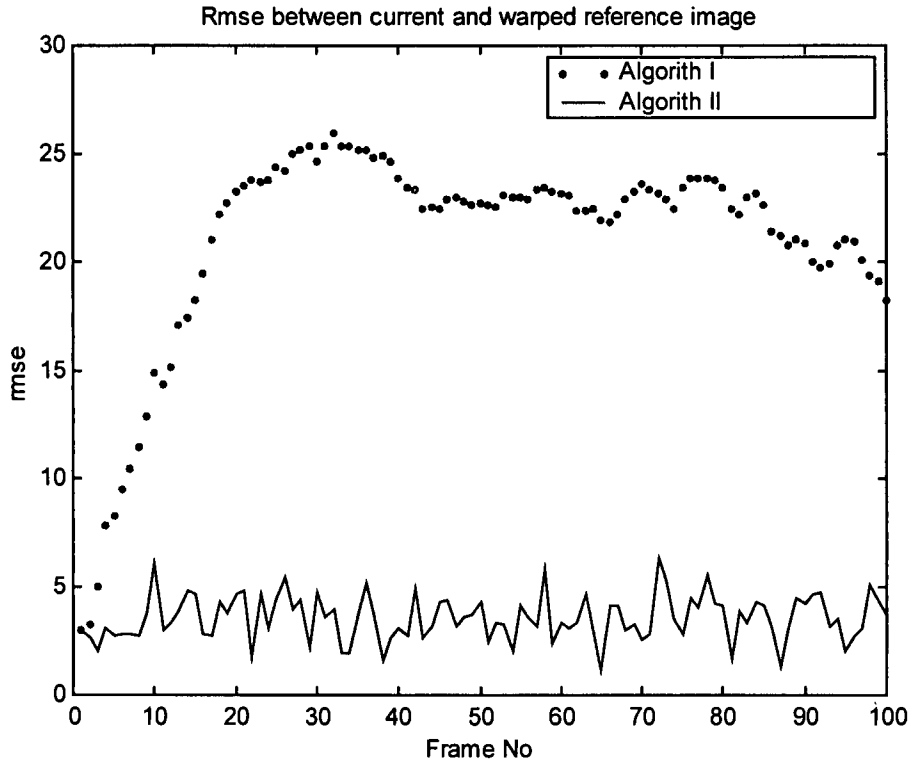
איור 4.6 – שיטת שערוך תנועה מבוססת חיזוי
Fig. 4.6: Prediction based motion estimation method

תוצאות הסימולציה

להלן מוצגות תוצאות סימולציה של הפעלת שתי השיטות לצורך שערוך תנועה בין תמונת הייחוס לבין תמונות הרצף.

שיטת השערוך מבוססת על מציאת נקודות התאמה ושערוך התמרה אפינית באמצעות פתרון משוואות Least Squares עבור תסריט ה – "Salesman" המסובב.

ככל ששערוך התנועה מדוייק יותר, אזי אנרגיית השגיאה של תמונת ההפרש בין כל תמונהה ברצף לבין חיזוייה באמצעות תמונת הייחוס (באמצעות מידע ההתמרה הגלובלית) צריך להיות קטן ככל האפשר.



איור 4.7 – השוואה כמותית בין שיטות שערוך תנועה ברצף תמונות

Fig 4.7: Quantitative comparison between motion estimation methods in image sequences

כפי שניתן לראות השיטה השנייה טובה יותר מהשיטה הראשונה. הדבר נובע משום שכאשר שערוך התנועה בין שתי תמונות ברצף גרוע (לדוגמא, משום שלא נמצאו מספיק נקודות התאמה טובות דיין) שגיאת השערוך משורשרת לכל תמונות הרצף (כלומר, אין מנגנון משוב שמסוגל לתקן את השגיאה שנפלה). בשיטת החיזוי, אם שוערכה התמרה לא – מדוייקת בשלב כלשהוא, ונבנתה תמונת חיזוי לא מדוייקת, אזי בשלב הבא תתוקן ההתמרה משום שמבוצע שערוך תנועה בין התמונה הנוכחית לבין התמונה החזוייה. זהו מנגנון פנימי של משוב.

דוגמא נוספת לביצועי שתי השיטות מבוסס על הערכה איכותית של ביצועי השערוך. מכיוון שרצף ה – "Salesman" המסובב הינו רצף סינטטי המורכב מהפעלת התמרת סיבוב על תמונה בודדת. אם תקוזז התנועה מכל אחת מתמונות הרצף , אזי אמורים לקבל את התמונה הראשונה. מיצוע כל התמונות לאחר קיזוז תנועה אמור לאפשר השוואה סובייקטיבית של איכות שערוך התנועה.

Averaged motion compensated image -Algorithm I

Averaged motion compensated image -Algorithm II

איור 4.8 – השוואה סובייקטיבית בין שיטות שערוך תנועה ברצף תמונות

Fig. 4.8: Subjective comparison between motion estimation methods in image sequences

כפי שניתן אכן לראות השיטה השנייה אכן נותנת תוצאות טובות יותר גם מהבחינה האיכותית.

פרק 5 החלקת פרמטרי התמרה אפינית באמצעות מסנן קלמן לינארי

שערוך פרמטרי ההתמרה האפינית שהצגנו מבוססת על מיצוי אוסף נקודות התאמה. נקודות ההתאמה נמצאות על סריג דיסקרטי ומציאת נקודת התאמה מורכבת ממציאת בלוק תואם מהתמונה הראשונה לשנייה (Block Matching, המסוגל לתאר הזזה בלבד). לכן הפרמטרים המשוערכים אינם מדוייקים מספיק ויש לשפרם. פרק זה כולל תיאור של שתי שיטות שונות המבוססות על מסנן קלמן לינארי לשיפור איכות השערוך של פרמטרי הטרנספורמציה האפינית.

5.1 מסנן קלמן לינארי

עבור משוואת מודל:

$$x_{k+1} = F_k \cdot x_k + w_k$$

ומשוואת מדידה המוגדרת באמצעות

$$z_k = C_k \cdot x_k + \varepsilon_k$$

יש למצוא את המשערך הלינארי האופטימלי של x_{k+1}, שנסמנו ב- $\hat{x}_{k+1|k+1}$, לפי קריטריון שגיאה ריבועית ממוצעת מינימלית, בהינתן $Z_{k+1} = \{z_0, z_1, \ldots, z_k, z_{k+1}\}$

x_k הינו וקטור המצב במימד n ו - z_k הינו וקטור המדידה במימד m.
w_k ו - ε_k הינם תהליכים אקראיים גאוסיים בעלי תוחלת אפס וקוואריאנס $E(w_k w_k^T) = Q_k$ $E(\varepsilon_k \varepsilon_k^T) = R_k$ בהתאמה, בת"ס זה בזה ובוקטור המצב x_k.

משוואות מסנן קלמן לינארי הינם [(AnM79, Ch 3)]:

שלב החיזוי:

$\hat{x}_{k+1|k} = F_k \cdot \hat{x}_{k|k}$ חיזוי וקטור המצב לזמן $k+1$.

$P_{k+1|k} = F_k \cdot P_{k|k} \cdot F_k^T + Q_k$ חיזוי קוואריאנס השגיאה לזמן $k+1$.

כאשר, $P_{k|k} = E(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^T$.

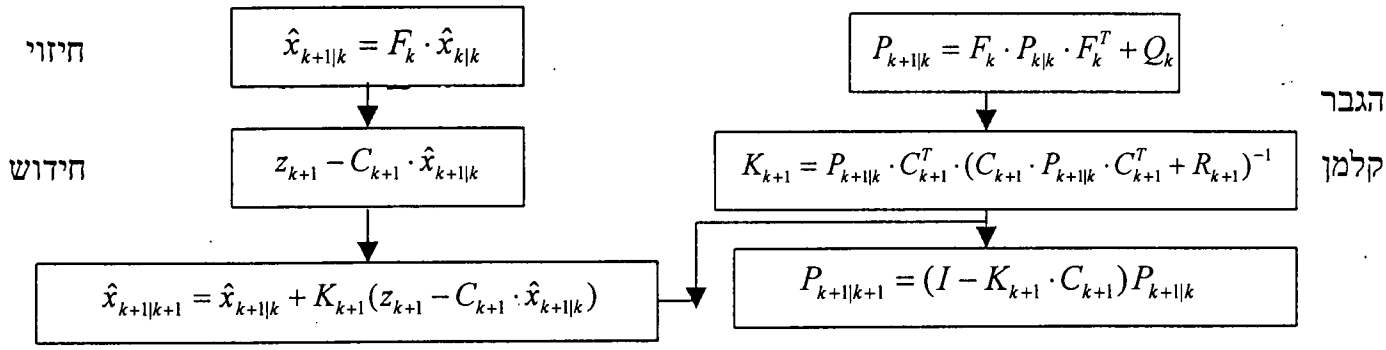
שלב השערוך:

$K_{k+1} = P_{k+1|k} \cdot C_{k+1}^T \cdot (C_{k+1} \cdot P_{k+1|k} \cdot C_{k+1}^T + R_{k+1})^{-1}$ הגבר קלמן.

$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1} \cdot (z_{k+1} - C_{k+1} \cdot \hat{x}_{k+1|k})$ שערוך וקטור המצב בזמן $k+1$

$P_{k+1|k+1} = (I - K_{k+1} \cdot C_{k+1}) P_{k+1|k}$ שערוך קוואריאנס בזמן $k+1$

דיאגרמת זרימה :



מסנן קלמן לינארי–איור 5.1

Fig. 5.1: The linear Kalman filter equations

5.2 אלגוריתם להחלקת נקודות התאמה

אם בין התמונה I_{k-1} ל - I_k חושבה טרנספורמציה אפינית T_k

$$X' = a_k X + b_k Y + \Delta_k^y$$
$$Y' = c_k X + d_k Y + \Delta_k^y \qquad (5.1)$$

אזי ההנחה הינה ש - $T_{k+1} \approx T_k$ (מכיוון שאנו מחשבים טרנספורמציה בין מסגרות עוקבות), ולכן לשערוך T_{k+1} נניח שנקודות העקיבה בתמונה I_k הועברו ל - I_{k+1} לפי T_k , לכן משוואת המודל עבור כל נקודת עקיבה ב - I_k הינה

$$\begin{bmatrix} X \\ Y \end{bmatrix}_{k+1} = \begin{bmatrix} a_k & b_k \\ c_k & d_k \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}_k + \begin{bmatrix} \Delta_k^x \\ \Delta_k^y \end{bmatrix} + w_k \qquad (5.2)$$

משוואת המדידה הינה

$$\begin{bmatrix} X' \\ Y' \end{bmatrix}_{k+1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}_{k+1} + \varepsilon_k \qquad (5.3)$$

כאשר T_k ידוע במדוייק $(w_k = 0)$, $\begin{bmatrix} X \\ Y \end{bmatrix}_k$ הינו וקטור נקודת עקיבה ב - I_k המותאם במדוייק ל - $\begin{bmatrix} X \\ Y \end{bmatrix}_{k+1}$ ב - I_{k+1} .

$\begin{bmatrix} X' \\ Y' \end{bmatrix}_{k+1}$ הינה המדידה המתקבלת מחיפוש בלוק תואם עבור $\begin{bmatrix} X \\ Y \end{bmatrix}_k$ ב - I_{k+1}.

w_k ו - ε_k הינם תהליכים אקראיים גאוסיים בעלי תוחלת אפס וקוואריאנס $Ew_k w_k^T = Q_k$ $E\varepsilon_k \varepsilon_k^T = R_k$ בת"ס זה בזה ובוקטור המצב $\begin{bmatrix} X \\ Y \end{bmatrix}_k$.

על המדידה $\begin{bmatrix} X' \\ Y' \end{bmatrix}_{k+1}$ מופעל מסנן קלמן לינארי לשערוך $\begin{bmatrix} X \\ Y \end{bmatrix}_{k+1}$. נקודות העקיבה המוחלקות משמשות לפתרון בעיית LS לחישוב פרמטרי הטרנספורמציה האפינית.

הערות

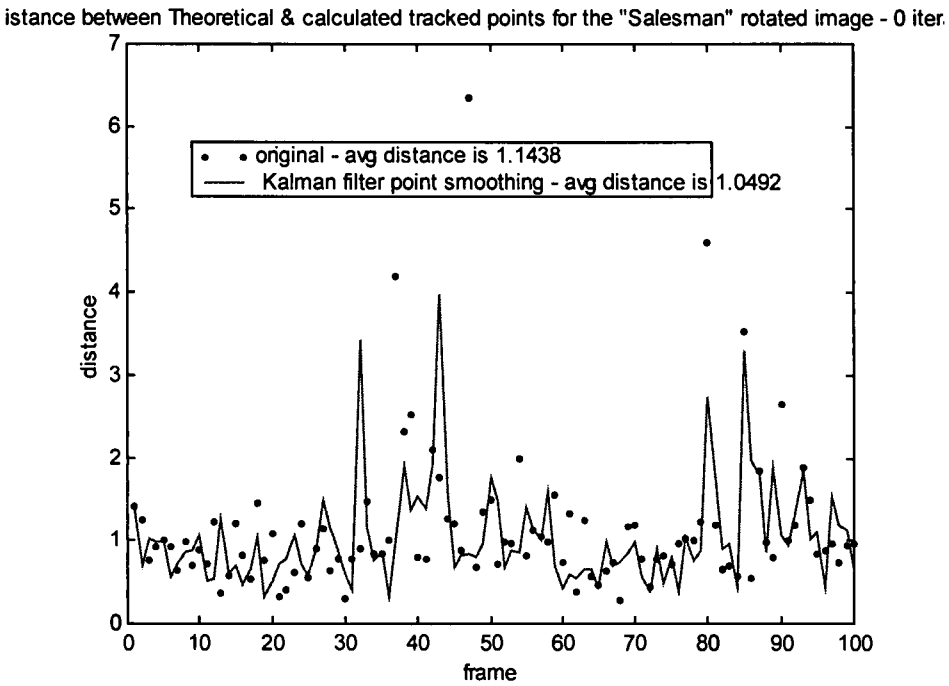
א. תנאי התחלה - כאשר $k = 1$ מחושבת T_1 בהתבסס על נקודות התאמה ללא הפעלת מסנן קלמן. עבור $k > 1$ מחושבת T_k באמצעות נקודות עקיבה שהוחלקו באמצעות מסנן קלמן. ניתן לבצע אתחול מתוחכם יותר, כגון מיצוע פרמטרי טרנספורמציה אפינית לאורך N מסגרות (ללא מסנן קלמן) ושערוך T_{N+1} תוך שימוש בתוצאה זו.

ב. ניתן לשפר את ביצועי האלגוריתם באמצעות הפעלת איטרצייה נוספת . לאחר ששוערכה הטרנספורמציה האפינית בהתבסס על נקודות העקיבה המוחלקות (באמצעות מסנן קלמן) יוחלקו שוב נקודות ההתאמה ((5.2) ו – (5.3)) בהתבסס על הטרנספורמציה האפינית החדשה ששוערכה T_{k+1}. נקודות העקיבה המוחלקות ישמשו לחישוב מחדש של הטרנספורמציה האפינית.

5.2.1 תוצאות סימולצייה

קריטריון אפשרי לבדיקה מבוקרת של ביצועי שערוך תנועה הינו סימון של מספר נקודות ידועות מראש בתוך התמונה, אשר ידועה תנועתן מראש (ולכן גם מיקומן כעבור זמן Δt) ולהשוות את מיקומן האמיתי (על – סמך מידע מוקדם) עם מיקומן המחושב לפי שערוך התנועה המתקבל. פרק זה משווה שיטות שונות להחלקת פרמטרי טרנספורמציה אפינית באמצעות השוואת המיקומים התיאורטיים של מספר נקודות למיקומיהן המחושבים.

אלגוריתם ההחלקה נבדק על רצף תמונות Salesman המסובב במהירות של 2° למסגרת. מטריצות רעש המודל והמדידה נבחרו כ - $R = 0.075 \cdot I_{2x2}$; $Q = 0.015 \cdot I_{2x2}$ תוצאות החלקת נקודות העקיבה מופיעות באיור 5.2.

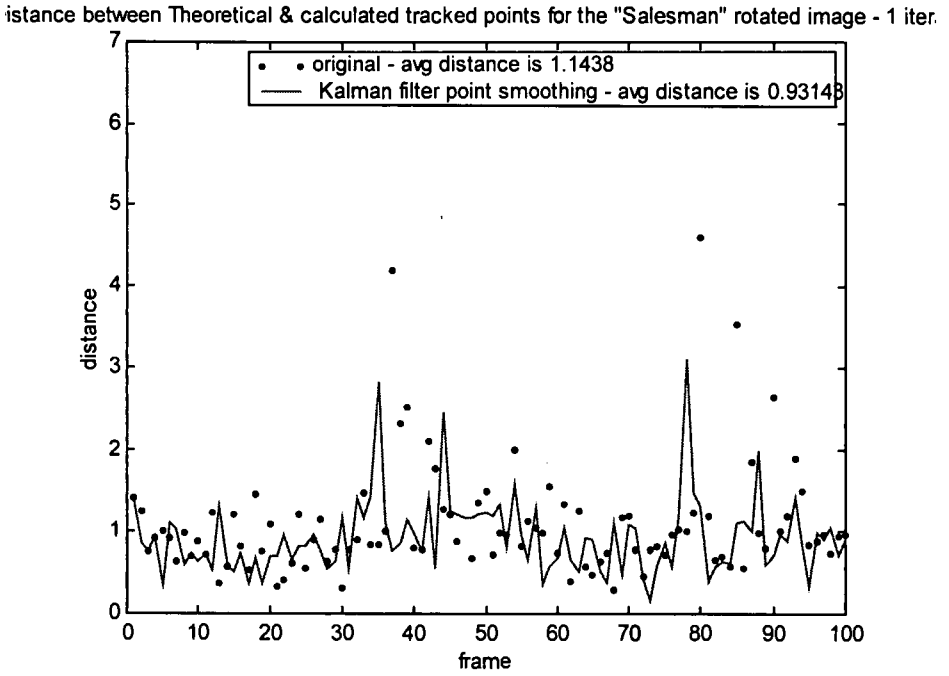


איור 5.2 – החלקת נקודות עקיבה באמצעות מסנן קלמן.

Fig 5.2: Matching points smoothing using linear Kalman filter

כפי שניתן לראות קיים שיפור של כ – 10%.

הבדיקה השנייה נערכה לפי האופציה בסעיף ב' לעיל. כאמור, השערוך מתבצע בשילוב איטרצייה נוספת שבמהלכה מוחלקות נקודות העקיבה לפי הטרנספורמציה האפינית שחושבה במסגרת הנוכחית. מתוך הנקודות המוחלקות בשנית משוערכת הטרנספורמציה האפינית במסגרת הנוכחית. תוצאות הבדיקה מופיעות איור 5.3.



איור 5.3 – החלקת נקודות עקיבה באמצעות מסנן קלמן – שתי איטרציות

Fig 5.3: Matching point smoothing using linear Kalman filter – two iterations

כפי שניתן לראות הפעלת האלגוריתם בפעם השנייה מניבה שיפור בממוצע של כ – 20% אולם עדיין ישנן מסגרות בהן השגיאה גדולה.

שיטה אחרת, המשתמשת במסנן קלמן, היא להחליק ישירות את פרמטרי הטרנספורמציה האפינית.

5.3 החלקת פרמטרי טרנספורמציה אפינית באמצעות מסנן קלמן

תחת ההנחה שפרמטרי הטרנספורמציה האפינית משתנים מעט בין שתי מסגרות עוקבות, ניתן לתאר את השתנותם לפי המודל הבא:

$$\begin{bmatrix} a \\ b \\ \Delta_x \\ c \\ d \\ \Delta_y \end{bmatrix}_{k+1} = \begin{bmatrix} 1&0&0&0&0&0 \\ 0&1&0&0&0&0 \\ 0&0&1&0&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&0&1&0 \\ 0&0&0&0&0&1 \end{bmatrix} \begin{bmatrix} a \\ b \\ \Delta_x \\ c \\ d \\ \Delta_y \end{bmatrix}_k + w_k \tag{5.4}$$

ומשוואת המדידה הינה

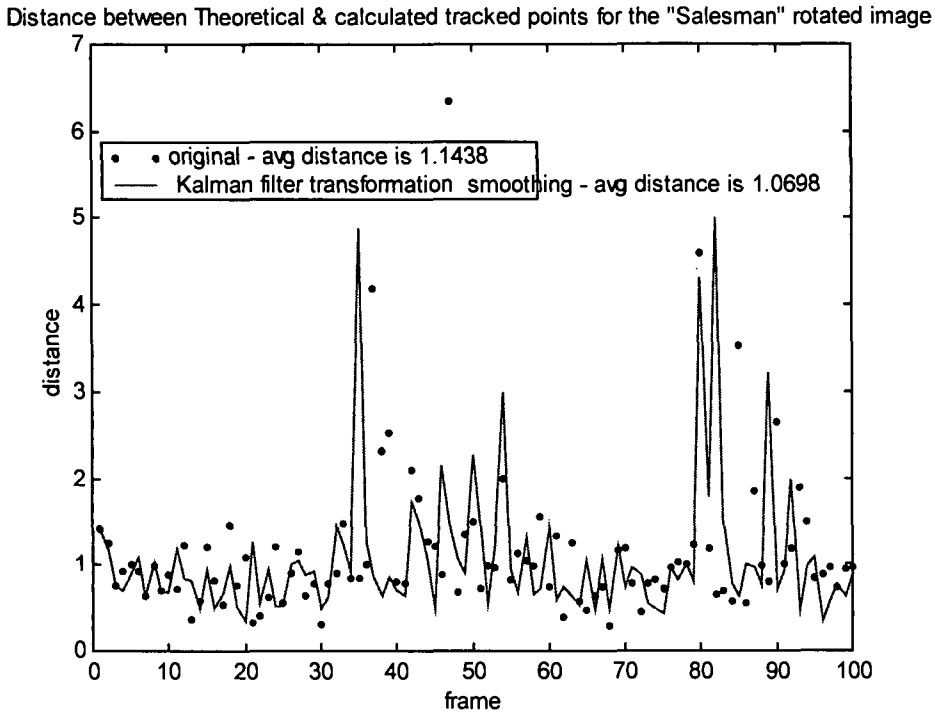
$$\begin{bmatrix} X_1' \\ Y_1' \\ \vdots \\ X_N' \\ Y_N' \end{bmatrix}_k = \begin{bmatrix} X_1 & Y_1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & X_1 & Y_1 & 1 \\ \vdots & & & & & \\ X_N & Y_N & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & X_N & Y_N & 1 \end{bmatrix}_k \begin{bmatrix} a \\ b \\ \Delta_x \\ c \\ d \\ \Delta_y \end{bmatrix}_{k+1} + \varepsilon_k \tag{5.5}$$

כאשר $\left[a\,,\,b\,,\,\Delta_x\,,\,c\,,\,d\,,\,\Delta_y\right]_{k+1}^T$ הינה הטרנספורמציה האפינית בין I_k ל - I_{k+1}, וכאשר $(\vec{X},\vec{Y})_k$ הינם נקודות ההתאמה בתמונה I_k להן נמצאה התאמה בתמונה I_{k+1} ע"י הנקודות $(\vec{X}',\vec{Y}')_k$.

w_k ו - ε_k מייצגים את רעש המודל והמדידה, בהתאמה. ההנחה היא שרעשים אלה הינם תהליכים אקראיים גאוסיים בעלי תוחלת אפס וקוואריאנס $Ew_k w_k^T = Q_k$ $E\varepsilon_k \varepsilon_k^T = R_k$ בת"ס זה בזה ובוקטור המצב $\left[a\,,\,b\,,\,\Delta_x\,,\,c\,,\,d\,,\,\Delta_y\right]_{k+1}^T$.

שערוך וקטור המצב $\left[a\,,\,b\,,\,\Delta_x\,,\,c\,,\,d\,,\,\Delta_y\right]_{k+1}^T$ מתבצע באמצעות הפעלת מסנן קלמן לינארי על – סמך נקודות ההתאמה $(\vec{X},\vec{Y})_k$ ו - $(\vec{X}',\vec{Y}')_k$.

אלגוריתם ההחלקה נבדק על רצף תמונות Salesman המסובב במהירות של 2° למסגרת. מטריצות רעש המודל והמדידה נבחרו כ - $R = 0.5 \cdot I_{2Nx2N}$; $Q = diag(0.03, 0.03, 0.3, 0.03, 0.03, 0.3)$ תוצאות החלקת פרמטרי ההתמרה אפינית מופיעות באיור 5.4.



איור 5.4 – החלקת פרמטרי התמרה אפינית באמצעות מסנן קלמן.

Fig 5.4: Affine transformation parameters smoothing using a linear Kalman filter

כפי שניתן לראות מתקבל שיפור של כ – 10%, ע"י השימוש במסנן קלמן.

חשוב לציין שאמנם בממוצע החלקה באמצעות מסנן קלמן הניבה שיפור בתוצאות, אולם עדיין ישנן נקודות שבהן ביצועי מסנן קלמן היו גרועים יותר.

תוצאות הסימולציות מראות שקיים פוטנציאל לשימוש במסנן קלמן. חשוב לציין שהאלגוריתמים המוצעים נבדקו על רצף סינטתי הנע במהירות זוויתית קבועה. יש לבדוק את האלגוריתמים על רצפים אמיתיים תוך חיפוש של ערכים מתאימים למטריצות הרעש R ו - Q.

הסיבות העיקרית לכך שלא נעשה שימוש במסנן קלמן בהמשך עבודה זו הנן הסיבוכיות החישובית הכבדה והעובדה שמציאת הפרמטרים המתאימים עבור מסנן קלמן בכל רצף תמונות אינה פשוטה ודורשת עבודה רבה.

פרק 6 שערוך פרמטרי התמרה מישורית

במושג של "סצינות תלת-מימדיות" מתייחסים לתמונות של עולם אשר איזורים שונים בו נמצאים בעומק שונה. באופן כללי, לכל נקודה בסצינה יכול להיות עומק כלשהוא. חישוב ההטלה הפרספקטיבית, בצורתה הכללית (4.15) אינו אפשרי בד"כ מכיוון שהינו מחייב ידיעת המרחקים של כל נקודה בסצינה מהמצלמה. במקרים מסויימים ניתן לקבל ביטוי כללי עבור ההתמרה. לדוגמא, כאשר הסצינה נמצאת על מישור אחד או שהמצלמה קבועה במקומה.

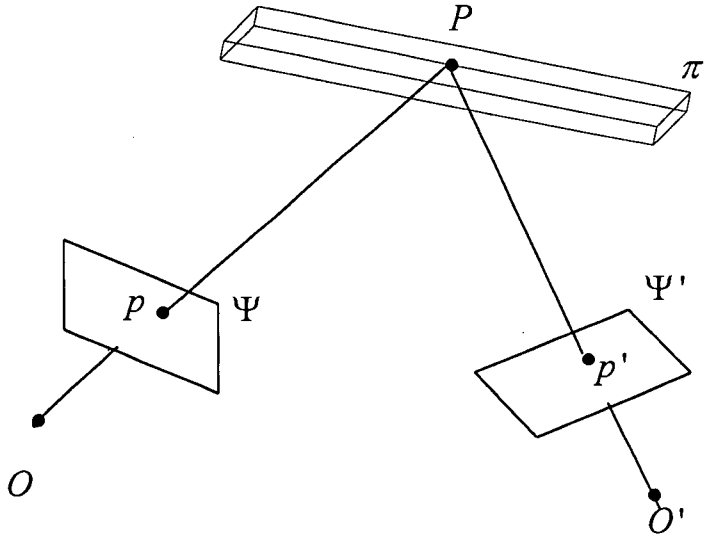
על – מנת שיתאפשר לתאר סצינה כללית באמצעות מספר קטן של פרמטרים, ניתן להגדיר את הסצינה כמורכבת ממספר מישורים. בחירה של מישורים רבים בגודל אינטיפיסימלי קטן נותנת תיאור מדוייק של הסצינה. עבודה זו מתמקדת בשערוך תנועה גלובלית במישור המצלמה כאשר הסצינה מורכבת ממישור אחד בלבד.

6.1 התמרה מישורית

מקרה פשוט אחד הינו עבור מצלמה קבועה במקומה. כלומר, מאפשרים סיבוב בלבד ללא הזזה ($\vec{T} \equiv 0$) – במקרה זה החלק הימני של המונה והמכנה ב – (4.15) מתאפס ומקבלים את הביטוי המוכר של ההתמרה המישורית:

$$X' = \frac{a_1 X + a_2 Y + a_3}{a_7 X + a_8 Y + 1}$$
$$Y' = \frac{a_4 X + a_5 Y + a_6}{a_7 X + a_8 Y + 1} \qquad (6.1)$$

כאשר הסצינה נמצאת על מישור אחד:



איור 6.1 – סצינה מישורית
Fig. 6.1: Planar scene

$P=(kx,ky,kz,k)^T$ הינה נקודה הומוגנית במרחב תלת – מימדי המוטלת על מישורי התמונות Ψ ו - Ψ' בנקודות ההומוגניות p ו - p', בהתאמה, המתקבלות מתנועת המצלמה מ - O ל - O'. אזי מתוך (4.11) ניתן לקבל

$$p = M_{\text{int}}[I_{3x3};0]P \tag{6.2}$$

$$p' = M_{\text{int}}[R;-\vec{T}]P \tag{6.3}$$

מערכת הייחוס (world reference frame) ומערכת הצירים של המצלמה בנקודה O מתלכדים.

שימוש ב (6.2) מניב $M_{\text{int}}^{-1}p = (kx\ ky\ kz)^T$ ומ (6.3) ניתן לקבל

$$p' = M_{\text{int}} R M_{\text{int}}^{-1} p - M_{\text{int}} kT \tag{6.4}$$

כאשר הנקודה $(x,y,z)^T$ נמצאת על מישור אזי מתקיימת משוואת המישור

$$n^T \cdot (x,y,z)^T = d_\pi \tag{6.5}$$

כאשר n שהינו הנורמל למישור ו - d_π שהינו סקלר נרמול כלשהוא, הינם פרמטרי המישור במערכת הקואורדינטות של O.

(6.5) יוצב בתוך (6.4) ויתקבל [Ro96]:

$$\begin{aligned} p' &= M_{\text{int}} R M_{\text{int}}^{-1} p - \frac{1}{d_\pi} M_{\text{int}} T n^T M_{\text{int}}^{-1} p \\ &= M_{\text{int}} (R - \frac{T n^T}{d_\pi}) M_{\text{int}}^{-1} p = Hp \end{aligned} \tag{6.6}$$

H ידועה כמטריצת ההומוגרפיה הקשורה למישור π.

מכיוון ש - p ו - p' הינם קואורדינטות הומוגניות, אזי לפי (4.12)

$$\begin{aligned} X' &= \frac{p_1'}{p_3'} = \frac{h_{11}p_1 + h_{12}p_2 + h_{13}p_3}{h_{31}p_1 + h_{32}p_2 + h_{33}p_3} \\ Y' &= \frac{p_2'}{p_3'} = \frac{h_{21}p_1 + h_{22}p_2 + h_{23}p_3}{h_{31}p_1 + h_{32}p_2 + h_{33}p_3} \end{aligned} \tag{6.7}$$

צמצום ב - p_3 מניב את ההתמרה המישורית הידועה

$$X' = \frac{h_{11}X + h_{12}Y + h_{13}}{h_{31}X + h_{32}Y + h_{33}}$$
$$Y' = \frac{h_{21}X + h_{22}Y + h_{23}}{h_{31}X + h_{32}Y + h_{33}} \qquad (6.8)$$

מכיוון ש - $X = \frac{p_1}{p_3}, Y = \frac{p_2}{p_3}$ (קואורדינטות התמונה בפיקסלים לפי (4.12)).

מכיוון ש – (6.6) הינה משוואה בין קואורדינטות הומוגניות אזי ניתן להגדיר את (6.6) כ:

$$\lambda m' = Hm \qquad (6.9)$$

כאשר $m' = (X', Y', 1), m = (X, Y, 1)$ ו - λ הינו קבוע נרמול כלשהוא.

6.2 סקר ספרות - שערוך פרמטרי התמרה מישורית

[TsH81] היו הראשונים שהראו שההתמרה המישורית הינה חד-ערכית (עד כדי קבוע נרמול). כלומר, בהנתן סט נקודות התאמה $(X,Y) \to (X',Y')$ קיים סט אחד של פרמטרי התמרה המישורית המתאים להתאמה זו (עד כדי קבוע נרמול). המחברים הראו שההתמרות הקשורות לסט הפרמטרים $a_i, (1 \le i \le 8)$ שייכים לקבוצה רציפה (Lie Group). מתוך תכונה זו ניתן לקבל סט משוואות ליניאריות לשערוך הפרמטרים. גם [ZhC93] מציעים להשתמש בשיטה הלינארית לשערוך פרמטרי ההתמרה המישורית.

שיטת השערוך באמצעות LS רגישה מאוד לשגיאות משום ששני הצדדים של מערכת המשוואות מכילים נתונים רועשים. אחת השיטות האפשריות להתמודד עם בעיית הרגישות לרעש הינה שיטת ה – TLS (Total Least Squares) (ראה סעיף 6.3). שיטת שערוך באמצעות TLS מורכבת חישובית, ומכיוון שבמטריצה A (6.26) קיימות שתי עמודות ללא רעש ניתן לחשב [KiK98] פתרון מדויק יותר ופשוט יותר חישובית המשלב TLS (עבור שש העמודות הרועשות) ואחריו פתרון LS של שתי העמודות ללא רעש.

קיימות גם מספר שיטות איטרטיביות:

מכיוון שדיוק שערוך פרמטרי ההתמרה המישורית אינו משביע רצון בד"כ, ניתן לשפר את השערוך באמצעות איטרציות נוספות [ScA99]. בתהליך זה, נקודות ההתאמה שהתקבלו באמצעות מציאת קורלציה מקסימלית בתמונה השנייה ואינן תואמות למודל המשוערך אינן נכללות בשערוך המודל באיטרציה הבאה. בשלב הבא מופעלת ההתמרה המישורית על התמונה הראשונה ליצירת תמונת חיזוי של התמונה השנייה. בין תמונת

חיזוי זו לתמונה השנייה משוערכים שוב פרמטרי ההתמרה המישורית וחוזר חלילה. ההתמרה הסופית מורכבת משרשור כל טרנספורמציות הביניים.

שיטות אחרות [Ho89] , [HoT88] מתבססות על שתי ההנחות הבסיסיות הבאות:
א. קיים מודל דו-מימדי להארת התמונה (רמות האפור של התמונה).
ב. שינויים ברמות האפור של התמונה בין תמונות סמוכות נובעים אך ורק מתנועת המצלמה או מתנועת עצם בתוך הסצינה – הנחה זו הינה ההנחה הבסיסית של מרבית האלגוריתמים לשערוך ה – Optical Flow.

בהנחה שלערכי התמונה S_k מודל ריבועי דו-מימדי:

$$S_k(X,Y) = A_1 + A_2X + A_3Y + A_4XY + A_5X^2 + A_6Y^2 \qquad (6.10)$$

כאשר (X,Y) הינן קואורדינטות התמונה,

והשינויים בין שתי תמונות נגרמים עקב ההתמרה המישורית:

$$(X',Y') = A(X,Y) = \left(\frac{a_1X + a_2Y + a_3}{a_7X + a_8Y + 1}, \frac{a_4X + a_5Y + a_6}{a_7X + a_8Y + 1}\right),$$

אזי, $S_{k+1}(X,Y) = S_k(X',Y') = S_k(A(X,Y))$

הניתן לקירוב מסדר שני של טור טיילור:

(6.11)

$$S_k(A(X,Y)) \approx S_k(X,Y) + \sum_{i=1}^{8} \frac{\partial S_k(X,Y)}{\partial a_i}\Delta a_i + 0.5\sum_{j=1}^{8}\sum_{i=1}^{8} \frac{\partial S_k(X,Y)}{\partial a_i}\frac{\partial S_k(X,Y)}{\partial a_j}\Delta a_i \Delta a_j$$

סביב התמרת יחידה מישורית $S_k(X,Y) = S_k(E(X,Y))$ כאשר:

$$E(X,Y) = \left(\frac{X + 0\cdot Y + 0}{0\cdot X + 0\cdot Y + 1}, \frac{0\cdot X + Y + 0}{0\cdot X + 0\cdot Y + 1}\right) = (X,Y)$$

ערכי ההפרשים בין תמונות סמוכות :

$$FD(X,Y) = S_{k+1}(X,Y) - S_k(X,Y) = S_k(A(X,Y)) - S_k(X,Y)$$

מניבים את המשוואה הלינארית הבאה [HoT88]:

$$
\begin{aligned}
FD(X,Y) &= G_x X \Delta a_1 + G_x Y \Delta a_2 + G_x \Delta a_3 \\
&\quad + G_y X \Delta a_4 + G_y Y \Delta a_5 + G_y \Delta a_6 \\
&\quad - X(G_x X + G_y Y)\Delta a_7 \\
&\quad - Y(G_x X + G_y Y)\Delta a_8 \\
&= H \cdot \Delta a
\end{aligned}
\tag{6.12}
$$

כאשר,

$$G_x = 0.5\left(\frac{\partial S_k(X,Y)}{\partial X} + \frac{\partial S_{k+1}(X,Y)}{\partial X}\right),\ G_y = 0.5\left(\frac{\partial S_k(X,Y)}{\partial Y} + \frac{\partial S_{k+1}(X,Y)}{\partial Y}\right)$$

$$a = (a_1,..,a_8)^T, e = (1,0,0,0,1,0,0,0)^T, \Delta a = a - e$$

$$H_i = (G_x X \ \ G_x Y \ \ G_x \ \ G_y X \ \ G_y Y \ \ G_y \ \ -X(G_x X + G_y Y) \ \ -Y(G_x X + G_y Y))$$

כש- H_i הינה השורה ה - i של H.

באמצעות פתרון LS עבור מספר נקודות ניתן לפתור את (6.12) אולם יש להשתמש בפתרון זה בד"כ בצורה איטרטיבית משום שמודל (6.10) איננו קירוב מדוייק דיו למודל האמיתי של התמונה.

ב – [MoK94] מבוצע ניתוח השוואתי בין הביצועים של אלגוריתם דומה ל – (6.12) לשערוך פרמטרים של מודלי תנועה דו – מימדיים מסדרים שונים (מודל מסדר רביעי – התמרה של גוף קשיח, התמרה אפינית בעלת שישה פרמטרים והתמרה מישורית בעלת שמונה פרמטרים). באמצעות שימוש בשתי תמונות, שעברו התמרה מישורית, שפרמטריה ידועים מראש, שוערכו מודלי התנועה וחושבה השגיאה הריבועית הממוצעת בין מיקומי נקודות התיאורטיים והמחושבים. לפי מסקנות המאמר, כאשר מישור הסצינה כמעט מאונך לציר האופטי, הן המודלים מסדר ארבע והן מסדר שש מניבים שגיאה קטנה, עבור מגוון רב של גדלי הזזה וסיבוב. רק המודל האפיני מצליח להניב שגיאה קטנה יחסית עבור מקרים בהם מישור הסצינה נטוי ביחס לציר האופטי. פרמטרי המודל המישורי לא התכנסו כלל כאשר איזור התמיכה לחישוב (6.12) היה בגודל $64x64$ פיקסלים (בניגוד למודלי התנועה הפשוטים יותר) והינו רגיש יותר לרעש.

ב – [DI91] מוצע להשתמש באלגוריתם האיטרטיבי של ניוטון - רפסון לצורך שערוך טרנספורמציות מרחביות כלשהן (ובפרט שערוך ההתמרה המישורית) תוך שימוש בקריטריון מינימום שגיאה ריבועית בין התמונה האמיתית ושערוכה (בהתבסס על ההתמרה המרחבית).

קיימים מספר מאמרים [Pa94], [HuC98] המציעים להשתמש באלגוריתם האיטרטיבי Levenberg-Marquardt שהינה שיטת קירוב LS – לא-לינארית

[Park94] לצורך מציאת וקטור הפרמטרים $\vec{a}=(a_1,..,a_8)$ המקיים:

$$\arg\min_{\vec{a}} \sum_i (X_i' - \frac{a_1X_i + a_2Y_i + a_3}{a_7X_i + a_8Y_i + 1})^2 + (Y_i' - \frac{a_4X_i + a_5Y_i + a_6}{a_7X_i + a_8Y_i + 1})^2 \quad (6.13)$$

מכיוון ש - [Park94] מתייחסים למקרה בו אין העתקה של המצלמה, אזי $\vec{a}=(r_{ij}/r_{33}),\ 1\le i,j\le 3$.

ב – [HuC98] מוצע לשפר את ביצועי אלגוריתם Levenberg-Marquardt באמצעות שימוש בפירוק מולטי – רזולוצייה. הפרמטרים המחושבים עבור רמת הפירוק הנמוכה ביותר הינם פרמטרי ההתחלה לאיטרציות ברמת הפירוק הגבוהה יותר.

ההבדל בין [SzC] לבין [HuC98] הינה ההנחה ששדה התנועה $(u=X'-x, v=Y'-Y)$ ניתן למידול באמצעות splines. הנחה זו מניחה אילוץ חלקות על שדה התנועה.

את הקואורדינטות בפיקסלים ניתן לקבל מתוך הקואורדינטות ההומוגניות באמצעות חלוקת השורות הראשונה והשנייה בשורה השלישית של (6.6).
חלוקת השורות הראשונה והשנייה בשורה השלישית של (6.6) והעברת אגפים מניבה את שתי המשוואות ההומוגניות:

$$\begin{aligned} V_1h &= 0 \\ V_2h &= 0 \end{aligned} \quad (6.14)$$

כאשר,

$$\begin{aligned} h &= \{h_1,h_2,..,h_9\} \\ V_1 &= \{X,Y,1,0,0,0,-XX',YX',X'\} \\ V_2 &= \{0,0,0,X,Y,1,-XY',YY',Y'\} \end{aligned}$$

ב – [LuF96],[LuF93] מוצגת השיטה הלינארית לשערוך ההתמרה המישורית באמצעות LS ע"י מזעור (6.14):

$$\min(\sum_i (V_{1i}h)^2 + (V_{2i}h)^2) \quad (6.15)$$

יתרונות שיטה זה הינה הסיבוכיות הנמוכה. לטענת – [LuF96],[LuF93] שיטת LS הינה מוטה (biased) מכיוון שנקודות ההתאמה אינן מנורמלות.

מיזעור הביטויים המנורמלים מניב ביטויים לא-לינאריים, שיש למזערם באמצעות פתרון איטרטיבי. המחברים מציעים שני קריטריוני מזעור :

א. מזעור המרחק האוקלידי שבין הנקודות (הקריטריון נבחר כך שלשתי התמונות משקל זהה) :

$$\min(\sum_i (d(m',Hm)^2 + d(m,H^{-1}m')^2) \quad (6.16)$$

$m \leftrightarrow m'$ הינה נקודת התאמה הומוגנית ו - H הינה מטריצת ההומוגרפיה (6.6).

המרחק האוקלידי שבין ישר $l=(a,b,c)$ ונקודה $m=[u_0,v_0,1]^T$ הינו

$$d(m,l)=\frac{au_0+bv_0+c}{\sqrt{a^2+b^2}} \tag{6.17}$$

וכפי שמופיע בנספח א', קו דו-מימדי מאופיין באמצעות קואורדינטה הומוגנית תלת-מימדית.

$A=Hm$ ו - $B=H^{-1}m'$ הינן קואורדינטות הומוגניות שמרחקן מ - m' ו - m הינו בהתאמה:

$$d(m',Hm)=\frac{m'^T A}{\sqrt{A_1^2+A_2^2}} \tag{6.18}$$

$$d(m,H^{-1}m')=\frac{m^T B}{\sqrt{B_1^2+B_2^2}}$$

X_n הינו האיבר ה - n של הוקטור X.

ג. מיזעור המשוואה הלינארית (6.14) הממושקלת לפי מידת אי-הדיוק של איברי הוקטור:

$$\min\left(\sum_i \frac{(V_{1i}h)^2}{\|\nabla(V_{1i}h)\|^2}+\frac{(V_{2i}h)^2}{\|\nabla(V_{2i}h)\|^2}\right) \tag{6.19}$$

כאשר,

$$\nabla(V_{1i}h)=\left(\frac{\partial(V_{1i}h)}{\partial X},\frac{\partial(V_{1i}h)}{\partial Y},\frac{\partial(V_{1i}h)}{\partial X'},\frac{\partial(V_{1i}h)}{\partial Y'}\right)$$

$$\nabla(V_{2i}h)=\left(\frac{\partial(V_{1i}h)}{\partial X},\frac{\partial(V_{1i}h)}{\partial Y},\frac{\partial(V_{2i}h)}{\partial X'},\frac{\partial(V_{2i}h)}{\partial Y'}\right)$$

על מנת לפתור את (6.16) או (6.19) יש להשתמש בשיטה מיזעור איטרטיבית. [LuF96],[LuF93] לא מתארים את שיטת המזעור, אלא מזכירים שהשתמשו בשגרות מספריות NAG (Numerical Algorithms Group) ו- Numerical Recipes [Pr86]. לטענתם השיטות האיטרטיביות הניבו שגיאה קטנה יותר מהשיטה הלינאריות, ללא בעיות התכנסות.

ה - Fundamental Matrix (F) הינה מטריצה המקשרת בין קואורדינטות בשתי תמונות כפונקציה של תנועת המצלמה בין שתי התמונות. נספח א' מכיל הסבר מפורט של

מבנה מטריצה זו אולם לצורך סקירת שיטות לשערוך פרמטרי התמרה מישורית מספיק לציין את הקשר בין הקואורדינטות $((X,Y),(X',Y'))$ בשתי תמונות:

$$(kX,kY,k)F(lX',lY',l)^T = 0 \qquad (6.20)$$

משוואה זו מניבה משוואה הומוגנית בתשעה נעלמים. את (6.14) ניתן לבטא כסט משוואות הומוגניות בתשעה נעלמים ולכן ניתן להשתמש במספר שיטות לינאריות שהוצעו עבור שערוך Fundamental Matrix לשערך את פרמטרי ההתמרה המישורית.

א. ב – [Zh98] ו - [LuF96] מוצע לפתור את בעיית המינימזצייה: מציאת וקטור $\vec{e}$ הממזער

$$\min_{\vec{e}} \| G\vec{e} \| \to 0 \qquad (6.21)$$

תחת האילוץ ש - $\| \vec{e} \| = 1$ הינו הוקטור העצמי השייך לערך העצמי המינימלי של $G^T G$ הוכחה לטענה זו נמצאת ב - [Zh98]. המטריצה G מוגדרת לפי (6.26) כ: $G = [A; -b]$

ב. כפי שניתן לראות מ (6.14), עמודות h שונות בסדרי גודל זו מזו, מכיוון ש - (X,Y,X',Y') הינם בערכי פיקסלים שלמים ולכן עמודות בסדרי גודל של $(1,X,X^2)$ שונות מאוד זה מזה וגורמות לחוסר יציבות נומרית. [Ha95] מציע לבצע נרמול של הקואורדינטות (X,Y,X',Y') למערכת צירים מרכזית כך שסטיית התקן של הקואורדינטה הממוצעת תהייה שווה ליחידה ($\sigma_X^2 = \sigma_Y^2 = \sigma_{X'}^2 = \sigma_{Y'}^2 = 1$) סביב המרכז.

לפי (6.6):

$$\begin{pmatrix} lX' \\ lY' \\ l \end{pmatrix} = H \begin{pmatrix} kX \\ kY \\ k \end{pmatrix}$$

ניתן למצוא מטריצת התמרה T_{3x3} כך ש -

$$\sigma_X^2 = \sigma_Y^2 = \sigma_{X'}^2 = \sigma_{Y'}^2 = 1, \bar{X} = \bar{Y} = \bar{X}' = \bar{Y}' = 0$$

ולכן

$$(T \begin{pmatrix} lX' \\ lY' \\ l \end{pmatrix}) = THT^{-1}(T \begin{pmatrix} kX \\ kY \\ k \end{pmatrix}) \qquad (6.22)$$

לכן ניתן לשערך את THT^{-1} במקום H . לאחר ששוערך $H'=THT^{-1}$ ההומוגרפיה הינה $H=T^{-1}H'T$.

אם ה - Fundamental Matrix (F, ראה נספח א') ידועה, אזי ניתן להשתמש ב- $Fv=0$ לצורך שערוך ה – epipoles (v). מתוך שערוך ה – epipoles ב - [Sh94], [ShW95], מוצע להשתמש ב $Hv=v'$ וב - $Hp=p'$ (p,p' הינן נקודות ההתאמה) לצורך פתרון מערכת משוואות לינארית לשם שערוך מטריצת ההומוגרפיה. ה- Fundamental Matrix מתארת את המיקום היחסי בין שתי המצלמות (או אותה מצלמה המצלמת במיקומים שונים) וניתן לחשבה באמצעות כיול תמונות.

ב - [VZR96] מוצע להשתמש במזעור איטרטיבי של מטריצת ההומוגרפיה כאשר מניחים שהמכנה של (6.8) קבוע. משוואה (6.6) בקואורדינטות ההומוגניות ניתנת להצגה (6.8) כ:

$$X'=\frac{h_0^T p}{h_2^T p},\ Y'=\frac{h_1^T p}{h_2^T p}$$

כאשר h_i הינה השורה ה - i של H (6.6), p הינה הקואורדינטה ההומוגנית בתמונה הראשונה המועתקת לקואורדינטה בפיקסלים (X',Y') בתמונה השנייה.

שגיאת שערוך נקודות ההתאמה $\zeta=(\bar{X}'-X',\bar{Y}'-Y')$ כאשר $(\bar{X}',\bar{Y}')$ הינם הנקודות המדודות (בצירוף רעש) של הנקודות התיאורטיות (X',Y') הינה:

$$\begin{pmatrix}\zeta\\0\end{pmatrix}=\frac{1}{h_2^T m}(\lambda m'-Hm) \qquad (6.23)$$

λ הינו קבוע נירמול הנובע מתוך (6.9).

$m'=(\bar{X}',\bar{Y}',1)^T, m=(X,Y,1)^T$ הינן הקואורדינטות ההומוגניות של נקודה במרחב כפי שהן נמדדות בשתי תמונות שונות.

אם ידוע שמטריצת הקוואריאנס של ζ הינה Λ הרי שעל מנת לחשב את ההומוגרפיה ניתן לבצע מזעור של מרחק Mahalanobis:

$$\begin{aligned}\varepsilon&=\sum_i \zeta_i^T\Lambda^{-1}\zeta_i^T\\&=\sum_i\frac{1}{h_2^T m_i}(\lambda_i m_i'-Hm_i)^T\Lambda^{-1}\frac{1}{h_2^T m_i}(\lambda_i m'-Hm)\end{aligned} \qquad (6.24)$$

m_i הינה נקודה מספר i .

קריטריון השגיאה (6.24) איננו לינארי מכיוון ש m_i נמצא במכנה של ε .

[VZR96] מציעים להניח ש h_2 הינו ניחוש ראשוני התחלתי הנבחר על – מנת למזער את (6.24) תחת הנחה זו. לאחר שערוך H מתקבל שערוך חדש של h_2 וחוזר חלילה. פתרון H מתקבל באמצעות סט משוואות לינאריות הנובעות מתוך גזירת ε לפי λ_i ולפי H.

ניתן לשלב מספר תמונות מזוויות שונות של אותו מישור לצורך שערוך מדוייק יותר של מטריצת ההומוגרפיה. ב – [ShW00] מציעים להשתמש בשלוש תמונות על – מנת לשערך מטריצת טנזור תלת – מימדית בגודל $3x3x3$. הטנזור משלב את שתי ההומוגרפיות כך שנוספים אילוצים אי-לינאריים נוספים – מעבר לאילוצים הלינאריים הקיימים בחישוב ההומוגרפיות בנפרד מתוך זוגות של שתי תמונות. ניתן לשערך את הטנזור באמצעות ארבע נקודות התאמה בשלוש התמונות.

בסקר ספרות זה נסקרו שיטות רבות לשערוך פרמטרי התמרה מישורית. חשוב להדגיש שהשוואה אמיתית בין השיטות צריכה להתבצע באמצעות אותם תמונות ותחת אותם אילוצי חישוביות. מכיוון שכל שיטה מוצעת נבדקה באמצעות תמונות שונות, ניתן רק להעריך את ביצועי האלגוריתמים מבחינה איכותית, תוך הסתמכות על ניתוח הביצועים שהופיעו במאמרים.

ניתן לחלק את השיטות לשתי קבוצות עיקריות: שיטות איטרטיביות ושיטות לא–איטרטיביות. באופן כללי ניתן לקבוע שהשיטות הלא – איטרטיביות אינן רובסטיות דיין ולא הניבו ביצועים מספקים תחת תנאי רעש. השיטות האיטרטיביות הניבו תוצאות טובות יותר, אם כי השיטות האיטרטיביות סובלות משלושה חסרונות עיקריים:

א. סיבוכיות גבוהה.

ב. לא מובטחת התכנסות.

ג. אם קיימת התכנסות, אזי לא מובטחת התכנסות למינימום גלובלי (אלא למינימום מקומי).

6.3 שיטות Least Squares לשערוך פרמטרי ההתמרה המישורית

ההתמרה המישורית המתארת העתקה בין הקואורדינטה (X,Y) ל - (X',Y') הינה

$$X' = \frac{a_1 X + a_2 Y + a_3}{a_7 X + a_8 Y + 1}$$
$$Y' = \frac{a_4 X + a_5 Y + a_6}{a_7 X + a_8 Y + 1} \qquad (6.25)$$

על – מנת להשתמש בפתרון לינארי יש להעביר את המכנים ב – (6.25) לצד השני כך שבעיית ה- LS המתקבלת הינה פתרון המשוואות ב – (6.26) לקבלת שגיאה ריבועית מינימלית בשערוך $\vec{a}$:

(6.26)

$$A\cdot\vec{a} = \begin{bmatrix} X_1 & Y_1 & 1 & 0 & 0 & 0 & -X_1X'_1 & -Y_1X'_1 \\ 0 & 0 & 0 & X_1 & Y_1 & 0 & -X_1Y'_1 & -Y_1Y'_1 \\ & & & \vdots & & & & \\ X_n & Y_n & 1 & 0 & 0 & 0 & -X_nX'_n & -Y_nX'_n \\ 0 & 0 & 0 & X_n & Y_n & 0 & -X_nY'_n & -Y_nY'_n \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_8 \end{bmatrix} = \begin{bmatrix} X_1' \\ Y_1' \\ \vdots \\ X_n' \\ Y_n' \end{bmatrix} = b$$

שתי בעיות עיקריות קיימות בפתרון של (6.25):

1. פתרון LS של (6.26) אינו מביא למזעור של הנורמות:

$$\left\|X' - \frac{a_1X + a_2Y + a_3}{a_7X + a_8Y + 1}\right\|, \quad \left\|Y' - \frac{a_4X + a_5Y + a_6}{a_7X + a_8Y + 1}\right\|$$

אלא למזעור של הנורמות:

$$\|X'(a_7X + a_8Y + 1) - (a_1X + a_2Y + a_3)\|$$
$$\|Y'(a_7X + a_8Y + 1) - (a_4X + a_5Y + a_6)\|$$

פתרונות אלה אינם מביאים לתוצאות זהות.

2. פתרון LS של מערכת המשוואות הוא

$$A\vec{x} = b$$
$$\Downarrow \qquad (6.27)$$
$$\vec{x}_{LS} = (A^T A)^{-1} A^T b$$

כאשר A ידועה במדוייק, בעוד b הינו וקטור מידע בצירוף רעש. (6.26) מכילה בצד שמאל של השוויון גם את האיברים X',Y' שהינם נקודות בצירוף רעש (הן משום התאמה לקויה והן משום שכל הנקודות נמצאות על סריג דיסקרטי, ולכן מתווספת שגיאת קוונטיזציה) ולכן פתרון LS איננו מתאים כאן מבחינה עקרונית.

פתרון מערכת משוואות לינארית, שבה שני הצדדים של השוויון מכילים רעש, מתאפשר באמצעות TLS [GoV80]. הרעיון עליו מבוסס פתרון בעיית TLS נובע מכך שקיימות שגיאות גם באיברי המטריצה A ולכן במקום לחפש את הפתרון

$$\hat{a} : \| A \cdot \hat{a} - b \| = \min_t \| A \cdot t - b \| \tag{6.28}$$

כפי שנעשה בפתרון בעיית LS, מדוייק יותר למצוא את הפתרון $\hat{a}$ כך שמתקיים

$$A' \hat{a} = b' \tag{6.29}$$

כאשר A', b' מתקבלים ע"י פתרון בעיית המינימיזצייה

$$\| [A' | b'] - [A | b] \|_F = \min_{range(A')=b'} \| [B | z] - [A | b] \|_F \tag{6.30}$$

$range(A')$ הינו הטווח הנפרש באמצעות עמודות A' ו - $\| \|_F^2$ היא הנורמה של פרוביניוס והינה מוגדרת כסכום ריבועי האיברים במטריצה. כאשר אם B הינה מטריצה בגודל $n_1 x n_2$ ו - z הינו וקטור עמודה באורך n_1 אזי $[B|z]$ היא מטריצה בגודל $n_1 x (n_2 + 1)$ ש - n_2 העמודות הראשונות בה הן עמודות המטריצה B והעמודה האחרונה היא וקטור העמודה z . את בעייה (6.30) ניתן לפתור באמצעות ביצוע Singular Value Decomposition של המטריצה $[A|b]$ ומשמעות הפעולה שמנסים לשנות גם את A בנוסף ל - b כך שנקבל התאמה טובה יותר של המדידות למודל.

המטריצה המאוחדת B מתקבלת מ – (6.26) ע"י הרחבת A ע"י b .

$$B = \begin{bmatrix} X_1 & Y_1 & 1 & 0 & 0 & 0 & -X_1X'_1 & -Y_1X'_1 & X_1' \\ 0 & 0 & 0 & X_1 & Y_1 & 0 & -X_1Y'_1 & -Y_1Y'_1 & Y_1' \\ & & & \vdots & & & & & \\ X_n & Y_n & 1 & 0 & 0 & 0 & -X_nX'_n & -Y_nX'_n & X_n' \\ 0 & 0 & 0 & X_n & Y_n & 0 & -X_nY'_n & -Y_nY'_n & Y_n' \end{bmatrix} \tag{6.31}$$

בעזרת פירוק SVD ,

$$B = U \cdot diag(\lambda_1, .., \lambda_9) \cdot V^T$$

U, V יוניטריות.

אזי לפי [GoV80]:

$$\vec{a} = -\frac{V_{1:8,9}}{V_{9,9}} \tag{6.32}$$

(שמונת האיברים הראשונים של העמודה התשיעית של V מחולקים באיבר (9,9) של V).

6.4 פתרון מקורב לשערוך פרמטרי התמרה מישורית

תוצאות הסימולציות, שיובאו בהמשך הסעיף , מראות כי הן שיטת LS והן שיטת TLS לשערוך פרמטרי התמרה מישורית רגישות לרעש. כלומר, רגישות לכך שנקודות ההתאמה אינן ידועות במדוייק ולכן שערוך פרמטרי ההתמרה המישורית אינו מדוייק דיו.

חישוב טרנספורמצית תנועה לצרכי עקיבה מורכבת מחישוב התנועה בין תמונת ייחוס לתמונה הנוכחית. בין תמונת הייחוס לתמונה הנוכחית קיימת תנועה רבה, אולם בין שתי תמונות עוקבות (לדוגמא, תמונות המתקבלות בקצב של 30 Hz) קיימת תנועה מעטה . ולכן ניתן לבצע שערוך מקורב של ההתמרה המישורית בין שתי תמונות ולשלב שערוך זה בשערוך הכולל של ההתמרה המישורית בין תמונת הייחוס לתמונה הנוכחית.
המכנה של ההתמרה המישורית (כפונקצייה של פרמטרי המצלמה - (7.12)) הינו:

$$(r_{31}+\frac{dz\cdot k_1}{D})X+(r_{32}+\frac{dz\cdot k_2}{D})Y+F_1r_{33}+F_1\frac{dz}{D} \tag{6.33}$$

כאשר עפ"י (7.13) :

$$\begin{aligned} r_{31} &= n_1n_3(1-\cos\alpha)-n_2\sin\alpha \\ r_{32} &= n_2n_3(1-\cos\alpha)+n_1\sin\alpha \\ r_{33} &= n_3^2+(1-n_3^2)\cos\alpha \end{aligned} \tag{6.34}$$

(n_1,n_2,n_3) הינו ציר הסיבוב של המצלמה, α הינה זווית הסיבוב ו- $\vec{T}=(dx,dy,dz)$ הינו וקטור ההזזה .

בין שתי מסגרות עוקבות ניתן להניח כי:

$$\alpha\to 0, dx\to 0, dy\to 0, dz\to 0$$
$$\Downarrow$$
$$r_{31}\to 0, r_{32}\to 0, r_{33}\to 1$$

ולכן ניתן להשתמש אז בקירוב של ההתמרה המישורית ע"י התמרות פשוטות יותר.

קירובים אפשריים:

1. התמרה אפינית:

$$\begin{aligned} X' &= aX+bY+\Delta_x \\ Y' &= cX+dY+\Delta_y \end{aligned} \tag{6.35}$$

2. התמרה בי–ליניארית [Te95] :

$$X' = aX + bY + cXY + d$$
$$Y' = eX + fY + gXY + h \qquad (6.36)$$

3. התמרה מסדר שני:

$$X' = b_1 X + b_2 Y + b_3 + b_7 X^2 + b_8 XY$$
$$Y' = b_4 X + b_5 Y + b_6 + b_7 XY + b_8 Y^2 \qquad (6.37)$$

מכיוון שקירוב זה הניב את הביצועים הטובים ביותר בסימולציות שיוצגו בהמשך, מובא בנספח ב' בפירוט פיתוח קירוב זה (עפ"י [Te95]).

השוואת ביצועים

על-מנת לבדוק מצד אחד את רגישות שיטות השערוך השונות לרעש, ומצד שני את היכולת של ביטויים מקורבים של ההתמרה המישורית לייצג בצורה מספקת את ההתמרה המישורית, בוצעו סימולציות על רצף תמונות סינטטי. בסימולציות הופעלה התמרה מישורית ידועה מראש על סט קואורדינטות

$$(X,Y) \xrightarrow{T_perspective} (X',Y')$$

ונבדקה היכולת של האלגוריתמים השונים לשערך את ההתמרה T בהתבסס על $(X,Y,X'+\eta_1,Y'+\eta_2)$ (נקודות התאמה שלהן נוסף רעש לבן אחיד η , בלתי תלוי בנקודות ההתאמה) . קריטריון האיכות הינו מזעור

$$\| T(X,Y) - (X',Y') \|$$

התוצאות הבאות מציגות את ביצועי האלגוריתמים עבור מיצוע של 40 הרצות. כאשר הציר האנכי הינו המרחק הממוצע בין הנקודות התיאורטיות (X',Y') לבין המיקום המשוערך של הנקודות $T(X,Y)$. הציר האופקי מתאר את עוצמת הרעש, כאשר שונות הרעש האחיד הינה $\sigma^2/12$.

איורים 6.2,6.3 ו – 6.4 מתארים את התוצאות עבור מודלים מקרבים שונים של ההתמרה המישורית (6.35),(6.36),(6.37) . כל גרף באיור מכיל מספר אופציות שונות עבור המונה של ההתמרה המישורית כאשר המכנה נשאר קבוע – חשוב להדגיש שככל שהמקדמים של המכנה a_7,a_8 גדולים יותר בערכם המוחלט, אזי הקירובים השונים של ההתמרה המישורית תקפים פחות, מכיוון שאז קשה להזניח את המכנה (היוצר את אי-הלינאריות).

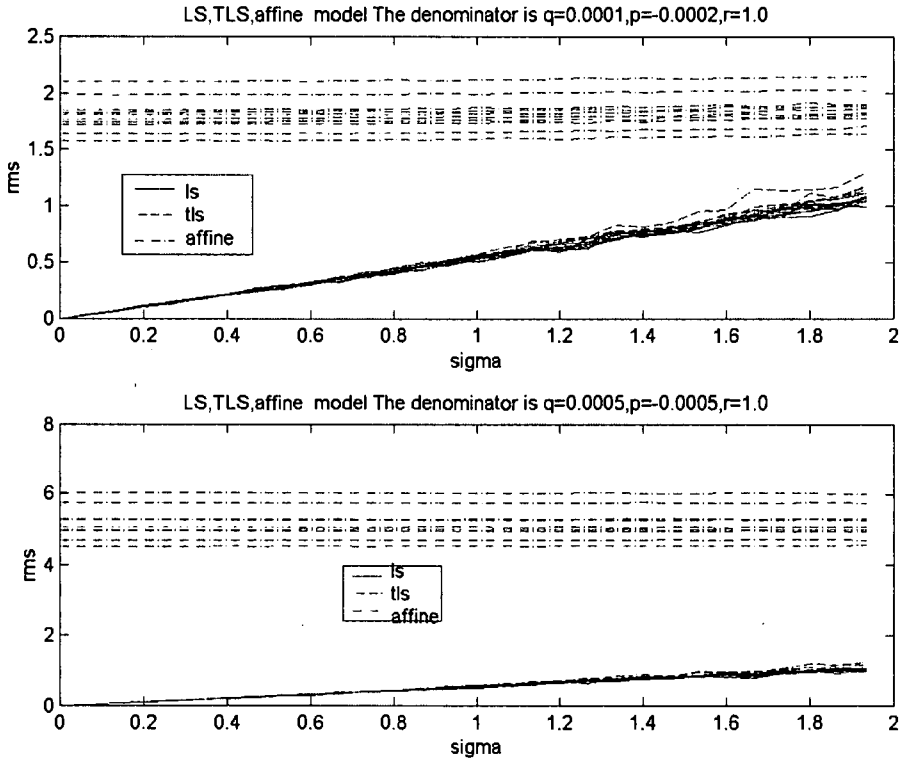
התוצאות (בכל האיורים) מראות:

1. ביצועי אלגוריתם LS דומים לביצועי ה – TLS ולא ניתן (באיורים) להבחין ביניהם.
2. כאשר אין רעש ($\sigma = 0$) אזי שיטת ה – LS (6.26) או ה – TLS (6.32) מניבות תוצאות מדוייקות. כשיש רעש, השגיאה פרופורציונית לעוצמת הרעש.
3. ככל שהמקדמים של המכנה a_7, a_8 גדולים יותר (לדוגמא, הגרף הראשון באיור

לעומת השני), ביצועי המודלים המקרבים גרועים יותר. ביצועי ה – TLS \LS לא הושפעו משינוי איברי המכנה.

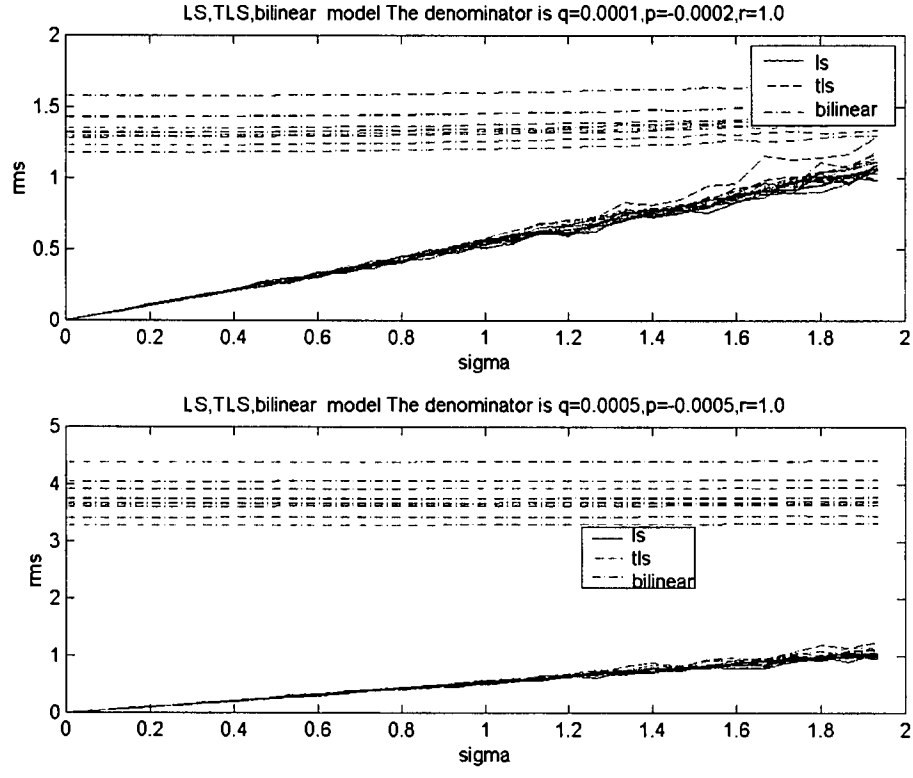
4. ביצועי המודל האפיני (איור 6.2) והמודל הבי-לינארי (איור 6.3). טובים פחות מביצועיו של המודל מסדר שני ((6.37), איור 6.4) ולכן המשך עבודה זו מתייחס אך ורק למודל מסדר שני.

כאשר אין רעש, המודל מסדר שני ((6.37)– איור 6.4) מניב ביצועים גרועים מזה של ה – LS או ה – TLS מכיוון שהינו מהווה קירוב בלבד של ההתמרה המישורית. אולם חשוב לציין שקירוב זה רגיש פחות לעוצמת הרעש, ולכן כאשר הרעש חזק ביצועיו טובים יותר משערוך של ההתמרה המישורית המדוייקת.



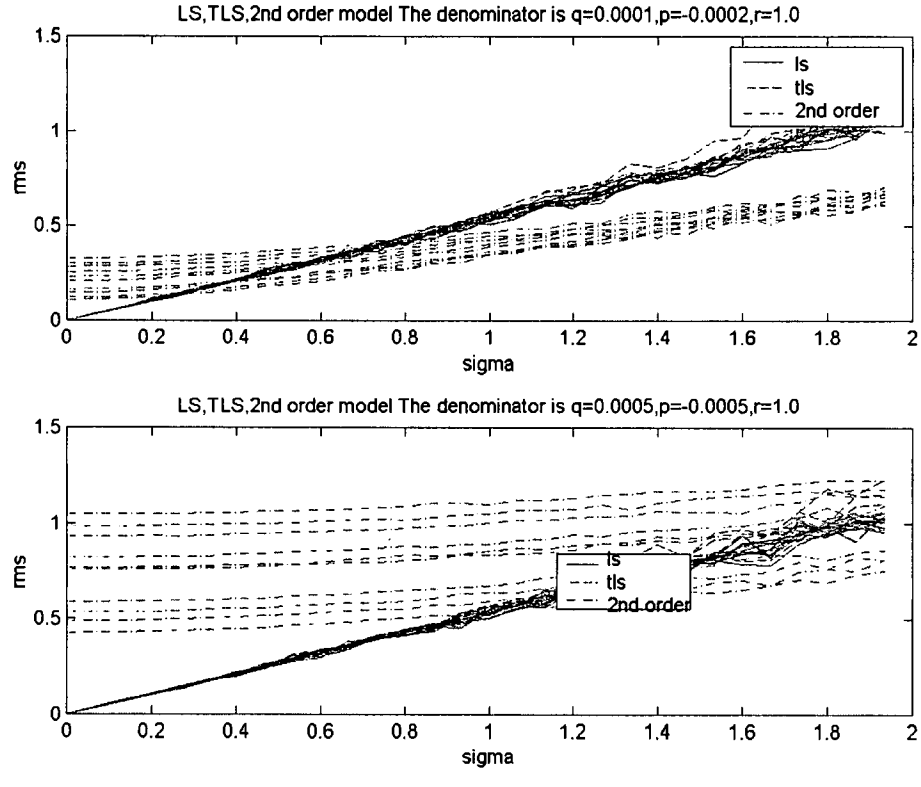
איור 6.2 התמרה אפינית

Fig 6.2: Affine model results



איור 6.3 - התמרה בי-לינארית

Fig 6.3: Bilinear model results



איור 6.4 – מודל מסדר שני

Fig 6.4: Second-order model results

6.5 אלגוריתם מוצע לשערוך פרמטרי התמרה מישורית ברצף תמונות

המודל המקורב מסדר שני של ההתמרה המישורית (6.37) הוכח כמשערך רובסטי של פרמטרי ההתמרה המישורית כאשר התנועה בין התמונות קטנה (כלומר, $a_7, a_8 \ll 1$). המקודד המוצע בעבודה זו מציע לייצג את הרצף באמצעות תמונת ייחוס ומידע התנועה בין תמונת הייחוס לבין כל תמונות הרצף. התנועה בין תמונת הייחוס לבין תמונות הרצף איננה בהכרח קטנה ולכן לא ניתן להשתמש במודל המקורב (6.37) בצורה ישירה לייצוג התנועה.

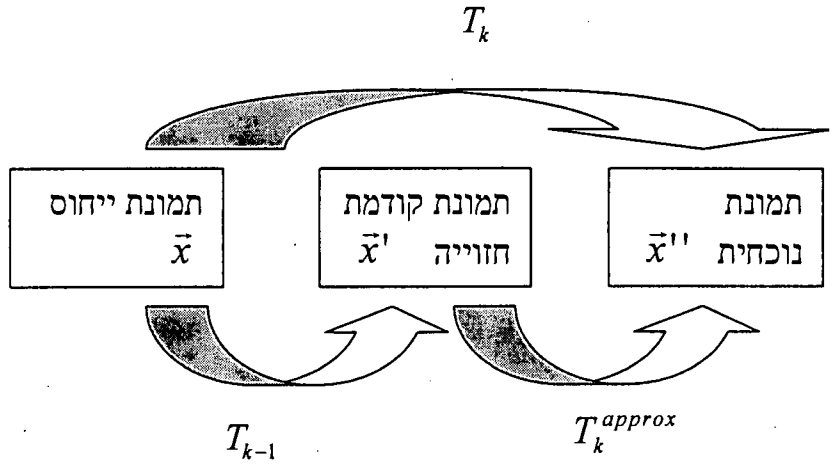
המסקנות בסעיף 4.4 קובעות ששערוך מדוייק יותר של מידע התנועה מתקבל באמצעות שרשור של שתי טרנספורמציות. הראשונה הינה בין תמונת הייחוס לבין התמונה הקודמת והשנייה בין התמונה הנוכחית לבין חיזוי התמונה הקודמת ע"י תמונת הייחוס (באמצעות ההתמרה הראשונה). מכיוון שלא ניתן לשרשר בצורה פשוטה בין ההתמרה המישורית לבין המודל המקורב מסדר שני (6.37) מוצעת השיטה הבאה. השיטה מתבססת על העובדה ששערוך המודל המקורב מבוצע בין תמונה נוכחית לחיזוי תמונה קודמת. ההנחה היא שבין תמונות סמוכות התנועה קטנה.

האלגוריתם המוצע מתואר בצורה כללית ומאפשר שימוש בכל התמרה מקורבת T_k^{approx} המניבה שערוך טוב בין תמונות עוקבות. עבודה זו התמקדה בעיקר בהתמרה מסדר שני (6.37) כהתמרה המקורבת.

אלגוריתם זה הינו החידוש העיקרי של עבודה זו . עבור תמונה נוכחית I_k, נבנה תחילה שערוך של התמונה $\hat{I}_{k-1} = T_{k-1}(I_0)$ באמצעות רגיסטרציה של תמונת הייחוס לפי ההתמרה המישורית שחושבה עבור המסגרת $k-1$ (ביחס לתמונת הייחוס).

בין $\hat{I}_{k-1}$ לבין I_k משוערך קירוב של ההתמרה המישורית T_k^{approx} , תוך שימוש באחת השיטות המקורבות (כגון (6.37)). מכיוון שבין I_k ל - I_0 קיימת תנועה רבה, וההתמרה המתארת תנועה זו הינה מישורית במדוייק ואיננה ניתנת לקירוב טוב באמצעות המודלים המקורבים יש לשלב את שתי הטרנספורמציות T_{k-1} ו - T_k^{approx} להתמרה מישורית אחת, T_k (איור 6.5).

הדבר מבוצע באמצעות העתקת מספר נקודות בקואורדינטות כלשהן בתמונת הייחוס $\vec{x}$ לפי $\vec{x}' = T_{k-1}(\vec{x})$, ולאחר – מכן העתקת התוצאה לפי $\vec{x}'' = T_k^{approx}(\vec{x}')$. מציאת T_k מתבצעת תוך פתרון בעיית LS למציאת פרמטרי התמרה מישורית בין $\vec{x}$ ו - $\vec{x}''$ (6.26).



איור 6.5 – אלגוריתם לשערוך התמרה מישורית
Fig. 6.5: Planar transformation estimation algorithm

אלגוריתם זה פותר את בעיית רגישות שערוך LS של פרמטרי התמרה המישורית לרעש מכיוון ששערוך ה – LS מבוצע על קואורדינטות מוחלקות (לפי T_k^{approx} הרגיש פחות לרעש, מכיוון שהמודל המקורב רובסטי יותר בתנאי רעש.

<u>הערות</u>

א. כאשר $k=1$ מתבצע בשלב הראשון שערוך מודל מקורב בין I_0 ל - I_1 . $\vec{x}'=T_k^{approx}(\vec{x})$ וחישוב ההתמרה המישורית מתבצע באמצעות פתרון בעיית LS למציאת פרמטרי התמרה מישורית בין $\vec{x}$ ו - $\vec{x}'$ (6.26).

ב. וקטור הקואורדינטות $\vec{x}$ (בד"כ כ – 15 נקודות) הינו אוסף ערכים (x,y) המחולקים על – פני התמונה כאשר הם אינם אמורים בהכרח להתאים לנקודות עניין בתמונה.

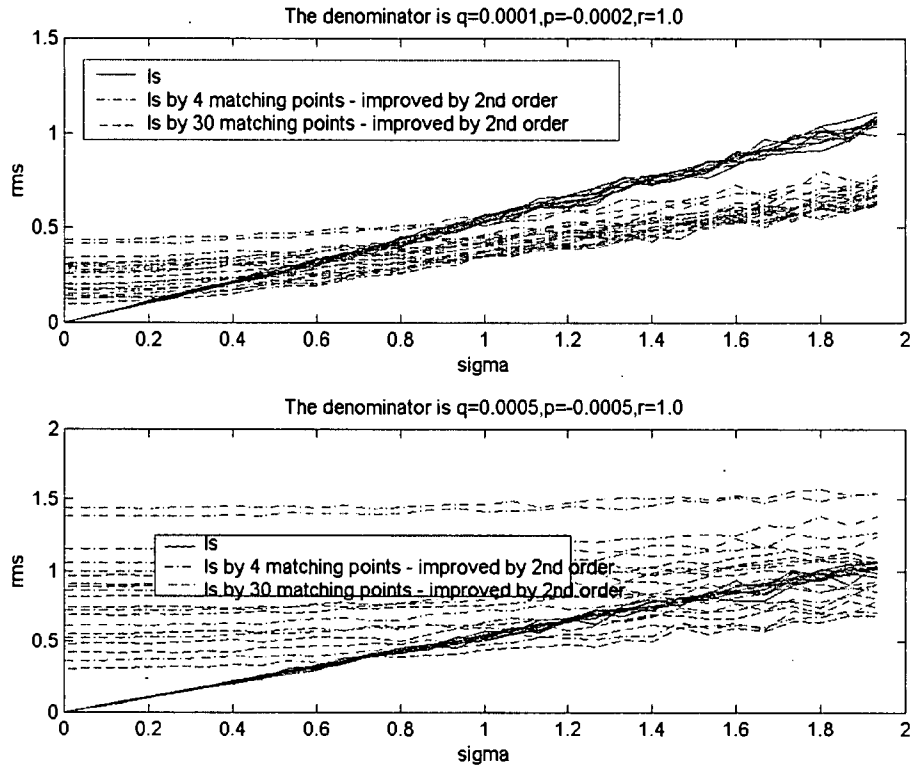
ג. תוצאות הביניים $\vec{x}'$ המתקבלים בשלב הראשון ו - $\vec{x}''$ המתקבלים בשלב השני הינם בייצוג float , ואינם קשורים למיקומי פיקסלים דיסקרטיים בתמונה. ולכן שערוך T_k המתבסס על קואורדינטות אלו יהיה מדוייק יותר כיוון שלא מתווסף רעש קוונטיזצייה לנקודות.

שיפור שערוך TLS\LS באמצעות מודל מקרב מסדר שני:

האלגוריתם המוצע בעבודה זו כולל שימוש בהתמרה מסדר שני על–מנת להחליק את נקודות ההתאמה , כך שבשלב השני יתבצע שערוך התמרה מישורית באמצעות TLS\LS (כזכור, כאשר עוצמת הרעש נמוכה הביצועים של ה TLS/LS טובים).

התוצאות המופיעות באיור (6.6) מתארות תהליך דו – שלבי: בשלב הראשון מבוצע שערוך באמצעות המודל מסדר שני ומשוערך מודל מקורב T_0 ובשלב השני משוערכת התמרה מישורית (LS) באמצעות $T_0(X,Y)$ ו – X,Y . התוצאות שהתקבלו מראות כי כאשר a_7,a_8 קטנים שאז המודל המקורב מדוייק דיו (הגרף הראשון באיור 6.6) השלב הראשון מצליח לשערך את מיקום הנקודות בצורה טובה, כך שמתקבל פתרון LS מדוייק. כאשר המודל המקרב איננו מדוייק דיו, אין טעם בהפעלת שני השלבים.

מכיוון שאחרי השלב השני, ידועה התמרה T_0, ושערוך השלב השני מתבסס על $T_0(X,Y)$ ו X,Y, נבדק האם רוב המידע הקיים ב - T_0 מוכל בארבע נקודות בלבד ולכן סביר ששערוך ההתמרה המישורית (בשלב השני) לא ישתפר אם השערוך ינצל יותר מארבע נקודות . בדיקה זו העלתה שמספר הנקודות בכל-זאת משפיע על התוצאות (יתכן שהדבר נובע מאי-הליניאריות של כל הטרנספורמציות), אולם לא התקבל ששימוש ביותר נקודות משפר בצורה משמעותית את איכות השערוך. התוצאות (איור 6.6) מראות שכאשר המודל המקורב מדוייק (כאשר a_7,a_8 קטנים - גרף ראשון) הייתה השפעת הגדלת מספר הנקודות זניחה. התוצאות השתנו בעקבות הגדלת a_7,a_8 אם כי לא התקבל שיפור עקבי עקב שימוש במספר רב יותר של נקודות.



איור 6.6 - שימוש במודל מסדר שני כאמצעי לשיפור שערוך התמרה מישורית

Fig. 6.6: Results of using second order model to improve planar transformation estimation

ביצועי האלגוריתם המוצע לקידוד רצף תמונות (להלן, אלגוריתם II) הושוו עם ביצועיו של האלגוריתם המבוסס על מסקנות סעיף 4.4 (להלן, אלגוריתם I). אלגוריתם זה מורכב משערוך התמרה מישורית באמצעות פתרון LS (שיטת TLS לא נבדקה מכיוון שלא נמצאו הבדלים משמעותיים בין תוצאות ה – TLS לבין תוצאות ה – LS, אולם TLS דורש כמות חישובים גדולה). שערוך התנועה בין תמונות הייחוס לבין תמונות הרצף מורכב משרשור של שתי טרנספורמציות מישוריות. ההתמרה הראשונה הינה בין התמונה הקודמת לתמונת הייחוס והשנייה בין חיזוי התמונה הקודמת (לפי תמונת הייחוס וההתמרה הראשונה) לבין התמונה הנוכחית. שרשור שתי טרנספורמציות מישוריות הינו מיידי:

אם סט הפרמטרים

$$X_2 = \frac{a_1 X_1 + a_2 Y_1 + a_3}{a_7 X_1 + a_8 Y_1 + 1}$$
$$Y_2 = \frac{a_4 X_1 + a_5 Y_1 + a_6}{a_7 X_1 + a_8 Y_1 + 1} \qquad (6.38)$$

מתאר את הקשר בין (X_1, Y_1) לבין (X_2, Y_2) ו –

$$X_3 = \frac{b_1 X_2 + b_2 Y_2 + b_3}{b_7 X_2 + b_8 Y_2 + 1}$$
$$Y_3 = \frac{b_4 X_2 + b_5 Y_2 + b_6}{b_7 X_2 + b_8 Y_2 + 1} \qquad (6.39)$$

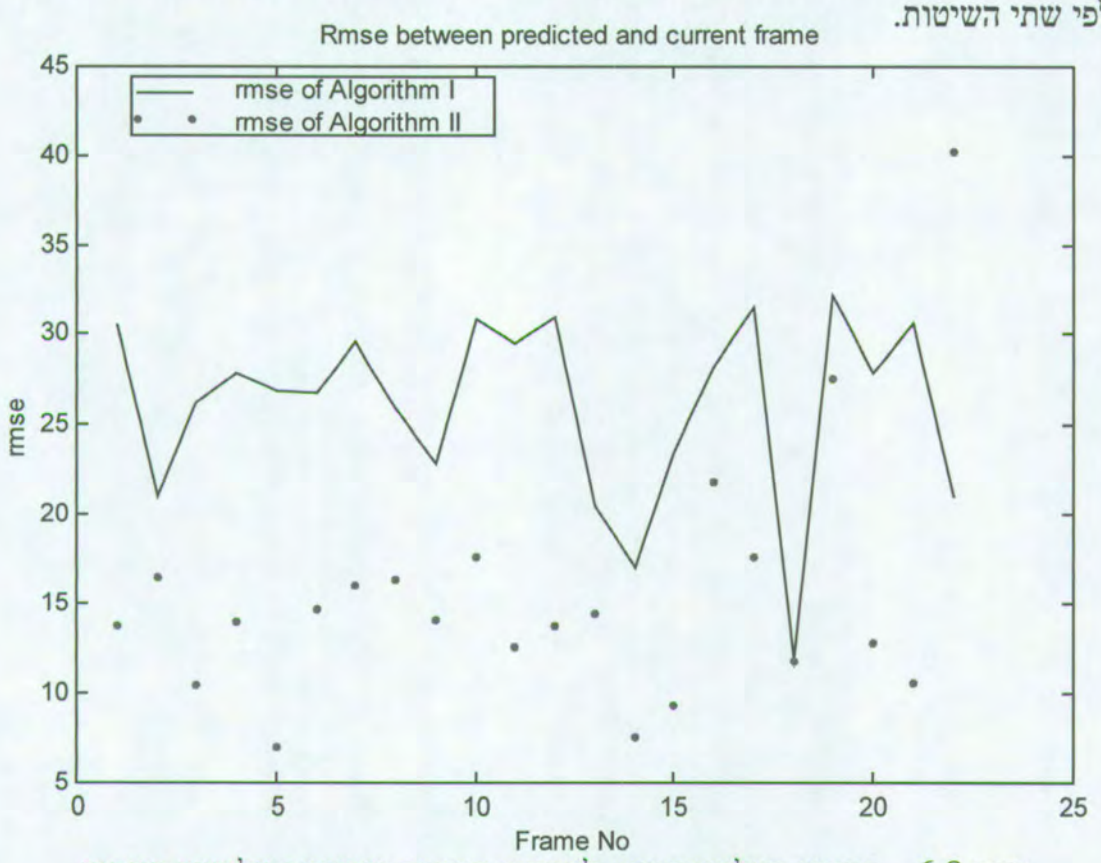
מתאר את הקשר בין (X_3, Y_3) לבין (X_2, Y_2) ו –
אזי מהצבת (6.38) לתוך (6.39) מתקבל

$$X_3 = \frac{(a_1 b_1 + a_2 b_4 + a_3 b_7) X_1 + (a_1 b_2 + a_2 b_5 + a_3 b_8) Y_1 + a_1 b_3 + a_2 b_6 + a_3}{(a_7 b_1 + a_8 b_4 + b_7) X_1 + (a_7 b_2 + a_8 b_5 + b_8) Y_1 + a_7 b_3 + a_8 b_6 + 1}$$
$$Y_3 = \frac{(a_4 b_1 + a_5 b_4 + a_6 b_7) X_1 + (a_4 b_2 + a_5 b_5 + a_6 b_8) Y_1 + a_4 b_3 + a_5 b_6 + a_6}{(a_7 b_1 + a_8 b_4 + b_7) X_1 + (a_7 b_2 + a_8 b_5 + b_8) Y_1 + a_7 b_3 + a_8 b_6 + 1} \qquad (6.40)$$

הביצועים נבדקו באמצעות רצף תמונות סינטתי ורצף תמונות אמיתי.
הרצף הסינטתי נוצר באמצעות הפעלת התמרה מישורית קבועה בין כל תמונות הרצף.
לדוגמא תמונה ראשונה ותמונה מספר 20 (איור 6.7):

איור 6.7 – רצף תמונות סינתטיות שנוצרו באמצעות התמרה מישורית
Fig. 6.7: Image Sequence created by a planar transformation

הגרף המצורף מציג את ה – rmse בין כל תמונה לבין חיזוייה באמצעות שערוך התנועה לפי שתי השיטות.



איור 6.8 – ביצועי האלגוריתמים לשערוך התמרה מישורית על רצף סינתטי
Fig. 6.8: Performance comparison between planar transformation estimation algorithms (synthetic sequence)

דוגמא נוספת לביצועי שתי השיטות מבוסס על הערכה איכותית של ביצועי השערוך. מכיוון שרצף ה – "Salesman" המסובב הינו רצף סינטטי המורכב מהפעלת התמרה מישורית על תמונה בודדת. אם יקוזזו מכל – אחת מתמונות הרצף התנועה, אזי אמורים לקבל את התמונה הראשונה. מיצוע כל התמונות לאחר קיזוז תנועה אמור להניב מדד איכותי לאיכות שערוך התנועה.

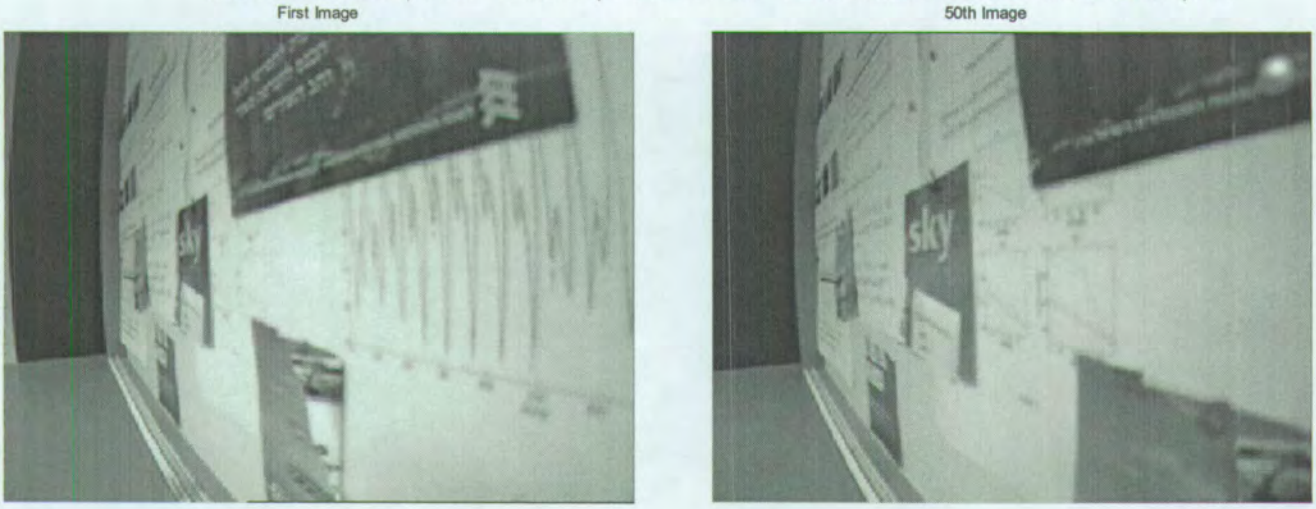
Averaged motion compensated image -Algorithm I

איור 6.9 – מיצוע תמונות הרצף לאחר קיזוז תנועה (20 תמונות) – אלגוריתם I (רצף סינתטי)
Fig. 6.9: Averaging motion compensated sequence images (20 images) – Algorithm I (synthetic sequence)

Averaged motion compensated image -Algorithm II

איור 6.10 – מיצוע תמונות הרצף לאחר קיזוז תנועה (20 תמונות) – אלגוריתם II (רצף סינתטי)
Fig. 6.10: Averaging motion compensated sequence images (20 images) – Algorithm II (synthetic sequence)

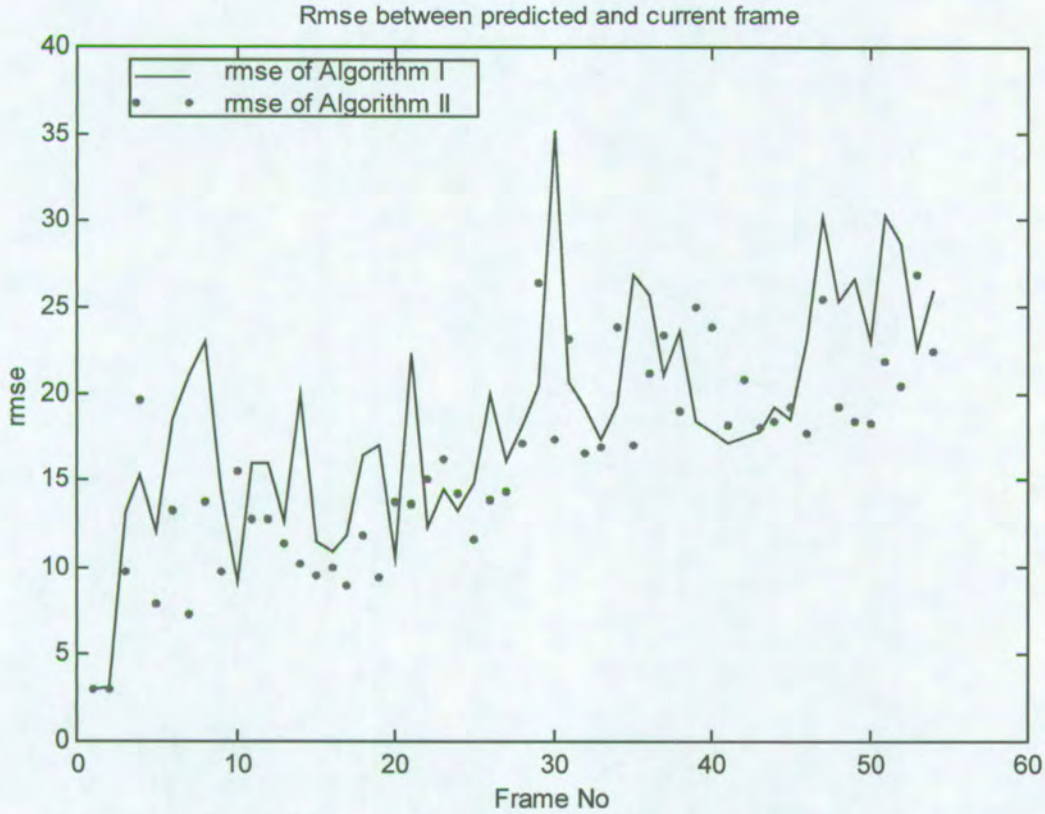
להלן התמונה הראשונה ותמונה מספר חמישים ברצף האמיתי שנלקח ממצלמה נעה:



איור 6.11 – רצף תמונות שצולמו ממצלמה נעה

Fig 6.11: Images from an image sequence taken by a moving camera

הגרף המצורף מציג את ה – rmse בין כל תמונה לבין חיזוייה באמצעות שערוך התנועה לפי שתי השיטות.



איור 6.12 – ביצועי האלגוריתמים לשערוך התמרה מישורית על רצף אמיתי

Fig. 6.12: Performance comparison between planar transformation estimation algorithms (real sequence)

Averaged motion compensated image -Algorithm I

איור 6.13 – מיצוע תמונות הרצף לאחר קיזוז תנועה (30 תמונות) – אלגוריתם I (רצף אמיתי)

Fig. 6.13: Averaging 20 motion compensated sequence images (30 images) – Algorithm I (real sequence)

Averaged motion compensated image -Algorithm II

איור 6.14 – מיצוע תמונות הרצף לאחר קיזוז תנועה (30 תמונות) – אלגוריתם II (רצף אמיתי)

Fig. 6.14: Averaging motion compensated sequence images (30 images) – Algorithm II (real sequence)

כפי שניתן לראות מתוצאות הפעלת האלגוריתמים השונים הן על הרצף הסינתטי והן על הרצף האמיתי, השיטה המשולבת של שרשור התמרה מישורית להתמרה מישורית עדיפה של שערוך התמרה מישורית בלבד.

פרק 7 שערוך פרמטרי תנועת מצלמה בסצינה מישורית

ראייה ממוחשבת עוסקת במיצוי מידע לגבי הסצינה המצולמת מתוך תמונות הסצינה. בדומה ליכולות ניתוח תנועה של מערכות ראייה אנושיות, מטרת הפענוח האוטומטי של תמונות הינה פענוח הסצינה: מיקום, אוריינטצייה, מהירות, מסלול וצורה של העצמים בתוך הסצינה.

פרק 2 סקר שיטות לשערוך תנועה דו-מימדית. דהיינו, בהינתן שתי תמונות יש למצוא את התנועה של תאי התמונה הראשונה כך שיותאמו בצורה הטובה ביותר לתאי התמונה השנייה. תנועה דו-מימדית זו הינה היטל של תנועה תלת-מימדית במרחב על מישור המצלמה, והיא משוערכת מתוך שתי התמונות עצמן.

שערוך תנועה תלת-מימדית הינו שערוך התנועה האמיתית התלת-מימדית של עצם במרחב בהתבסס על הטלת התנועה על מישור המצלמה.
ברור שמבנה הסצינה (דהיינו מרחק כל נקודה בסצינה ממישור המצלמה) משפיע על התמונה המוטלת על מישור המצלמה. מכיוון שבד"כ מבנה הסצינה (scene structure) איננו ידוע, יש לבצע בד"כ שערוך סימולטני של התנועה התלת-מימדית ומבנה הסצינה.

קיים מגוון רב של יישומים אפשריים לשערוך תנועה תלת-מימדית ומבנה סצינה. אחד השימושים הינו ניווט אוטונומי (ידיעת המצב היחסי של המצלמה ביחס לסצינה (מיקום ואוריינטציה) מאפשר ניווט אוטונומי של רובוטים), שימושים נוספים הינם דחיסת תמונות באמצעות שידור של תמונות ההפרשים לאחר קיזוז תנועה ובקרת כבישים (שערוך תנועת כלי רכב מאפשר שערוך של מידת העומס בכבישים).

חשוב להדגיש שהמצב בו המצלמה נעה והסצינה סטטית שקול למצב בו הסצינה (או עצם) נעה והמצלמה סטטית . כלומר, מבחינת התמונות המתקבלות אין הבדל מהותי בין שני התרחישים הללו.
מטרת השערוך, בשני המקרים, הינה לזהות את התנועה היחסית ומבנה הסצינה בהתבסס על רצף תמונות.

אין צורך לציין שבעיית שערוך תנועה ומבנה תלת-מימדיים מתוך רצף תמונות דו – מימדיות הינו ill-posed, כך שיתכן שאין פתרון יחיד, ללא מספר הנחות מוקדמות מקלות. לדוגמא, הנחות מוקדמות הינן הנחת תנועה קשיחה (כל העצם נע כגוף אחד – סיבוב תלת-מימדי, והזזה זהה – המרחק האמיתי שבין הנקודות התלת-מימדיות נשאר קבוע) או הנחת מודל מסויים לגבי מבנה העצם. ידוע שתנועת עצם קשיח ניתנית למידול באמצעות ששה פרמטרים (וקטור הזזה תלת-מימדי ומטריצת סיבוב בעלת שלוש דרגות חופש: ציר סיבוב תלת מימדי וזווית סיבוב). ניתן להניח לדוגמא, שהסצינה הינה מישורית, מישורית למקוטעין או שניתנת למידול כפונקציה ריבועית. שערוך מבנה הסצינה הינו מציאת העומק של כל נקודה או מציאת פרמטרי משטח הסצינה.

קיימות שתי גישות עיקריות לשערוך תנועה מתוך רצף תמונות[AgN88]:

- מציאת אוסף נקודות בולטות בתמונה (נקודות דו-מימדיות) שהינן היטל של מאפיינים בולטים בעצמים תלת-מימדיים כגון פינות, ושפות. נקודות אלו, ישרים או קווים עקומים, נבחרים מתוך שתי התמונות ומותאמים זה לזה בין שתי התמונות. באמצעות אילוצים הנובעים מתוך הנחות קשיחות העצם ניתן לקבל בד"כ סט של משוואות אי – ליניאריות (שבהן מוצבות נקודות העניין המותאמות), שמהן ניתן לקבל את פרמטרי התנועה.
- חישוב ה – optical flow: מציאת וקטור המהירות הדו-מימדי עבור רוב נקודות בתמונה. במקום להשתמש בשינויים הזמניים של רמת האפור (שינוי ערכי הפיקסלים בין תמונות עוקבות), ניתן להשתמש בשינויים הזמניים שבין תמונות התוצאה של הפעלת אופרטרים לוקלים מסוגים שונים כגון נגזרות מרחביות או חישוב קונטרסט. בכל מקרה התוצאה המתקבלת הינה מיפוי צפוף (dense) של התנועות הדו-מימדיות בתמונה. באמצעות שדה ה - optical flow ואילוצים נוספים (ו\או הנחות נוספות לגבי הסצינה) מחושבת התנועה היחסית בין המצלמה והסצינה.

כאשר סצינה מכילה מספר עצמים\תנועות, אזי יש לחלק את הנקודות הבולטות או את שדה ה - optical flow לאיזורים, כך שכל איזור יתאר תנועה אחת. את מבנה הסצינה התלת-מימדי יש לחשב מתוך איזורים אלו. ולכן לעיתים, ניתן לחשב את מבנה הסצינה והתנועה התלת-מימדית בנפרד. לדוגמא, באמצעות שתי מצלמות (stereo) ניתן לשערך את מבנה הסצינה (באמצעות triangulation) ללא שום תלות בתנועה עצמה. אולם בד"כ רצוי להשתמש במידע לגבי התנועה, אם קיים, על – מנת לשפר את דיוק שערוך מבנה הסצינה. באמצעות זיהוי העצמים השונים בסצינה (באמצעות זיהוי התנועות השונות בסצינה) ניתן לשערך בצורה מדוייקת יותר את מבנה הסצינה.

שערוך מדוייק של תנועת מצלמה עשוי לסייע בשימושים רבים. ואכן עבודות רבות עסקו בשערוך פרמטרי מצלמה. מכיוון שעבודה זו מתמקדת בעיקר בשערוך פרמטרי מצלמה כאשר התסריטים הינם מישוריים, או מישוריים למקוטעין (כאשר מספר המישורים הינו מצומצם יחסית), עיקר הסקירה תתמקד בתסריטים כאלו. אם כי תנתן סקירה קצרה על שיטות כלליות לשערוך פרמטרי תנועת מצלמה.

השיטות לשערוך פרמטרי התמרה מישורית התבססו על מציאת פרמטרים באמצעות התאמת נקודות (קואורדינטות) בפיקסלים. שיטה זו מדוייקת כאשר ההתמרה מיועדת להתאים תמונה אחת לתמונה שנייה, מכיוון שההתמרה המרחבית מעבירה קואורדינטה (פיקסלים) אחת למיקום (בפיקסלים) שונה.
מכיוון שפרמטרי התנועה הממשיים הינם גדלים פיזיקלים וגיאומטריים מדידים, חישוב פרמטרי התנועה איננו מתבצע בערכי פיקסלים, אלא מתבצע באמצעות הקואורדינטות האמיתיות במערכת הצירים של המצלמה - כלומר קואורדינטות במימדים של אורך.
המעבר מקואורדינטות בפיקסלים לקואורדינטות האמיתיות במישור המצלמה מתבצע באמצעות הפרמטרים הפנימיים של המצלמה(4.10).

7.1 סקר ספרות - שערוך פרמטרי תנועת מצלמה

7.1.1 שערוך פרמטרי תנועת מצלמה בהתבסס על שערוך optical flow.

הביטוי הכללי עבור תנועת המצלמה הרגעית הינה (עבור תנועות קטנות) [Ad85]:

$$\begin{bmatrix} u(X,Y) \\ v(X,Y) \end{bmatrix} = \begin{bmatrix} +(\frac{V_x}{z}+\Omega_y) - X\frac{V_z}{z} - y\Omega_z + X^2\Omega_y + XY\Omega_x \\ +(\frac{V_y}{z}+\Omega_x) + X\Omega_z - y\frac{V_z}{z} + XY\Omega_y - Y^2\Omega_x \end{bmatrix} \tag{7.1}$$

כאשר:

$u(X,Y), v(X,Y)$ הינן המהירויות של תאי התמונה בקואורדינטות (X,Y).

הוקטור $V=(V_x,V_y,V_z)$ הינו וקטור המהירות הקווית של המצלמה.

הוקטור $R=(\Omega_x,\Omega_y,\Omega_z)$ הינו המהירות הזוויתית של המצלמה.

z הינו מרחק נקודה (X,Y) מהמצלמה.

ניתן בד"כ לקרב באמצעות התמרה דו-מימדית יחידה את שדה התנועה הכללי. התמרה מקורבת זו מתקבלת תחת הנחות גיאומטריות מסויימות.
להלן מופיעות מספר דוגמאות:

א. הנחת תנועת משטח מישורי:

במקרה זה

$$\frac{1}{z} = -\frac{A}{C}X - \frac{B}{C}Y + \frac{1}{C} \Leftrightarrow z = AX + BY + C \tag{7.2}$$

כאשר מציבים משוואה זו ב – (7.1) מתקבל (ראה נספח ב'):

$$\begin{bmatrix} u(X,Y) \\ v(X,Y) \end{bmatrix} = \begin{bmatrix} a + bX + cY + gX^2 + hXY \\ d + eX + fY + gXY + hY^2 \end{bmatrix} \tag{7.3}$$

ב. כאשר המצלמה מרוחקת מאוד מהסצינה כלומר, סטיות הסצינה ממשטח מישורי קטנות יחסית למרחק מהמצלמה (7.3) עדיין הינו קירוב טוב [IrA98]. במקרה כזה, כש- $z \to \infty$, $\frac{T_x}{z}, \frac{T_y}{z}, \frac{T_z}{z} \to 0$ ולכן מקבלים את הביטוי המקורב

$$\begin{bmatrix} u(X,Y) \\ v(X,Y) \end{bmatrix} = \begin{bmatrix} \Omega_y - y\Omega_z + X^2\Omega_y - XY\Omega_x \\ -\Omega_x + X\Omega_x + XY\Omega_y + Y^2\Omega_x \end{bmatrix} \tag{7.4}$$

ג. משוואה (7.4) נכונה גם עבור המקרה בו המצלמה מסתובבת בלבד, כלומר ללא הזזה ($\|T\|_2 \approx 0$) או שההזזה זניחה ביחס למרחק המצלמה מהסצינה ($\|T\|_2 \ll \|z\|_2$).

בהינתן שדה התנועה, ניתן למצוא את הפרמטרים הדו-מימדיים (7.4) או (7.3).

קיימות בספרות שיטות רבות לשערוך שדה התנועה. למשל שיטות איטרטיביות המוצעות ב- ([IrA98],[IAH95]) שבאמצעות אופטימיזציית גאוס–ניוטון מבקשים למצוא את סט הפרמטרים a הממזער את

$$\varepsilon^2(a) = \sum_{(X,Y)\in A} ((I(X,Y,t) - I(X - u(X,Y,a), Y - v(X,Y,a), t-1))^2 \qquad (7.5)$$

כאשר:
a הינו סט הפרמטרים שיש לשערך.
$I(X,Y,t)$ הינו ערך פיקסל (X,Y) בזמן t.

אחת הסיבות לבעיתיות של שערוך פרמטרי תנועה תלת-מימדית כללית מתוך תמונות דו-מימדיות הינה הקושי להבדיל בין התנועה הסיבובית לבין התנועה הקווית של המצלמה ([DaN93]).

כאשר חלק מהתמונה נמצא על משטח כלשהוא, אזי מוכח ב - [IRP94], [IRP97] שכיוון שדה התנועה (optical flow) השיורי בין התמונה החזוייה (לפי פרמטרי התנועה הדו-מימדית של המשטח) לבין התמונה השנייה מרוכז בנקודה יחידה בתמונה. נקודה זו הינה היטל ווקטור ההזזה על מישור התמונה (FOE – Field Of Expansion) המוגדרת כ -

$$\vec{e} = (\frac{T_x}{T_z}, \frac{T_y}{T_z}) \quad .$$

בעבודות אלו מניחים שהסצינה ניתנת לתיאור כמשטח מישורי בקירוב (כך ששדה התנועה הינו (7.3)). בשלב הראשון ניתן לחשב את ווקטור ההזזה התלת-מימדי מתוך $\vec{e}$ (עד כדי קבוע) . השלב השני כולל שערוך פרמטרי שדה התנועה (7.3). מתוך פרמטרי התנועה(7.3) ניתן לשערך את פרמטרי הסיבוב התלת-מימדי (B.6) תוך הנחה שידוע ווקטור ההזזה התלת-מימדי.

7.1.2 שערוך פרמטרי תנועת מצלמה המבוסס על נקודות התאמה

אחת השיטות העיקריות למציאת פרמטרי המצלמה מבוססת על חילוץ "essential parameters" [ZaL],[TsH2] המקיימים את האילוץ

$$\varepsilon^2 = [X' \ Y' \ 1] E [X \ Y \ 1]^T = 0 \qquad (7.6)$$

כאשר

$$E=\begin{bmatrix} r_{21}dz\text{-}r_{31}dy & r_{22}dz\text{-}r_{32}dy_2 & r_{23}dz\text{-}r_{33}dy \\ r_{31}dx\text{-}r_{11}dz_1 & r_{32}dx\text{-}r_{12}dz & r_{33}dx\text{-}r_{13}dz \\ r_{11}dy_2\text{-}r_{21}dx & r_{12}dy\text{-}r_{22}dx & r_{13}dy\text{-}r_{23}dx \end{bmatrix}=\begin{bmatrix} e1 & e2 & e3 \\ e4 & e5 & e6 \\ e7 & e8 & e9 \end{bmatrix}$$

כאשר $r_{i,j},\ i,j=1,2,3$ - איברי מטריצת הסיבוב ו - $\vec{T}=[dx\ \ dy\ \ dz]^T$ הינו וקטור ההזזה.

לאחר שחושבה E, ניתן לקבל את פרמטרי התנועה בצורה מיידית ([TsH84b] , [ZaL93]) ע"י פירוק SVD (Singular Value Decomposition) של E:

אם U,V יוניטריות, כך ש –

$$E=U\cdot diag(\lambda_1,\lambda_2,\lambda_3)\cdot V^T \tag{7.7}$$

הרי מטריצת הסיבוב:

$$\vec{R}=\begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}=U\cdot\begin{bmatrix} 0 & \pm 1 & 0 \\ \mp 1 & 0 & 0 \\ 0 & 0 & s \end{bmatrix}\cdot V^T \tag{7.8}$$

כאשר $s=\det(U)\det(V)$

ו -

$$\vec{T}=\begin{bmatrix} dx \\ dy \\ dz \end{bmatrix}=\alpha\begin{bmatrix} \sigma_1^T\sigma_2/\sigma_2^T\sigma_3 \\ \sigma_1^T\sigma_2/\sigma_1^T\sigma_3 \\ 1 \end{bmatrix} \tag{7.9}$$

σ_i הינה שורה ה - i של E, ו - α הינו גורם נירמול כלשהוא.

רק פתרון אחד ל - $\vec{R}$ וסימן מתאים של α יניבו מרחקי המצלמה מהסצינה z ו - z' חיוביים ([ZaL93], [TsH84b]).

על - מנת לשערך את המטריצה E (מתוך (7.7)) הוצעו שיטות רבות (כגון [ZHH86] ,[Te95]).חסרונות שיטות אלה הינן רגישותן הגבוהה לרעש (כלומר, לשגיאות במיקום נקודות ההתאמה). כמו כן, המטריצה E מתאפסת כאשר וקטור ההזזה הינו זהותית אפס ולכן לא ניתן לשערך את E בסיטואציה זו.

ה – Fundamental Matrix (F, ראה נספח א') הינה מקרה כללי של המטריצה E, כאשר לא ידועים הפרמטרים הפנימיים של המצלמה (intrinsic parameters) (4.10). כאמור, ללא פרמטרים אלו לא ניתן לשערך את פרמטרי תנועת המצלמה. חרף זאת בוצעו מספר עבודות לשערוך ה – Fundamental Matrix. מתוך F ניתן לקבל מידע לגבי מבנה הסצינה המצולמת. מכיוון שלפי (A.5) ניתן לחשב בקלות את E מתוך F ראוי להזכיר שלושה מאמרי סיכום מקיפים [LuF96],[Zh98] ו- [ToM97] הסוקרים שיטות לשערוך F.

7.1.3 שערוך פרמטרי תנועת מצלמה בתסריטים מישוריים

ב - [TsH81],[THZ82] מציעים שתי שיטות לשערוך פרמטרי תנועת משטח מישורי מתוך שמונת הפרמטרים של ההתמרה המישורית. בספרות, פרמטרים אלו נקראים "pure parameters" (כזכור, המקרה בו המישור נע (והמצלמה סטטית) דומה מאוד למקרה בו מצלמה נעה מצלמת משטח מישורי).

המקרה הכללי ביותר של תנועה תלת-מימדית של עצם מתוארת ע"י מטריצת סיבוב תלת-מימדית R ווקטור הזזה תלת-מימדי $\vec{T}$. הקשר בין קואורדינטות העצם לפני ואחרי התנועה הינו:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = R \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \vec{T} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix} \tag{7.10}$$

כאשר $(x,y,z)^T$, $(x',y',z')^T$ הינם הקואורדינטות של העצם (במערכת הצירים של המצלמה הקבועה) לפני ואחרי התנועה. נקודות אלו מוטלות על מישור המצלמה כ:

$$X' = F_1 \frac{x'}{z'} \quad Y' = F_1 \frac{y'}{z'} \tag{7.11}$$

כאשר F_1 הינו אורך מוקד המצלמה.

כאשר המישור מתואר ע"י המשוואה $k_1 x + k_2 y + z = D$ אזי ההתמרה המישורית הינה [Te95]:

$$\begin{aligned} X' &= \frac{(r_{11} + T_x \frac{k_1}{D})X + (r_{12} + T_x \frac{k_2}{D})Y + r_{13}F_1 + \frac{T_x F_1}{D}}{(r_{31} + T_z \frac{k_1}{D})X + (r_{32} + T_z \frac{k_2}{D})Y + r_{33}F_1 + \frac{T_z F_1}{D}} \\ Y' &= \frac{(r_{21} + T_y \frac{k_1}{D})X + (r_{22} + T_y \frac{k_2}{D})Y + r_{23}F_1 + \frac{T_y F_1}{D}}{(r_{31} + T_z \frac{k_1}{D})X + (r_{32} + T_z \frac{k_2}{D})Y + r_{33}F_1 + \frac{T_z F_1}{D}} \end{aligned} \tag{7.12}$$

מטריצת סיבוב - R ניתן לביטוי כפונקצייה של שלושה משתנים:

(7.13)

$$R=\begin{pmatrix} n_1^2+(1-n_1^2)\cos\alpha & n_1n_2(1-\cos\alpha)-n_3\sin\alpha & n_1n_3(1-\cos\alpha)+n_2\sin\alpha \\ n_1n_2(1-\cos\alpha)+n_3\sin\alpha & n_2^2+(1-n_2^2)\cos\alpha & n_2n_3(1-\cos\alpha)-n_1\sin\alpha \\ n_1n_3(1-\cos\alpha)-n_2\sin\alpha & n_2n_3(1-\cos\alpha)+n_1\sin\alpha & n_3^2+(1-n_3^2)\cos\alpha \end{pmatrix}$$

המתארים סיבוב בזווית α סביב ציר סיבוב תלת-מימדי - $\vec{n}=(n_1,n_2,n_3), \|\vec{n}\|=1$.

לאחר שערוך פרמטרי ההתמרה המישורית (ראה פרק 6), (7.12) מניבה שבע משוואות בשבעה נעלמים. ב - [TsH81] הראו שלאחר מספר פעולות מתימטיות ניתן לקבל פולינום מסדר שישי של פרמטר אחד בלבד ($T_x'=T_x/T_z$). לאחר שפרמטר זה מחושב, שאר הפרמטרים מתקבלים בקלות. למרות שבאופן תיאורטי יתכנו ששה פתרונות ממשיים לפולינום זה, הסימולציות של המחברים הניבו לא יותר משני פתרונות.

ב – [THZ82] הוכחו מתימטית תוצאות הסימולצייה של [TsH81] , המראות כי קיימים עד שני פתרונות לשערוך פרמטרי התנועה מתוך פרמטרי ההתמרה המישורית. תיאור מפורט של המשוואות מופיע בסעיף 7.2.

ב - [TsH84a] הוכח שלהתאמת ארבע נקודות התאמה כפי שמתקבלות בשלוש תמונות (כלומר, צילום מישור משלוש זוויות שונות) קיים פתרון אחד בלבד של פרמטרי התנועה.

ב [HaK98] מוצע להשתמש בשיטת הקירוב האיטרטיבית Levenberg-Marquardt על מנת לחשב מתוך ההומוגרפייה (6.6) את פרמטרי התנועה ופרמטרי מישור הסצינה.

ניתן להשתמש בשתי מצלמות (שמרחקן זו מזו ידוע – סטריאו) לשם שיפור חישוב פרמטרי התנועה [AlH85], [HuU91].
שיטות נוספות לחישוב פרמטרי התנועה של משטח מישורי מוצגות ב – [Le88],[Le90].

7.1.4 שימוש במספר מישורים לצורך שערוך פרמטרי תנועת המצלמה

ב – [LuF93], [LuF96] הוצגו מספר שיטות לשערוך ה –Fundamental Matrix מתוך נקודות התאמה הנמצאות על יותר ממישור אחד. ניתן להראות שעבור מישור אחד מתקיים הקשר

$$H^TF+HF^T=0 \tag{7.14}$$

בין ההומוגרפיה H לבין ה –Fundamental Matrix F.

בהנתן H, (7.14) מניבה שש משוואות ליניאריות לשערוך F. על מנת לשערך את שמונת איברי F (עד כדי קבוע נרמול) ישנן שתי אפשרויות:

א. להשתמש בשש נקודות התאמה על מישור אחד ובשתי נקודות התאמה שאינן על המישור (לקבלת שתי משוואות $(kX,kY,k)F(lX',lY',l)^T = 0$ (6.20)).

ב. להשתמש במישורים נוספים לפתרון (7.14).

כמו – כן מוצעות שתי אופציות נוספות לשיפור השערוך:

א. שימוש באילוץ ש - $rank(F) = 2$ באמצעות ייצוג F באמצעות שבעה פרמטרים בלבד.

ב. באמצעות החלפה של התפקידים של שתי התמונות ואז (7.14) הינו

$$FH^{-1} + H^{-1T}F^T = 0 \qquad (7.15)$$

ניתן לבצע מזעור סימולטני של קריטריון סימטרי (7.14) ו – (7.15).

[SuA91] מציעים לשפר את השיטה שהוצעה ב – [TsH81] ע"י שימוש במספר תמונות (תחת ההנחה שהמהירות הסיבובית והקווית קבועות). המחברים הוכיחו קשר בין פרמטרי התנועה לבין מישור הסצינה. לאחר שערוך ראשוני של מטריצת הסיבוב (לפי [TsH81]) מבוצע חיפוש במרחב הפרמטרים התלת-מימדי של המהירות הזוויתית (באיזור של מטריצת הסיבוב). באמצעות מספר מסגרות ניתן למצוא את סט פרמטרי התנועה המתאים ביותר למישור אחד (תחת הנחת פרמטרי תנועה קבועים בכל רצף התמונות) וממזער את המרחק שבין פרמטרי התנועה.

מספר עבודות מציעות להשתמש במספר תמונות בבת – אחת לצורך חישוב פרמטרי תנועת המצלמה, כגון [MaI99] ו – [Ro96].

7.2 שערוך פרמטרי תנועת מצלמה באמצעות פרמטרי ההתמרה המישורית

בסעיף זה מוצג כיצד ניתן לשערך את פרמטרי תנועת המצלמה מתוך פרמטרי ההתמרה המישורית. מכיוון שעבודה זו מציגה שיטה לשערוך רובוסטי יותר של ההתמרה המישורית, אזי שערוך פרמטרי תנועת המצלמה אמור להיות אף הוא רובוסטי יותר.

השיטה לשיערוך פרמטרי תנועת המצלמה מבוססת על [THZ82]:

בהנתן סט פרמטרי התמרה מישורית:

$$X' = \frac{a_1X + a_2Y + a_3}{a_7X + a_8Y + 1}$$
$$Y' = \frac{a_4X + a_5Y + a_6}{a_7X + a_8Y + 1} \qquad (7.15)$$

המטריצה A תוגדר כ:

$$A = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & 1 \end{bmatrix}$$

אזי פירוק SVD (Singular Value Decomposition) של המטריצה A מוגדר

$$A = U \cdot diag(\lambda_1, \lambda_2, \lambda_3) \cdot V^T \tag{7.17}$$

כאשר $U = [u_1 | u_2 | u_3]$ ו - $= [v_1 | v_2 | v_3]$ מטריצות אורתוגנליות בגודל 3x3.
$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$ הינם הערכים העצמיים של המטריצה A.

ישנן שלוש אפשרויות לשערוך פרמטרי התנועה התלויים בערכים העצמיים:

I. $\lambda_1 > \lambda_2 > \lambda_3$ - מקרה זה מתאר מצב בו ניתן לתאר את התנועה כסיבוב סביב ציר העובר דרך המרכז ולאחריה הזזה בכיוון שונה מכיוון הנורמל למישור הסצינה. במקרה כזה ישנם שני פתרונות שונים אפשריים:

$$R = U \cdot \begin{bmatrix} \alpha & 0 & \beta \\ 0 & 1 & 0 \\ -s\beta & 0 & s\alpha \end{bmatrix} \cdot V^T \tag{7.18}$$

$$T = k(-\beta u_1 + (\frac{\lambda_3}{\lambda_2} - s\alpha)u_3)$$

$$[a\ b\ c]^T = \frac{1}{k}(\delta v_1 + v_3)$$

כאשר

$$\delta = \pm\sqrt{(\frac{\lambda_1^2 - \lambda_2^2}{\lambda_2^2 - \lambda_3^2})}, \alpha = \frac{\lambda_1 + s\lambda_3\delta^2}{\lambda_2(1+\delta^2)}, \beta = \frac{1}{\delta}(\alpha - \frac{\lambda_1}{\lambda_2}), s = \det(U) \cdot \det(V)$$

$[a\,b\,c]$ הינם פרמטרי המישור $ax + by + cz = 1$ ו - k הינו קבוע שרירותי שסימנו נבחר כך שמרחק המצלמה מהסצינה - $z > 0$ לכל נקודות הסצינה.
כאשר ישנם שני פתרונות אפשריים לפרמטרי התנועה ניתן לפסול פתרון אחד באמצעות אחת משתי השיטות הבאות:

א. באמצעות שלוש תמונות של אותו מישור (מזוויות שונות כמובן) יש לבחור את הפתרון שנותן קבועי מישור זהים בין התמונה הראשונה לשנייה ובין השנייה לשלישית.

ב. לפי נספח א' , שלושת הנקודות ההומוגניות T, $R \cdot X$ ו - X' נמצאות על מישור אחד (A.2). תכונה זו משותפת לכל נקודות הסצינה, כולל נקודות שאינן על המישור. הפתרון השגוי לא מקיים תכונה זו ולכן ניתן באמצעות נקודות שאינן על מישור הסצינה לזהות את הפתרון הנכון.

II. $\lambda_1 = \lambda_2 \neq \lambda_3$ - מקרה זה מתאר מצב בו ניתן לתאר את התנועה כסיבוב סביב ציר העובר דרך המרכז ולאחריה הזזה בכיוון הנורמל למישור הסצינה. במקרה כזה ישנו סט פרמטרים אחד בלבד:

$$
\begin{aligned}
R &= \frac{1}{\lambda_1} A + (\frac{\lambda_3}{\lambda_1} - s) u_3 v_3^T \\
T &= k(\frac{\lambda_3}{\lambda_1} - s) \\
[a\ b\ c]^T &= \frac{1}{k} v_3 \\
s &= \det(u) \cdot \det(V)
\end{aligned}
\tag{7.19}
$$

III. $\lambda_1 = \lambda_2 = \lambda_3$ במקרה זה התנועה הינה סיבוב בלבד סביב ציר דרך המרכז ולא ניתן לשערך את מישור הסצינה או את וקטור ההזזה:

$$
R = \frac{1}{\lambda_1} A \tag{7.20}
$$

ניתן לשערך את הפרמטרים הפיזיקלים, זווית וציר סיבוב מתוך מטריצת הסיבוב באמצעות [Te95]:

$$
\begin{aligned}
trace\{R\} &= 1 + 2\cos(\Delta\theta) \\
R - R^T &= 2\sin(\Delta\theta) \begin{pmatrix} 0 & -n_3 & n_2 \\ n_3 & 0 & -n_1 \\ -n_2 & n_1 & 0 \end{pmatrix}
\end{aligned}
\tag{7.21}
$$

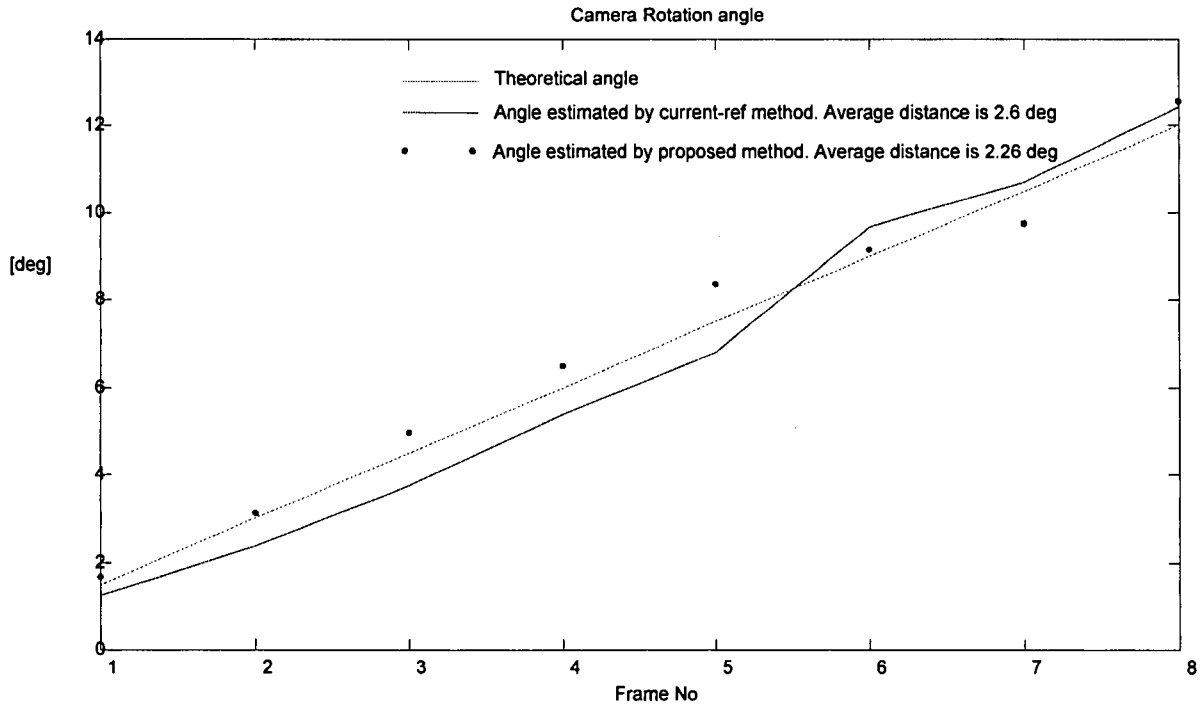
כדי לבדוק את שיפור שיערוך פרמטרי הסצינה. הורצו שני האלגוריתמים שתוארו בפרק 6 על הרצף הסינתטי הבא:

First Image

8'th Image

איור 7.1 – רצף תמונות סינתטיות שנוצרו באמצעות התמרה מישורית
Fig. 7.1: Image Sequence created by planar transformation

הרצף מורכב מסיבוב של 1.5° למסגרת. שני האלגוריתמים לשערוך פרמטרי ההתמרה המישורית (שפורטו בסעיף 6.4) הופעלו על רצף תמונות זה. באמצעות (7.18) ו – (7.21) שוערכה מטריצת הסיבוב וזווית הסיבוב. פרמטרי ההתמרה המישורית ששוערכו מתארים את התנועה בין תמונת הייחוס לתמונות הרצף ולכן זווית הסיבוב מתארת את הזווית בין תמונת הייחוס לתמונת הרצף.
להלן גרף המציג את הזווית האמיתית והמשוערכת לפי כל אחת משתי השיטות:



איור 7.2 - ביצועי אלגוריתם לשערוך פרמטרי תנועת מצלמה בתסריטים מישוריים

Fig. 7.2: Performance comparison between camera motion estimation algorithms in planar scenes.

כפי שניתן לראות סטיית התקן של שגיאת השערוך שהתקבלה בשיטה המוצעת בעבודה זו ($2.26°$) קטנה יותר לעומת השגיאה שהתקבלה באמצעות אלגוריתם הייחוס (אלגוריתם I בפרק 6) - $2.6°$.

פרק 8 סיכום ומסקנות

בעבודה זו מוצגים אלגוריתמים לשערוך תנועה גלובלית בתמונות רצף למטרות קידוד רצף תמונות. שערוך של התנועה הגלובלית מאפשר לייצג רצפי תמונות מסויימים באמצעות שידור של התמונה הראשונה ברצף ומידע התנועה בין כל תמונות הרצף לבין התמונה הראשונה. ייצוג כזה של הרצף המסויים הינו יעיל הרבה יותר משיטות דחיסה מקובלות כגון תקני ה – MPEG השונים.

קיימות שתי גישות העיקריות לשערוך פרמטרי תנועה. האחת, שערוך שדה ה - optical flow המבוסס על שערוך תנועה לכל פיקסל (או קבוצה קטנה סמוכה של פיקסלים). שיטת Block Matching (BM), המוצאת את וקטורי התנועה האופטימליים (במובן, לדוגמא, של מזעור השגיאה הריבועית הממוצעת) בין בלוקים ריבועיים הינה ואריאנט של שערוך optical flow. שיטת BM הינה השיטה הבסיסית לשערוך התנועה בתקני דחיסת הוידאו הנוכחיים. הגישה השנייה מבוססת על מיצוי של מספר מועט של מאפיינים בולטים, השייכים לנקודות בולטות של עצמים בתמונה, כגון פינות וכדומה. בעקבות מציאת ההתאמה בין מאפיינים בשתי תמונות ברצף, ניתן לשערך את תנועת המצלמה/האובייקט בין שתי התמונות. שיטה זו יעילה כאשר התנועה בין שתי התמונות הינה גלובלית בלבד וניתנת לתיאור באמצעות סט פרמטרים אחד לכל אזורי התמונה. בעבודה זו הושוו ביצועי שיטת BM לשיטה המבוססת על נקודות העניין ומתקבל שבתסריטים המתקבלים ממצלמה נעה, כלומר תנועה שאיננה הזזה בלבד, השיטה המבוססת על נקודות העניין מדוייקת יותר מאשר שיטת BM.

הוצג אלגוריתם לייצוג רצפי תמונות באמצעות שידור של התמונה הראשונה ברצף ומידע התנועה בין כל תמונות הרצף לבין התמונה הראשונה ברצף. אחד ממודלי התנועה שנבדקו הוא מודל אפיני בעל שישה פרמטרים. ההתמרה האפינית מסוגלת לתאר בצורה טובה סצינות בהן השינוי בעומק הסצינה קטן יחסית למרחק שבין המצלמה והסצינה.

פרמטרי התנועה הגלובלית מתקבלים באמצעות התאמה בין תמונות עוקבות תוך מציאת מאפיינים דומים בשתי התמונות ומציאת ההתמרה האופטימלית בין תמונות אלו. תהליך זה איננו מדוייק בהכרח ולכן סביר שקיזוז תנועה יעיל יותר (ולכן יחס דחיסה גבוה יותר) יתקבל באמצעות הפעלת אלגוריתם להחלקה וסינון של מידע התנועה הגולמי. הסינון שנבחן מסתמך על מודל דינמי של פרמטרי ההתמרה האפינית תוך שימוש במסנן קלמן לינארי . תוצאות עבודה זו מראות שלא התקבלו שיפורים מהותיים בעקבות הפעלת מסנן קלמן לינארי עקב חוסר היכולת לאפיין במדוייק את משוואות המודל והמדידה של פרמטרי התנועה.

בתסריטים מסויימים לא ניתן להניח שהסצינה מקבילה בקירוב למישור המצלמה (שאז ההתמרה האפינית הינה קירוב טוב של תנועת הסצינה). במקרה הכללי ביותר (תנועה תלת – מימדית כלשהי של מצלמה המצלמת משטח גיאומטרי כללי) לא קיים סט פרמטרים אחד המסוגל לתאר את תנועת הסצינה. כאשר הסצינה המצולמת הינה מישור ניתן לייצג את תנועת הסצינה באמצעות התמרה מישורית בעלת שמונה פרמטרים. שערוך ההתמרה המישורית הלא-לינארית איננו רובסטי בשיטות LS בתנאים בהן נקודות ההתאמה אינן ידועות במדוייק. כאשר התנועה בין התמונות קטנה ניתן להשתמש בשערוך מודל מקורב של ההתמרה המישורית. המודל המקורב לבדו אינו יעיל לייצוג תנועה בין תמונות שביניהן תנועה רבה (כמו בין תמונות הרצף לתמונה הראשונה ברצף). החידוש העיקרי המוצג בעבודה זו הינו השילוב בין שערוך המודל המקורב והמודל המדוייק לצורך שערוך רובסטי

של התמרה מישורית (ללא אילוצים לגבי גודל התנועה בין התמונות). בדיקות, הן על רצפי תמונות סינתטיות והן על רצפי תמונות אמיתיות, הראו ששערוך התמרה מישורית בשיטה המוצעת רובסטי יותר משיטות ה – LS המקובלות.

מתוך פרמטרי ההתמרה המישורית ניתן לחלץ את פרמטרי תנועת המצלמה, מטריצת הסיבוב ווקטור ההזזה. שערוך מדוייק יותר של פרמטרי ההתמרה המישורית מניב שערוך מדוייק יותר פרמטרי תנועת המצלמה, כפי שהודגם בעבודה זו.

כיווני המשך אפשריים הם:

א. בעבודה זו הוצגו עיקריו של מקודד לקידוד רצף תמונות. כלומר שידור תמונה ראשונה ומידע לגבי התנועה ברצף התמונות. פרטי המימוש עדיין לא נבחנו: שיטת קידוד התמונה הראשונה, צורת ייצוג של מידע התנועה ושידור מידע צד נוסף עבור תמונות בהן מידע התנועה בלבד (ותמונת הייחוס) איננו מסוגל לייצג את רצף התמונות היטב.

ב. שיטות שערוך התנועה שהוצגו בעבודה זו מתאימות לתרחישים בהן ניתן לתאר את הסצינה כולה באמצעות סט פרמטרים אחד. ישנם תרחישים בהן נדרשים יותר מסט פרמטרים אחד לייצוג הסצינה. לדוגמא, כאשר המצלמה מתקרבת לסצינה המצולמת, צפוי שיהיה צורך לבצע סגמנטציה של התנועה, כלומר נדרש לחלק את התמונה לאזורים, כך שלכל איזור בתמונה יתאים סט אחר של פרמטרי תנועה.

נספח א' ראיית סטריאו

שערוך תנועה בין שתי תמונות שהתקבלו ממצלמה נעה דומה למצב בו שתי מצלמות מצלמות איזור משותף (ראיית סטריאו) . שיטות לשערוך האוריינטציה בין שתי מצלמות מאפשרות למעשה לשערך את פרמטרי תנועת המצלמה בין שתי התמונות (פרק 7) .

בספרות מקובל להשתמש במונח ראיית סטריאו כיכולת לשערך את המרחק והמבנה של הסצינה תוך שימוש בלפחות שתי תמונות הנלקחות מזוויות שונות. לאחר שנמצאה התאמה בין נקודות בשתי התמונות, היכולת לשערך את הסצינה תלוייה בכמות המידע האפריורי הידוע בזמן החישוב.

המקרים החשובים הינם[TrV98]:

1. כאשר ידועים במדוייק הפרמטרים הפנימיים ($M_{\text{int}} = f(F,s,o)$ - אורך מוקדי המצלמות, כיול הקוארדינטות במערכת הצירים של המצלמות לפיקסלים ותיקון היסט בין הקוארדינטות במערכת הצירים לבין הפיקסלים) והחיצוניים ($M_{ext} = f(R,T)$ - ידיעה מדוייקת של הסיבוב וההזזה בין שתי המצלמות) ניתן לשערך במדוייק את הסצינה ע"י טריאנגולצייה.
2. כאשר ידועים הפרמטרים הפנימיים של המצלמה ($M_{\text{int}} = f(F,s,o)$) בלבד ניתן לשערך את הסצינה עד כדי קבוע נרמול. ישנן שיטות לשערך את פרמטרי המצלמה החיצוניים (הסיבוב וההזזה בין שתי המצלמות) מתוך התאמת נקודות, אולם לא ניתן לדעת אם השוני בסצינה הנראית בין שתי התמונות נגרם מכיוון שהמרחק בין מוקדי המצלמות קטן בקבוע או שמרחק הסצינה גדול באותו הגורם.
3. כאשר לא ידועים מראש לא הפרמטרים החיצוניים ולא הפרמרים הפנימיים ניתן לשערך את הסצינה עד כדי מטריצת מודל מצלמה (4.10,4.11) [TrV98].

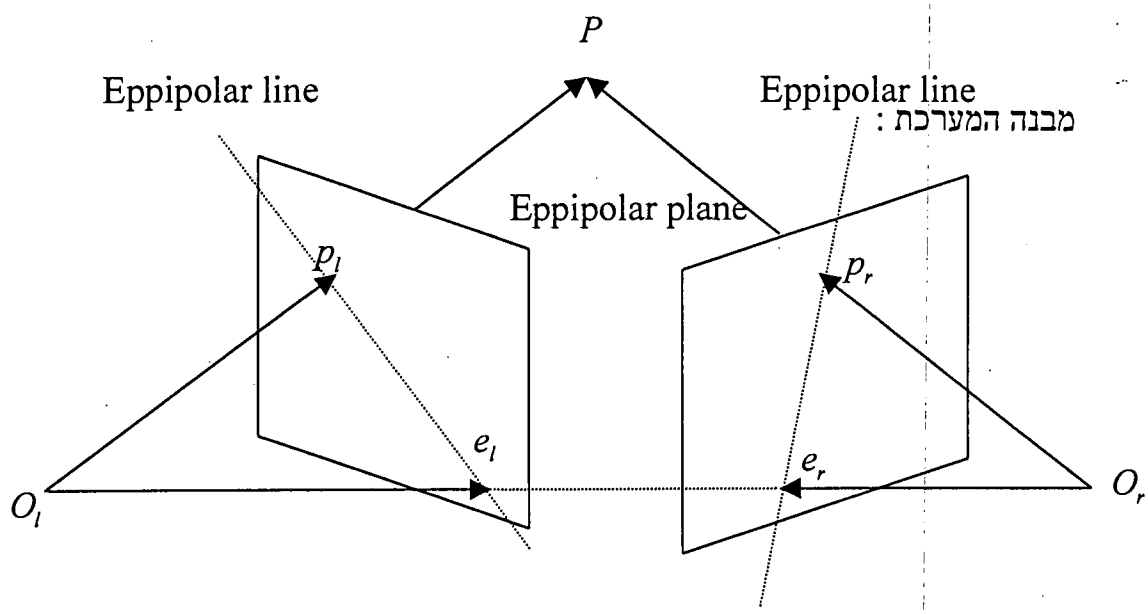
כאשר המצלמה נעה לא ידועים בד"כ מראש הפרמטרים החיצוניים שבין שתי המצלמות ולכן לא ניתן לשערך את הסצינה בצורה מדוייקת.

מכיוון שחלק מהשיטות לשערוך פרמטרים שדווחו בעבודה זו מנצלות את הגיאומטריה והמונחים של מערכת סטריאו (הידועה גם כ – Epipolar Geometry) מובאת בנספח זה סקירה קצרה של הנושא.

Epipolar Geometry

המערכת מורכבת משתי מצלמות pinhole שמוקדיהן ב - O_l ו - O_r, מישורי המצלמות הינם הינם π_l ו - π_r נקודה P מוגדרת במערכות הצירים של שתי התמונות כ - $P_l = (X_l, Y_l, Z_l)$ ו - $P_r = (X_r, Y_r, Z_r)$ ומוטלת על מישור המצלמה כ - $p_l = (x_l, y_l, f_l)$ ו - $p_r = (x_r, y_r, f_r)$.

מבנה המערכת :



איור A.1 – מערכת צילום סטריאו
Fig. A.1: Stereo cameras configuration

שתי מערכות הצירים שמרכזן O_r ו - O_l קשורות זו לזו באמצעות הפרמטרים החיצוניים של המצלמות (מטריצת סיבוב R ווקטור הזזה תלת – מימדי $T = O_r - O_l$), כך שהקשר בין נקודה כלשהיא במרחב P בשתי מערכות הצירים הינו:

$$P_r = R(P_l - T) \tag{A.1}$$

קונפיגורציה כזו של שתי מצלמות קרויה epipolar geometry משום שהנקודות בהן חותך הישר המחבר את מרכזי המצלמות נקראות epipoles (e_r ו - e_l). ה – epipoles הינם תמונות מרכזי המצלמות בתמונה הנגדית.

החשיבות המעשית של ה - epipolar geometry נובעת מכיוון שהמישור $P - O_r - O_l$ (epipolar plane) חותך כל תמונה בקו epipolar line . מתוך איור (A.1) ניתן לראות שהקרן $P - p_l$ מוטלת על ה - epipolar line בתמונה הימנית ללא תלות במיקום P ולכן ההתאמה של p_l בתמונה הימנית נמצאת על - epipolar line בתמונה הימנית. אילוץ זה ידוע כ - epipolar constraint. חשוב לציין שאילוץ זה קיים בצורה זהה גם עבור התמונה השנייה

משמעות ה - epipolar constraint שחיפוש נקודת התאמה בתמונה השנייה הינו חיפוש חד – מימדי לאורך קו (epipolar line).

מושג חשוב נוסף הינו ה – epipole (e_l ו - e_r) המהווים את היטל מרכז המצלמה השנייה על מרכז המצלמה הראשונה. ניתן לראות בקלות שכל ה - epipolar lines של המצלמה עוברים דרך ה - epipole של המצלמה – משום שכל נקודה במרחב P' יוצרת מישור $P'-O_l-O_r$ החותך בשתי נקודות כל מישור מצלמה (בנקודה p' - המהווה את היטל P' על מישור המצלמה ובנקודה e שהינה ה – epipole).

אם O_r יסומן כראשית הצירים $O_r=\vec{0}$ אזי האילוץ על קיום המישור $P_l-O_l-O_r$ שבו נמצאים הוקטורים P_l , $O_r-O_l=T$ ו - P_l-T מוגדר כ:

$$(P_l-T)^T T\times P_l=0$$

תוך שימוש ב – (A.1) מקבלים

$$(R^T P_r)^T T\times P_l=0 \qquad \text{(A.2)}$$

מכיוון שמכפלה וקטורית ניתנת לכתיבה באמצעות הכפלה במטריצה חסרה (rank deficient matrix) המוגדרת כ –

$$S=T_x^T=[T_1\ T_2\ T_3]_x^T=\begin{bmatrix} 0 & -T_3 & T_2 \\ T_3 & 0 & -T_1 \\ -T_2 & T_1 & 0 \end{bmatrix}$$

אזי (A.2) ניתנת לכתיבה כ - $P_r^T E P_l = P_r^T R S P_l = 0$

מכיוון ש - $rank(S)=2$ ו - $rank(R)=3$ אזי $rank(E)=2$.

באמצעות חלוקה של (A.2) ב - $Z_l Z_r$ ניתן לקבל את הקשר הידוע בין קורדינטות במישור התמונה לפרמטרים החיצוניים:

$$p_r^T E p_l = 0 \qquad \text{(A.3)}$$

והמטריצה E הידועה כ essential matrix מתארת את הקשר בין הפרמטרים החצוניים (extrinsic parameters) - סיבוב והזזה בין שתי המצלמות לבין ב – epipolar constraint. כמו כן מטריצה זו הינה המיפוי בין נקודות בתמונה לבין ה- epipolar line בתמונה השנייה.

ניתן להראות [TvR98] שה- (epipolar line) u_r העובר דרך p_r ו - e_r הינו

$$u_r = E p_l \qquad \text{(A.4)}$$

משוואה (A.3) ניתנת לרישום כ - $p_l E^T p_r^T = 0$ ולכן $u_l = E^T p_r$.

חשוב לציין שהקואורדינטות p הינן קואורדינטות במערכת הצירים של התמונה ומכיוון שנקודות ההתאמה הינן בערכי פיקסלים יש לבצע התאמה בין קואורדינטות לבין פיקסלים לפי $\tilde{p} = M_{\text{int}} p$ (4.10) ומתקבל ש

$$\tilde{p}_r^T F \tilde{p}_l = 0 \qquad \text{(A.5)}$$

$F = M_{\text{int}}^{-T} E M_{\text{int}}^{-1}$.

המטריצה F ידועה כ – fundamental matrix מכילה מידע הן לגבי הפרמטרים הפנימיים והן על הפרמטרים החיצוניים של המצלמה.

$rank(F) = 2$ מכיוון ש - $rank(E) = 2$ ו $rank(M_{\text{int}}) = 3$.

בצורה דומה ל – (A.4) ניתן לקבל את ה – epipolar line הקשור לנקודה $\tilde{p}_l$, דהינו

$$\tilde{u}_r = F \tilde{p}_l \qquad \text{(A.6)}$$

ישר $l = (a, b, c)$ בתמונה העובר דרך נקודה הומוגנית $m = [ku, kv, k]^T$ מקיים $l^T \cdot m = m^T \cdot l = 0$ (הכפלה של l בכל סקלר כלשהוא מניבה את אותו הישר הדו-מימדי, לכן ישר דו-מימדי מוגדר באמצעות וקטור הומוגני תלת-מימדי) .

מתוך ההסבר לעיל ניתן לראות שה – (epipolar line) $\tilde{u}_r$ העובר דרך ה – epipole $\tilde{e}_r$ מקיים

$$\tilde{u}_r^T \tilde{e}_r = \tilde{p}_r^T F^T \tilde{e}_r = 0 \qquad \text{(A.7)}$$

מכיוון ש (A.7) נכון עבור כל נקודה $\tilde{p}_r^T$ משום שה- epipole $\tilde{e}_r$ חותך כל – epipolar line. לכן:

$$F^T \tilde{e}_r = 0 \qquad \text{(A.8)}$$

בצורה דומה ניתן להראות ש

$$F \tilde{e}_l = 0 \qquad \text{(A.9)}$$

מכיוון ש $rank(F) = 2$ אזי הפתרון ל- (A.8) ו (A.9) הינו הוקטור העצמי המתאים לערך העצמי אפס של F^T ו - F בהתאמה.

נספח ב' התמרה מסדר שני

נספח זה מכיל פיתוח מפורט של התמרה מסדר שני מכיוון שמודל זה הינו הבסיס לאלגוריתם לשערוך פרמטרי התמרה מישורית המוצע בעבודה זו (סעיף 6.3).

בהנתן מצלמה המסתובבת בזוויות קטנות לפי מטריצת סיבוב R ונעה לפי וקטור הזזה T אזי נקודה במרחב $\vec{P}=(x,y,z)^T$ נעה במהירות ([Te95])

$$\begin{pmatrix}\dot{x}\\ \dot{y}\\ \dot{z}\end{pmatrix}=\begin{bmatrix}0 & -\Omega_3 & \Omega_2\\ \Omega_3 & 0 & -\Omega_1\\ -\Omega_2 & \Omega_1 & 0\end{bmatrix}\begin{pmatrix}x\\ y\\ z\end{pmatrix}+\begin{bmatrix}V_1\\ V_2\\ V_3\end{bmatrix} \tag{B.1}$$

כאשר $\Omega=(\Omega_1,\Omega_2,\Omega_3)^T$ הינה המהירות הזוויתית , ו - $V=(V_1,V_2,V_3)^T$ הינה המהירות הקווית.

היטל הנקודה $\vec{P}$ במישור התמונה הינו

$$X=F_1\frac{x}{z} \quad Y=F_1\frac{y}{z} \tag{B.2}$$

F_1 הינו אורך מוקד המצלמה.

מהירות של הנקודה במישור המצלמה (בצירוף (B.2)) הינה

$$\begin{aligned}\dot{X}&=v_x=F_1\frac{z\cdot\dot{x}-x\cdot\dot{z}}{z^2}=F_1\frac{\dot{x}}{z}-X\frac{\dot{z}}{z}\\ \dot{Y}&=v_y=F_1\frac{z\cdot\dot{y}-y\cdot\dot{z}}{z^2}=F_1\frac{\dot{y}}{z}-Y\frac{\dot{z}}{z}\end{aligned} \tag{B.3}$$

הכנסה של (B.1) לתוך (B.3) וייצוג (x,y) באמצעות (X,Y,F_1,z) מניבה

$$\begin{aligned}v_x&=F_1(\frac{V_1}{z}+\Omega_2)-\frac{V_3}{z}X-\Omega_3Y-\frac{\Omega_1}{F_1}XY+\frac{\Omega_2}{F_1}X^2\\ v_y&=F_1(\frac{V_2}{z}-\Omega_1)+\Omega_3X-\frac{V_3}{z}Y+\frac{\Omega_2}{F_1}XY-\frac{\Omega_1}{F_1}Y^2\end{aligned} \tag{B.4}$$

מכיוון שההנחה היא שנקודות הסצינה נמצאות על מישור, אזי

$$k_1 x + k_2 y + z = D \tag{B.5}$$

הצבה של (B.5) לתוך (B.4) מניבה:

$$\begin{aligned} v_x &= F_1(\frac{V_1}{D}+\Omega_2)-(F_1\frac{V_1k_1}{D}+\frac{V_3}{D})X-(F_1\frac{V_1k_2}{D}+\Omega_3)Y \\ &+(\frac{V_3k_1}{D}+\frac{\Omega_2}{F_1})X^2+(\frac{V_3k_2}{D}-\frac{\Omega_1}{F_1})XY \\ v_y &= F_1(\frac{V_2}{D}-\Omega_1)-(F_1\frac{V_2k_1}{D}-\Omega_3)X \\ &-(F_1\frac{V_2k_2}{D}+\frac{V_3}{D})Y+(\frac{V_3k_1}{D}+\frac{\Omega_2}{F_1})XY+(\frac{V_3k_2}{D}-\frac{\Omega_1}{F_1})Y^2 \end{aligned} \tag{B.6}$$

או לחילופין

$$\begin{aligned} v_x &= b_1X+b_2Y+b_3+b_7X^2+b_8XY \\ v_y &= b_4X+b_5Y+b_6+b_7XY+b_8Y^2 \end{aligned} \tag{B.7}$$

עתה נותר להוכיח כי

$$\begin{aligned} v_x &= \dot{X} = \frac{X'-X}{\Delta t} \\ v_y &= \dot{Y} = \frac{Y'-Y}{\Delta t} \end{aligned} \tag{B.8}$$

עבור Δt סופי ולכן יתקבל (6.37)

$$\begin{aligned} X' &= b_1X+b_2Y+b_3+b_7X^2+b_8XY \\ Y' &= b_4X+b_5Y+b_6+b_7XY+b_8Y^2 \end{aligned} \tag{B.9}$$

הוכחה של (B.8):

בזווית סיבוב קטנות, מטריצת הסיבוב R ניתנת לקירוב באמצעות [Te95]:

$$R \approx \begin{bmatrix} 1 & -\Delta\phi & \Delta\psi \\ \Delta\phi & 1 & -\Delta\theta \\ -\Delta\psi & \Delta\theta & 1 \end{bmatrix} \tag{B.10}$$

θ,ψ,ϕ הינם זוויות הסיבוב האוקלידיות סביב הצירים X,Y,Z בהתאמה.
הצבה של (B.10) ו – (4.15) לתוך (B.8) ($F_1 = 1$ לצורך הפשטות) מניבה

$$\frac{X'-X}{\Delta t} = \frac{X+\Delta X}{\Delta t(1+\Delta z)} - \frac{X}{\Delta t}$$

$$\Delta X = -\Delta\phi\cdot Y + \Delta\psi + \frac{dx}{z} \qquad \text{(B.11)}$$

$$\Delta z = -\Delta\psi\cdot X + \Delta\theta\cdot Y + \frac{dz}{z}$$

תוך שימוש בקירוב $\frac{1}{1+x} \approx 1-x, x << 1$,

$$\frac{X+\Delta X}{1+\Delta z} \approx (X+\Delta X)\cdot(1-\Delta z) \approx X+\Delta X - X\cdot\Delta z \qquad \text{(B.12)}$$

הכנסת (B.12) לתוך (B.11) תוך הנחה ש - $\Delta z << 1, \Delta X\cdot\Delta z \to 0$.

$$\frac{X'-X}{\Delta t} = \frac{\Delta X}{\Delta t} - X\frac{\Delta z}{\Delta t} = \frac{-\Delta\phi\cdot Y + \Delta\psi + \frac{dx}{z} - X(-\Delta\psi\cdot X + \Delta\theta\cdot Y + \frac{dz}{z})}{\Delta t} \qquad \text{(B.13)}$$

$$= -\Omega_3 Y + \Omega_2 + \frac{V_1}{z} + X\cdot(\Omega_2 X - \Omega_1 Y - \frac{V_3}{z})$$

שהינו זהה ל – (B.6) כאשר $F_1 = 1$ וכך הוכח השויון (B.8). הוכחה דומה מאוד מתקבלת עבור v_y .

References מקורות ספרות

[Ad85] Adiv G., "Determining Three-Dimensional Motion and Structure From Optical Flow Generated by Several Moving Objects", *IEEE Trans*. PAMI, Vol. 7, No. 4, pp. 384-401, July 1985.

[AgN88] Aggarawal K. & Nandhakumar N., "On the Computation of Motion from Sequence of Motion", Proc. Of the IEEE, Vol. 76, No. 8, August 1988.

[AlH85] Aloimonos J. & Herve J.Y., "Correspondenceless Stereo and Motion: Planar Surface", *IEEE Trans*. PAMI, Vol. 12, No. 5, pp. 504-510, July 1985.

[An87] Anandan P. ,"Measuring Visual Motion for Image Sequences", Coins Technical Report 87-21,Uni. Of Massachusetts at Amherst, 1987.

[AnM79] B.O.Anderson & J.B.Moore, "Optimal Filtering", Prentice Hall, 1979.

[Br92] Brown, l. G., "Image Registration Technique", *ACM computing survey*, Vol. 24, No. 4, Dec. 1992.

[Ch92a] Chen, L.H., "A New Approach for Feature Point Classification, Aggregation, and Description", *Int. J. of Pattern Recognition and artificial Intelligence*, Vol. 6, No. 5, pp. 849-871, 1992.

[Ch92b] Chen, L.H., "A Two-Phase Area Level Line/Edge Detector", *Pattern Recognition*, Vol. 25, No. 1,pp. 55-63 1992.

[CrG94] Christopher, R. J. & Gregson P. H. ,"Detecting Corners Using the 'Patchy' Hough Transform", Conf. Proc. Canadian Conference in Electrical and Computer Engineering, pp. 576-579, 1994.

[DaN93] Daniilidis K. & Nagel H.H., "The Coupling of Rotation and Translation in Motion Estimation of Planar Surfaces", ICVPR, Los Alamitos ,Ca.,USA, pp. 188-193, 1993.

[DeG90] Deriche, R. & Giraudon G., "Accurate Corner Detection: An Analytical Study", Proc. Int conf. On Computer Vision, pp. 66-70, 1990.

[DeM87] De Castro, & Morandi, C., "Registration of Translated and Rotated Images using Finite Fourier Transform", *IEEE Trans. Patt. Anal Machine Intell.* PAMI-9, No. 5, pp. 1186-1189, 1987.

[Di91] Diehl N., "Object Oriented Motion Estimation and Segmentation in Image Sequences", *Signal Processing:Image Communications,* 3, pp. 23-56,1991.

[ErG85] Eric W & Grimson L.," Computational Experiments with a Feature Based Stereo Algorithms", *IEEE Trans. PAMI* -7(1), pp. 17-34 , Jan., 1985.

[Ev88] Evans et al, "Anatomical-Functional Correlation Using an Adjustable MRI – Based Region of Interest Atlas with Positron Emission Tomography", *J. Cerebral Blood Flow Metabol.* 8, 513-530, 1988.

[FDS98] Fayolle J. , Ducottet C & Schon J. P., "Application of Multiscale Characterization of Edges to Motion Determination", *IEEE Trans. Image Proc.* Vol. 46(4), pp. 1174-1179 , April 1998.

[FDFS96]Fayolle J. , Ducottet C, Fournel T. & Schon J. P., "Motion Characterization of Unrigid Objects by Detecting And Tracking Feature Points", Proc. Int. Conf. Image Proc. Vol. 3,pp. 803-806,1996.

[Fl92] Flusser J., "An Adaptive Method for Image Registration", *Pattern Recognition*, Vol. 25, No. 1, 1992.

[Fr79] Franke E.,"A critical Comparison of Some Methods for Interpolation of Scattered Data", Tech. Rep. NPS-53-79-003, Naval Postgraduate School, 1979.

[Go85] Goshtasby A.,"Template Matching in Rotated Images", *IEEE Trans. Patt.*

Anal Machine Intell. PAMI-7, pp. 338-334, 1985.

[Go86] Goshtasby A.," Piecewise Linear Mapping Function for Image Registration", *Pattern Recognition*, Vol. 19, No. 6, 1986.

[Go88] Goshtasby A., "Image Registration by Local Approximation Methods", Image Vision Compt., Vol. 6, No. 4, pp. 255-261,1988.

[GoV80] G.H.Golub & C.F.Van – Loan, "An analysis of the Total Least Squares Problem", *Siam. J. Numer. Anal.* Vol. 17, pp. 883-893, No. 6, December 1980.

[Ha95] Hartley R. ,"In Defence of the 8-Point Algorithm", Proc. of the fifth Int. Conf. On Computer Vision, MA, pp. 1064-1070, 1995.

[HaK98] Han M. & Kanade T., "Homography-Based 3D Scene Analysis of Video Sequences", Proc. of the Image Understanding Workshop, 1998.

[Ho89] Hotter M. , "Differential Estimation of the Global Motion Parameters Zoom & Pan", *Signal Processing* 16, pp. 249-265,1989.

[HoS81] Horn, B. K. P. & Schunk, B.G., "Determining Optical Flow", *artificial intelligence,* Vol. 17, pp.185-203, 1981.

[HoT88] Hotter M. & Thoma R. , " Image Segmentation based on Object Oriented Mapping Parameters, *Signal Processing* 15, pp. 315-334,1988.

[Hs96] Hsieh J.W et al, "A Fast Algorithm for Image Registration without Predetermined Correspondences" , Proc. Int. Conf. Pattern Recognition pp. 765-769, 1996.

[Hs97] Hsieh J.W et al, "Image Registration Using a New-Edge Based Approach", *Comp. Vis. And Image Understanding,* Vol 67, No.2, August, pp. 112-130 ,1997.

[HuC98] Huang C.L. & Chang P.Y., "A multi Resolution Image Registration Method for Multimedia Application", Proceedings of the 1998 IEEE International Symposium on Circuits and Systems, pp. 13-16, New York, NY, USA,1998.

[HuU91] Hudaithi A.A. & Udpa S.S. "Recovering Structure and 3-D Motion Parameters of Rigid Planar Patch Using Moments", IEEE. Int. Symp. On Circuits & systems ,pp. 674-677,Vol. 1, 1991.

[IAH95] Irani M., Anandan P & Hsu S., "Mosaic Based Representation of Video Sequences and Their Applications",fifth Int. Conf. Computer Vision, pp. 605-611, 1995.

[IrA98] Irani M. & Anandan P, "A unified Approach to Moving Object Detection in 2D & 3D Scenes", *IEEE Trans.* PAMI, Vol. 20, No. 6, pp. 577-588, June 1998.

[IRP94] Irani M., Russo B. & Peleg S., "Recovery of Ego-Motion Using Image Stabilization", Proc. Conf. CVPR Seatle, pp. 454-460,June 1994.

[IRP97] Irani M., Russo B. & Peleg S., "Recovery of Ego-Motion Using Region Alignment", *IEEE Trans* PAMI , Vol. 19, No. 3, pp. 268-272, March 1997.

[Ja89] Jain, A. K., Fundamentals of Digital Image Processing, Prentice-Hall, 1989.

[KiK98] Kim E. T. & Kim H.M., "Efficient Linear Three dimensional Camera Motion Estimation Method with Application to Video Coding" , *optical Eng.* Vol. 37, pp. 1065-1077, March, 1998.

[KTB87] Kearney, J.K., Thompson, W.B. & Boley, D.L., "Optical Flow Estimation: an Error Analysis of Gradient-Based Methods with Local Optimization", *IEEE Trans. Patt. Anal Machine Intell.* PAMI-9, 5, pp. 229-244, 1987.

[KuH75] Kuglin, C. D.., & Hines, D.C. , *"The Phase Correlation Image Alignment Method" ,In Proceedings of the IEEE 1975 International Conference on Cybernetics and Society",* (Sept.) IEEE, New – York, pp.163-165, 1975.

[Le88] Lee C.H., " Structure and Motion from Two Perspective Views via Planar Patch", Second Int. Conf Comp. Vision, pp. 158-164, 1988.

[Le90] Lee C.H., "Motion Estimation of Rigid Planar Patch: A robust Computational technique", Proc. AI Systems in Government Conf., pp. 199-206, 1990.

[LuF93] Loung Q.T. & Faugeras O. D., "Determining the Fundamental Matrix from Planes: Unstability and New Algorithms", In Proc . Conference on CVPR, New-York, pp. 489-494, 1993.

[LuF96] Loung Q.T. & Faugeras O. D., "The Fundamental Matrix: Theory Algorithms and Stability Analysis", ", *Int. J. of Computer Vision,* 17, pp. 43-75, 1996.

[MaI99] Manor L.Z. & Irani M., "Multi-Frame Alignment of Planes", CVPR IEEE Conf. Los Alamitos, Ca.,USA, pp. 151-156, 1999.

[MaS88] Mao Z. & Strickland R.N., "Image Sequence Processing for Target Estimation in Forward Looking Infrared(FLIR) Imagery", Proceedings-of-the-SPIE, Vol. 901, pp. 44-53, 1988.

[MeN84] Medioni G. & Nevatia R., "Matching Images using Linear Features", *IEEE Trans. Patt. Anal Machine Intell.* PAMI-6, pp. 675-685, 1984.

[MoC96] Moigne J.L. & Cromp R.F., "The Use of Wavelets for Remote Sensing Image Registration and Fusion", *Proc. Of SPIE*, Vol. 2762, pp.535-544, 1996.

[MoK94] Moon J.H. & Kim J.K., "On the Accuracy and Convergence of 2-D Motion Models using Minimum MSE Motion Estimation", *Signal Processing:Image Communications,* 6, pp. 19-333,1994.

[MoY88] Morishita K. & Yokoyama T., "Image Registration Method Using Adaptive Nonlinear Filter", *Systems & Computers in Japan*, Vol. 19,No. 9, pp.41-50, 1988.

[MP1] Motion Pictures Experts Group, "MPEG video committee draft", MPEG 90/176.

[Na92] Nasrabadi N.M., "A Stereo Vision Technique Using Curve Segments and a Relaxation Matching", *IEEE Trans. PAMI* -14(5), pp. 566-575 , May 1992.

[Pa94] Park J. I. et al, "Estimation of Camera Parameters from Image Sequence for Model-Based video Coding", *IEEE Trans. Circuits and Sys. For Video Technologies,* , Vol. 4, No. 3, pp. 288-297, June 1994

[Pr86] Press W.H. et al, "Numerical Recipes in C" Cambridge University press, Cambridg, GB, 1986.

[Ro96] Rousso B., Avidan S., Shashua A. & Peleg S.," Robust Recovery of Camera Rotation from Three Frames", Fifth Int. Conf. Comp. Analysis of Images and Patterns, pp. 796-802,Berlin, Germany, 1996.

[ScA99] Schultz R. R. & Alford M. G., "Automated Image Registration using the Projective Transformation Model and Block Matching Feature Point Pair Selection", Proceedings-of-the-SPIE, Vol. 3719, pp.140-151, 1999.

[SzC94] Szeliski R. & Coughlan J. " Hierarchical Spline-Based Image Registration", Proc. *IEEE.* Conf. Computer Vision & Pattern Recognition", pp.194-201, Seattle, 1994.

[Sh94] Shashua A., "Projective Structure from Uncalibrated Images: Structure from Motion and Recognition", ", IEEE Trans. PAMI 16(8), pp. 778-790 , 1994.

[ShW95] Shashua A. & Werman M" , "Trilinearity of Three Perspective Views and its Associated Tensor", Proc. Fifth Int. Conf. Computer Vision,1995.

[ShW00] Shashua A. & Wolf L.,"Homography Tensors: On Algebric Entities That Represent Three Views of Static or Moving Planar Patches", Proc. of the European Conference on Computer Vision (ECCV), June 2000, Dublin, Ireland.

[ShT94] Shi J. & Tomasi C. "Good features to Track", *CVPR 94,* pp. 593-600, Seattle, June 1994.

[SoE98] Sutskover I. & Elad M., Personal Correspondance, 1998.

[SuA91] Sull S. & Ahuja N., "Estimation of Motion and Structure of Planar Surface from a Sequence of Monucular images", Proc. CVPR pp. 732-733, 1991.

[Te95] Tekalp A.M., "Digital Video Processing", Prentice Hall 1995.

[THZ82] Tsai R.Y. , Haung T.S. & Zhu W. "Estimating Three Dimensional Parameters of a Rigid Planar Patch, II: singular Value Decomposition", *IEEE Trans.* Acoustics, Speech and Signal Proc., Vol. ASSP-30, No. 4, August 1982.

[ToJ89] Ton, J. & Jain, A.K., "Registering LandSat Images by Point Matching", *IEEE Trans. Geosc. Remote Sensing,* Vol. 27, pp. 642-651, Sep . 1989.

[ToM97] Torr P.H.S & Murray D.W, "The Development and Comparison of Robust Methods for Estimating the Fundamental Matrix", *Int. J. of Computer Vision,* 20(3), pp. 271-300, 1997.

[TrV98] Trucco E. & Verri A., "Introductory techniques for 3-D Computer Vision", Prentice Hall, 1998.

[TsH81] Tsai R.Y. & Haung T.S. "Estimating Three – Dimensional Motion Parameteres of a Rigid Planar Patch", *IEEE Trans. Acoustics, Speech & Signal Processing*, Vol. ASSP-29, No' 8, Dec. 1981.

[TsH84a] Tsai R.Y. & Haung T.S "Estimating Three Dimensional Parameters of a Rigid Planar Patch, III Finite Point Correspondence and the Three-View Problem", *IEEE Trans*. Acoustics, Speech and Signal Proc., Vol. ASSP-32, No. 2, April 1984.

[TsH84b] Tsai R.Y. & Haung T.S "Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces", *IEEE Trans.* PAMI, Vol. 6, No. 1, pp. 13-, January 1984.

[VZR96] Vieville T, Zeller C. & Robert L., "Using Collineations to Compute Motion and Structure in an Uncalibrated Image Sequence", *Int. J. of Computer Vision,* 20(3), pp. 213-242, 1996.

[WHA93] Weng, J.,Haung, T.S. & Ahuja, N., "Motion and structure from Image Sequences", Springer-Verlag, 1993.

[WHA89] Weng J., Haung T. S. & Ahuja N. ,"Motion and Structure from Two Perspective Views: Algorithms, Error Analysis and Error Estimation", *IEEE Trans.* PAMI 11(5), pp. 451-476, 1989.

[WuK90] Wu S.F. & Kittler J., "A Differential Method for Simultaneous Estimation of Rotation, Change of Scale and Translation", Image Communication, Vol. 2, pp. 69-80,1990.

[WuK93] Wu S.F. & Kittler J., "A Gradient-Based Method for General Motion Estimation and Segmentation", *J. Visual Comm. & Image Proc.*, Vol. 4, No. 1, pp. 25-38,1993.

[Wi97] Winter A et al, E.,"Entropy & Multiscale Analysis: A New Feature Extraction Algorithm for Aerial Images", Int. Conf. Acoustics, Speech & Image Proc., ICASSP – 97,pp.2765-2768,1997.

[ZaL93] Zakor A. & Lari F., "3D Camera Motion Estimation with Applications to Video Compression & 3D Scene Reconstrction", Proc. Of the SPIE, Vol. 1903, 1993.

[Zh98] Zhang Z., "Determining the Epipolar Geometry and its Uncertainty: A Review", *Int. J. of Computer Vision,* 27(2), pp. 161-195, 1998.

[ZhC93] Zheng C. & Chellappa R., "A computational Vision Approach to Image Registration", *IEEE Trans. Image Processing*, Vol. 2 No. 3, July 1993.

[ZhC94] Zheng C. & Chellappa R., "Automatic Registration of oblique Ariel Images",Proc. ICIP, pp. 218-222, Los Alamitos Ca. 1994.

[ZHH86] Zhaung X., Haung T.S. & Haralick R.M., "Two-view Motion Anlysis: A unified Approach", J. Opt. Soc. American, Vol. 3. No. 9, pp.1492-1500, September 1986.

motion information for a group of frames. This work deals with estimating the global motion parameters under low computational load and robust estimation constraints.

When the scene plane is nearly parallel to the camera plane, the global motion can be described by an affine transformation having six parameters. If the scene is located on a single plane which isn't parallel to the image plane, the scene motion should be described by a planar transform with eight parameters. The non-linear planar transform estimation is not robust under noisy conditions. This work suggests a more robust method for planar transform estimation.

The camera motion parameters, such as rotation and translation, can be evaluated from the planar transform parameters. Thus, robust estimation of the planar transform parameters, as proposed in this work, enables a robust estimation of the camera motion parameters for planar scenes.

Efficient Coding of Image Sequences of Planar Scenes

by

Udi Shrot

000006459321

Efficient Coding of Image Sequences of Planar Scenes

FINAL PAPER

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE
IN ELECTRICAL ENGINEERING

Udi Shrot

SUBMITTED TO THE SENATE OF THE TECHNION –
ISRAEL INSTITUTE OF TECHNOLOGY
AV, 5760 HAIFA AUGUST 2000

This final paper was done under the supervision of Prof. David Malah and Dr. Meir Bar-Zohar in the faculty of Electrical Engineering.

I would like to thank Prof. David Malah and Dr. Meir Bar-Zohar for their devoted and close supervision.

I would also like to thank the staff of the Signal and Image Processing Lab, Ziva Avni, Nimrod Peleg and the rest of the lab's staff, for their help in creating an excellent working environment.

The support of Rafael and Samuel Neaman Institute is gratefully acknowledged.

Contents

Abstract 1

List of Symbols and Abbreviations 4

1 Introduction 5
- 1.1 Image-Sequence coding 5
- 1.2 Research goals 6
- 1.3 Thesis organization 7

2 Image Registration 9
- 2.1 Introduction 9
- 2.2 Image registration definition 10
- 2.3 Spatial transform estimation 10
  - 2.3.1 Using image model assumptions 10
  - 2.3.2 Estimation of optical flow 11
  - 2.3.3 Using correlation 12
  - 2.3.4 Using Fourier transform 12
  - 2.3.5 By matching distinct points 13
- 2.4 Local spatial transform estimation 16
- 2.5 Summary 17

3 Matching Points Extraction 18
- 3.1 Literature survey of matching points extraction methods 18
- 3.2 Algorithms for matching points extraction 20
  - 3.2.1 Algorithm I - "gradient method" 20
  - 3.2.2 Algorithm II - "eigenvalue method" 21
- 3.3 Simulation results 25

4 Coding Image Sequences Using Global Affine Transformation Estimation 28
- 4.1 Introduction 28
- 4.2 Camera motion 30
- 4.3 Estimation of global motion parameters using Least Squares 35
- 4.4 Simulations of global motion parameters estimation using LS 38
- 4.5 Coding image sequences using reference image and motion information 39

5 Improving Affine Transformation Estimation Using Linear Kalman Filter 44
5.1 Linear Kalman filter 44
5.2 Matching points smoothing algorithm 45
5.2.1 Simulation results 46
5.3 Affine transformation parameters smoothing using linear Kalman filter 48

6 Planar Transformation Estimation 50
6.1 The planar transformation 50
6.2 Literature survey of planar transformation estimation methods 52
6.3 Estimation of planar transformation using LS 60
6.4 Approximated estimation of the planar transformation 62
6.5 The proposed algorithm for the planar transformation estimation in an image sequence 66

7 Estimation of the Camera's Motion Parameters in Planar Scenes 74
7.1 Literature survey of camera's motion parameters estimation methods 76
7.1.1 Estimation methods based on optical flow estimation 76
7.1.2 Estimation methods based on matching points 77
7.1.3 Estimation methods for planar scenes 79
7.1.4 Estimation methods utilizing more than a single plane 80
7.2 Camera's motion parameters estimation using the planar transformation parameters 81

8 Summary and Conclusions 86

Appendix A: Stereo Vision 88

Appendix B: Second Order Motion Model 92

References 95

English Abstract I

Figure List

Figure	page
3.1: Algorithm I ("Gradient Algorithm") for finding matching points	20
3.2: The output of Algorithm I ("Gradient Algorithm")	21
3.3: The "Salesman" image rotating with angular velocity of 2.5° per frame	26
3.4: Choosing matching points	26
3.5: Energy of difference image – motion estimation based upon matching points	27
4.1: Pin-hole camera model	28
4.2: Orientation between camera's coordinate system and world reference coordinate system	30
4.3: Least squares solution geometry	36
4.4: Comparison between different motion estimation methods	39
4.5: Incremental motion estimation method	40
4.6: Prediction based motion estimation method	41
4.7: Quantitative comparison between motion estimation methods in image sequences	42
4.8: Subjective comparison between motion estimation methods in image sequences	43
5.1: The linear Kalman filter equations	45
5.2: Matching points smoothing using a linear Kalman filter	47
5.3: Matching points smoothing using a linear Kalman filter - two iterations	47
5.4: Affine transformation parameters smoothing using a linear Kalman filter	49
6.1: Planar scene	50
6.2: Affine model results	64
6.3: Bilinear model results	65
6.4: Second – order model results	65
6.5: Planar transformation estimation algorithm	67
6.6: Results of using second order model to improve planar transformation estimation	68
6.7: Image Sequence created by a planar transformation	70
6.8: Performance comparison between planar transformation estimation algorithms (synthetic sequence)	70
6.9: Averaging motion compensated sequence images (20 images) – Algorithm I (synthetic sequence)	71
6.10: Averaging motion compensated sequence images (20 images) – Algorithm II (synthetic sequence)	71
6.11: Images from an Image sequence taken by a moving camera	72
6.12: Performance comparison between planar transformation estimation algorithms (real sequence)	72

6.13: Averaging motion compensated sequence images (30 images) – Algorithm I (real sequence) 73
6.14: Averaging motion compensated sequence images (30 images) – Algorithm II (real sequence) 73
7.2: Performance comparison between camera motion estimation algorithms in planar scenes 85
A.1: Stereo cameras configuration 89

Abstract

Recent advances of digital technology and communications present new challenges and requirements for high-quality visual communications, on one hand, while maintaining low transmission bit-rate and storage requirements, on the other. These are conflicting demands requiring therefore a suitable trade-off. The coding methods have to keep low transmission rate and memory demands but still make sure that the pictures' quality will be acceptable so that the viewer will not be disturbed by the quality loss.

Common image sequence coding methods exploit both spatial and temporal redundancy existing in the sequence. The temporal redundancy is due to the fact that consecutive frames in the sequence are very similar. It can be exploited by coding the difference between consecutive frames instead of the frame itself. Since the energy of the difference frame is smaller than that of the original frame signal, less bits have to be transmitted to the decoder. The spatial redundancy is due to the strong correlation between close pixels in the frame and is being exploited by transform coding. A transform is applied to each block in the frame in order to get a compact representation of the block information with a small number of coefficients. The transform coefficients are being quantized in the frequency domain in a manner which is consistent with the human visual system characteristics. The quantization sets the rate and the quality of the coded frame. By using an appropriate quantization scheme, a gain in the bit-rate can be achieved without noticeable quality loss to the human observer. Further rate reduction is achieved by losslessly coding the quantized transform coefficients and side information in an efficient manner using variable length coding (VLC).

In order to improve the prediction of the frame to be coded, motion compensation is used. Many of the changes between consecutive frames are due to motion of objects in the scene. Therefore, a gain in performance can be obtained by coding the difference between the coded frame and its prediction, which is obtained from the previous frame using motion compensation. In that case, the motion information also has to be coded but the prediction improvement makes it worth while. The motion compensation, which is common in standard coders of image-sequences, is of local motion only and assumes that all the pixels in a block move with the same translatory motion. This assumption is done in order to enable motion description which requires a small amount of side information and in order to reduce the computational load to enable simple hardware implementation. In image-sequences of 3-D scenes this assumption is inaccurate, even when the camera is just panned, because the magnitude of the motion of pixels in the image is changing gradually in accordance with their distance from the camera. When zoom is present or when the camera is being rotated this assumption is not valid even for 2-D scenes. In those cases it is better to apply techniques for global motion estimation and compensation which rely on a more appropriate motion model, providing an improved motion description. This way the prediction of the frame to be coded is improved, and the bit-rate is lowered or the quality of the frame improves, at the same bit-rate.

Global motion in certain scenes, such as scenes far from the camera or scenes located on a single plane (planar scenes), can be described by a single parameter set which describes the motion on the image plane. In addition to the reduced side information needed, this motion description is usually more accurate in the mentioned scene. In this work an approach for efficient coding of image-sequences of planar scenes is proposed. The coding is performed by transmission of the first image and the global

motion information for a group of frames. This work deals with estimating the global motion parameters under low computational load and robust estimation constraints.

When the scene plane is nearly parallel to the camera plane, the global motion can be described by an affine transformation having six parameters. If the scene is located on a single plane which isn't parallel to the image plane, the scene motion should be described by a planar transform with eight parameters. The non-linear planar transform estimation is not robust under noisy conditions. This work suggests a more robust method for planar transform estimation.

The camera motion parameters, such as rotation and translation, can be evaluated from the planar transform parameters. Thus, robust estimation of the planar transform parameters, as proposed in this work, enables a robust estimation of the camera motion parameters for planar scenes.

[ZhC94] Zheng C. & Chellappa R., "Automatic Registration of oblique Ariel Images",Proc. ICIP, pp. 218-222, Los Alamitos Ca. 1994.

[ZHH86] Zhaung X., Haung T.S. & Haralick R.M., "Two-view Motion Anlysis: A unified Approach", J. Opt. Soc. American, Vol. 3. No. 9, pp.1492-1500, September 1986.