DEBRIS TRACKING IN A SEMISTABLE BACKGROUND

by

KARTHIKKALATHI VANUMAMALAI
B.E. University of Madras, 2003

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science
in the Department of Electrical and Computer Engineering
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Fall Term
2005

ABSTRACT

Object Tracking plays a very pivotal role in many computer vision applications such as video surveillance, human gesture recognition and object based video compressions such as MPEG-4. Automatic detection of any moving object and tracking its motion is always an important topic of computer vision and robotic fields.

This thesis deals with the problem of detecting the presence of debris or any other unexpected objects in footage obtained during spacecraft launches, and this poses a challenge because of the non-stationary background. When the background is stationary, moving objects can be detected by frame differencing. Therefore there is a need for background stabilization before tracking any moving object in the scene.

Here two problems are considered and in both footage from Space shuttle launch is considered with the objective to track any debris falling from the Shuttle. The proposed method registers two consecutive frames using FFT based image registration where the amount of transformation parameters (translation, rotation) is calculated automatically. This information is the next passed to a Kalman filtering stage which produces a mask image that is used to find high intensity areas which are of potential interest.

Dedicated to my Family

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Takis Kasparis who has given me immense support and guidance throughout the course of this work. I would also like to thank the members of my thesis committee Dr.Michael Georgiopoulos and Dr. Michael Haralambous who have accepted to be in my committee with in a short period of notice. I would also like to thank my colleagues, friends and well wishers who have supported throughout my life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ACRONYMS

Fps……………………Frames per second

Corr……………………Correlation

FFT……………………Fast Fourier Transform

2D………………………Two Dimensional

KLT……………………Kanade-Lucas-Tomasi

MR……………………..Magnetic Resonance

PET……………………Positron Emission Tomography

SMC……………………Sequential Monte Carlo

Pdf………………………Probability distribution function

LeRP……………………Length- R Paths

CT………………………Computed tomography

PDE……………………..Partial Differential Equation

MAP……………………..Maximum Aposteriori estimate

CHAPTER ONE: INTRODUCTION

There are many methods and techniques available in the literature for Image Registration such as Fourier Methods, Point Mapping, Correlation based and Elastic Model Based Matching and for Visual Tracking such as Optical Flow, Feature based and Contour based. While all these methods work well for a variety of tracking applications they also have limitations.

The general overview of different techniques reported in the literature for Image Registration has been presented in the Chapter 2.It mainly involves description of the most commonly used methods, namely, Point Mapping, Fourier Methods and Correlation Methods.

Chapter 3 deals with various visual tracking methods, including feature based methods using KLT Tracker and IPAN Tracker, finding optical flow using Horn and Schunk and prediction based methods using a Kalman filter. Chapter 4 describes the approach used in this work and Chapter 5 provides conclusions inferred from the experiments. Finally, chapter 6 proposes future work.

Motivation for this Work

This main objective of this work is to device an algorithm to detect any falling debris in footage obtained during lift-off that may pose a threat to Space Shuttle. The driving force behind this work is the increased safety awareness policy adopted by NASA after the Columbia Space Shuttle tragedy.

The tragic disintegration of the Columbia occurred on its re-entry, only sixteen minutes before the scheduled landing. The videos and images obtained during lift-off showed a piece of foam falling from the external tank and hitting the left wing, and is considered the primary cause for the shuttle disintegration. Frames from this video are presented in this work. The impact velocity of the debris was estimated to be between 500 frames per second and 800 frames per second relative to the Space Shuttle. A lot of image enhancement techniques such as deconvolution filters, stabilization, contrast stretching, frame averaging and band ratioing have been tried out by the NASA image analysis team to determine the debris size and color, reduce motion blur and to sharpen the details. Change in the color of the debris was attributed to a rotational movement.

Problem Statement

The two videos which have been used in this work has a semi stable background and so it is required to register the frames to make the background stable before further processing is done. Differencing of the consecutive frames fails when the background also varies since there will be residual noise associated with the motion. The objective is to reduce this noise and this can be done by registering consecutive frames of the image sequence. A Kalman filter is used because it is best suited when illumination changes occur and it does not consider intensity changes as a foreground, but rather it can be used for background estimation. It also takes into account occlusions a little bit .Since there is a clear distinction of the foreground and the background the occluded object reconstruction is detected from its boundary. The whole process is divided into two major steps:

1) Image Registration : Align the Images to stabilize the background and this is done by using the Phase Correlation method.
2) Kalman Filter : The filter is used to detect the foreground and also to estimate the background. Based on the information from the Kalman filter, all the moving objects which appear as a foreground are tracked.

CHAPTER TWO: IMAGE REGISTRATION

Image registration is the method of matching a target image with respect to the reference image so that corresponding pixels in the two images represent the same region of interest. The need for registration arises due to many cases such as images taken from different viewpoints or different sensors or different times or from moving camera. The application of image registration is widely used in remote sensing as well as medical and computer vision applications. Registration methods can be broadly classified into two types:

1) Manual registration methods

2) Automatic registration methods

Manual registration methods involve mapping ground control points in both images and this method requires a large number of control points for accurate matching of images. Since this method involves human interaction this is tedious and labor intensive.

Automatic registration methods can again be classified into two types; area based and feature based. In the area based methods, blocks or windows of smaller size are compared between the two images and a usual similarity metric is given by the normalized cross correlation which can be implemented using Fourier transform. When the amount of translation or rotation is large, this method becomes unreliable. On the other hand, feature based methods are more robust to large displacements. These methods extract features from each image and then try to match them. Typical feature extraction methods include the Canny operator, LoG operator and region growing. Depending on the method by which the image has been acquired feature based methods can be divided into the following types:

- Multi view analysis
- Multi temporal analysis
- Multi modal analysis

Multi view analysis refers to images of the same scene obtained from different view points. Examples of applications are shape recovery in computer vision and mosaicing of the images of the surveyed area.

In multi temporal Analysis image of the same scene are obtained at different times and also at different conditions so that changes can be detected. This is mostly used in medical imaging where constant monitoring of tumor or any healing therapy is required. It can also be applied to computer vision application such as motion tracking and video surveillance.

Multi modal analysis refers to images of the same scene obtained by different sensors and this is mainly done to get a detailed understanding of the scene under observation. Typical applications include MRI, PET, SPECT or MRS where the metabolic activity of the body is recorded by different sensors.

Definition of registration

Given two images $I_1(x, y)$ and $I_2(x, y)$ that share some common mutual information, then the mapping between the two can be represented as follows:

$$I_2(x, y) = t(I_1(s(x, y))) \tag{1}$$

where, *s* is a 2-D Spatial coordinate transformation and *t* is a 1-D intensity transformation.

The most common general transformations are rigid, affine, projective, perspective and polynomial. A transformation T is linear if

$$T(ax_1+bx_2) = aT(x_1) + bT(x_2) \tag{2}$$

The transformation is said to be affine if T(x) – T (0) is linear. Affine transformation is composed of Cartesian operations of translation, rotation and scaling. Affine transformation normally has four parameters namely translation parameters along the x and y direction, a rotation parameter and a scaling parameter which map a point (x_1, y_1) in the reference with the point (x_2, y_2) in the target image as follows:

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} t_x \\ t_y \end{pmatrix} + s * \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \tag{3}$$

The 2-D affine transformation is given by

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} a_{1,3} \\ a_{2,3} \end{pmatrix} + \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \tag{4}$$

which takes care of skew and aspect ratio in addition to the aforementioned geometric transformations.

Perspective transformation is a mapping from 3D to 2D. If the coordinates of an object in 3D is x_0, y_0, z_0 then its corresponding point in the 2D image is given by

$$x_i = \frac{-f x_0}{z_0 - f} \tag{5}$$

$$y_i = \frac{-f y_0}{z_0 - f} \tag{6}$$

where, f is the focal length of the lens.

If the scene is composed of a flat plane tilted with respect to the image plane then there is a need for the projective transformation which maps a coordinate on the plane (x_p, y_p) on to the coordinates in the image (x_i, y_i) given by

$$x_i = \frac{a_{11} x_p + a_{12} y_p + a_{13}}{a_{31} x_p + a_{32} y_p + a_{33}} \tag{7}$$

$$y_i = \frac{a_{21} x_p + a_{22} y_p + a_{23}}{a_{31} x_p + a_{32} y_p + a_{33}} \tag{8}$$

Global alignment is needed in cases where the distortions in the scene are not taken into account or there is not sufficient information about the camera geometry. So this can be determined using polynomial transformation. In case of non linear distortion due to the sensors or object deformations local transformation is sufficient which is taken care by piecewise interpolation like splines where the features to be matched are known or model based methods like elastic warping, object/motion models.

Image Distortion can be classified as static, internal/external, geometric/photometric. Static distortions do not change with the image; hence it can be corrected by calibration techniques. Internal distortions are due to the sensor caused by camera shading effects, lens distortions, sensor imperfections and sensor induced filtering. External errors mainly occur due to changing sensor operations or individual scene characteristics.

Image Registration Methods

Correlation and Sequential Methods

Cross correlation is often used for template matching and it is a measure of the correspondence between the template and the reference image. The template image is always small when compared to the reference image. If the template image is represented by T and the image by I we represent the 2-D normalized cross correlation as

$$Corr(s,t) = \sum_x \sum_y \frac{T(x,y)I(x-s,y-t)}{\left[\sum_x \sum_y I^2(x-s,y-t)\right]^{1/2}} \tag{9}$$

Cross correlation is related to the Euclidean distance which calculates the sum of the squared differences between the template and the image given by

$$ED(s,t) = \sum_x \sum_y (T(x,y)I(x-s,y-t))^2 \tag{10}$$

This measure is going to give the smallest value when the difference between the template and the reference image is little. When this expression is expanded there are three terms: template energy term, product of the template and the reference image and the reference image energy term. The normalized version of the product of the template and the energy is the one which gives the degree of correlation.

An important property of Correlation is the Fourier transform of the correlation of the two images is the product of the Fourier transform of one image with the complex conjugate of the Fourier Transform of the other image. Correlation is normally applied to the problems which involves small translation, rotation and scaling because when the transformation becomes large it becomes computationally expensive. Correlation computed using Fourier methods is described in detail in Chapter 4. Fourier based correlation is robust to images having low frequency noise mainly caused by atmospheric or lighting variations; it tends to fail for high frequency noise like white noise.

A relatively better method which is also computationally simpler than correlation is called Sequential Similarity detection algorithm calculates the similarity metric based on L_1 norms given by:

$$E(u,v) = \sum_x \sum_y |T(x,y) - I(x-u, y-v)| \tag{11}$$

The normalized measure is given by:

$$E(u,v) = \sum_x \sum_y |T(x,y) - \hat{T} - I(x-u, y-v) + \hat{I}(u,v)| \tag{12}$$

Where $\hat{T}$ and $\hat{I}$ are the means of the template and the local image respectively.

Even though sequential methods have the advantage of being computationally efficient with minimal degradation in performance, suffer from the problem of becoming complex as the search space increases in size.

Sequential methods are found to be far superior compared to correlation in terms of matching similar images, but both of these methods fail when the images are dissimilar since both depend on correlation coefficients and SSD(Sum of squared differences) respectively which is a property of identical images.

Point Mapping

When the matching process requires robustness for example when there are occlusions in the images or any other distortions, this method seems to produce good results. Point mapping generally works in 3 stages:

- Feature extraction from the image to be registered
- Map the features or the control points of the reference image with the test image
- Determine a 2 D polynomial function of a specified order from the registered image that calculates the transformation parameters.

Selecting control points is the most important step in point mapping since it determines the accuracy of the method. Therefore extra care needs to be taken in the first step. There are generally two types of control points: Intrinsic and extrinsic. Intrinsic points refer to those which are not related to the scene. They are generally used to aid the registration process. As an example, chemical markers are used in MRI systems. Extrinsic control points are those present in a scene and selection requires a well understanding of the scene. Features which are usually used are corners, contours, line intersections etc. Selection of optimal number of features is required for the success of this method since fewer features can result in improper alignment and too many features can make the registration technique complicated.

Point mapping can also be done using feedback between the selections of control points and finding the optimal transformation, and this includes methods like relaxation, clustering, graph matching etc. In relaxation the displacement is the match between each feature points and it is given a rating. The procedure is iterated till the highest rating is achieved which represents the optimal transformation. It has an advantage of sustaining local and global variations but suffers from high computational complexity. For the purpose of clustering, the parameters of the feature point are represented as a point in the feature space and the best point is chosen by statistical methods. Point mapping can also be done without feedback and it can be global or local polynomial methods. Global methods can be done by approximation or interpolation given an adequate number of points. While the approximation can be done by statistical methods such as least squares regression or clustering, interpolation is done by deriving a system of equations using the coefficients of the polynomials.

Figure 2.1 is a typical example of an image registration being done using the point mapping method. The top row shows control points that are established or identified in both the images. The second row shows the second stage of the point mapping establishing the relationship between the control points of both the images. The left figure of the third row shows that the estimation of the transform parameters is done which is used for mapping of images which is done in the right half of the third row.

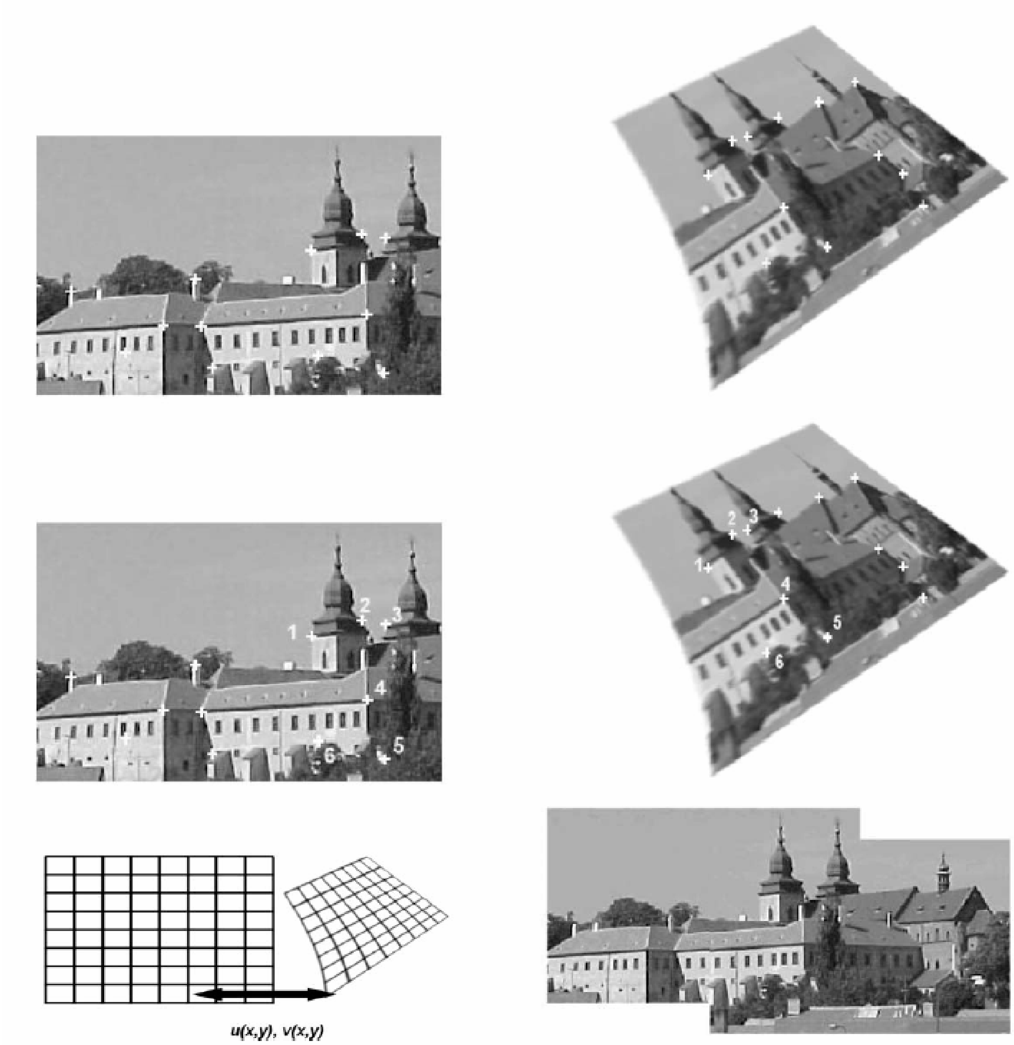


Figure 2.1 Image Registration Example

Parametric and Non Parametric Methods

Another form of classifying image registration methods is parametric and non parametric, and these methods are predominantly used in medical imaging.

Parametric registration methods are of the following types:

1) Landmark Based Registration

2) Principal axes Based Registration

3) Optimal Linear Registration

Non Parametric registration methods can be classified as follows:

1) Elastic Registration

2) Fluid Registration

3) Diffusion Registration

4) Curvature Registration

Parametric Methods

Landmark Based Registration

Landmarks are easily located easily by the user based on the anatomical properties of the image or based on geometrical properties such as corners, curvature maxima by the commonly used image segmentation methods. The set of points used for comparison in this method is less thereby reducing the computation needed. The optimization of this algorithm is done by measuring the norm distance between each landmark and its neighbor or by iterating the minimized landmark distance. For this Iterative Closest point algorithm [3] is used. Landmark based methods are normally used for rigid and affine transformations. [25] deals with the landmark registration achieved with a deterministic procedure called Length- r paths (LeRP) Algorithm for finding the corresponding features. In a nutshell the method proposed by the author for registration involves in finding the feature points, refining it, forming a graph using those feature locations, and matching these graphs using LeRP algorithm to find corresponding features. Then the residual error and the absolute rotation are found out. [26] deals with landmark registration of the curves using continuous wavelet transform. It is based on aligning the main features which occur along with the noise (structural intensity) of the zero crossings of the wavelet transform.

Voxel property based Registration

This method plays an important role in the autonomous registration of the MR-PET Images. This type of method falls under Reductive registration methods where in the image center of gravity and the principal axes or orientation is calculated from the zeroth and the first order moments. After finding out the moments center of gravity and the principal axes are aligned to register the images. Sometimes higher order moments are found out for calculating the center of gravity and principal axes. Although this method does not give you highly accurate results it has an advantage of registering images automatically. This method is normally used as a pre processing step before any better registration method is used and in some cases for re aligning the images related to cardiac studies. They use almost all the information available from the images unlike other methods which uses less information for calculating transformation. This property limits this method for use in 3-D images because of the computational cost.

Optimal Linear Registration

Optimal Linear Registration can be Intensity based registration, Correlation based registration or Mutual Information based registration. Correlation based registration is explained in the previous section so let us concentrate on the other two. There are various methods proposed in the literature for registration of images using mutual information. The method proposed by Woods [5] matches the corresponding regions in both the images and tries to minimize the average variance of the grayscale value to achieve registration. An extension of the Woods method is by Hill [6] where in instead of regions they represent it as a feature space which gives the 2-D plot or the histogram.

Non Parametric Methods

Elastic Registration

This method models the distortion or the misalignment in the image as a deformation of the elastic material with the properties of bending and stretching. The energy in the elastic material determines the amount of bending and stretching. This method uses a vector spline regularization to solve the two problems of elasticity namely divergence and curl. There is also this method of using statistical features, which is, combining affine and elastic transformation mentioned in [10]. The affine transformation is dealt by affine Principal Component Analysis which is like a preprocessing step to obtain the maximum spatial correlation. The elastic transformation is calculated by getting the accurate feature boundaries by segmentation using Monte Carlo methods. In this type of registration methods the transformation is due to bending and stretching of an elastic material and the energy associated with the material is related to the bending. The main objective of this method is to find the minimum energy state which defines the transformation. Automatic elastic registration was developed by Burr which iteratively finds the distance between a point in the first image with the corresponding point in the nearest neighborhood in the second image. After each iteration, the neighborhood closes thereby resulting in a perfect match. The method adopted by Burr includes three approaches: Iteration, Hierarchical structure and cooperation. Iteration changes the effective neighborhood used for matching every step, hierarchical nature considers the global variations first and later the local variations resulting in coarse to fine estimation and finally cooperation takes into account how well a feature and its neighbor matches. Misregistration occurring due to local and global distortion is avoided by the coarse to fine nature of the hierarchical approach.

Fluid Registration

The Fluid Registration algorithm is based on the linear deformation of the elastic material and this principle is used in getting a convolution filter as mentioned in [11]. The algorithm uses an Eulerian Reference frame where in the particles are tracked from their current position rather than from the previous position as done in Lagrangian frames. Here the template itself is modeled as undergoing viscous fluid deformation. The Eulerian velocity as mentioned in [11] is given by

$$v(x,t) = \delta u(x,t)/\delta t + \nabla u(x,t)v(x,t) \tag{13}$$

where ∇ represents the gradient operator u(x,t) represent the displacement along x

and the product $\nabla u(x,t)v(x,t)$ represent the kinematic non linearities of the particles.

The Viscous Fluid deformation model can be written in PDE as follows:

$$\mu\Delta v(x)+(\mu+\lambda)\nabla(\nabla.v(x)) = f(x,u(x)) \tag{14}$$

where Δ is the laplacian operator

$\nabla(\nabla.v)$ is the divergence

Δv is the viscous term

Euler integration is used to solve the PDE which is the solution of the viscous fluid registration.

Diffusion Registration

This registration can be related to the most commonly used registration in Medical Imaging called as Thirion's Demons Algorithm. It is based on the fact that if we consider the boundaries of an image as a semi permeable membrane then the other image which is used for matching is represented as a deformable grid model which is made to diffuse through the membrane. It is based on Maxwell's concept in the field of thermodynamics. This algorithm applies an extension of the existing demons algorithm, which is using the gradient information for registration. If the gradient is too low the general algorithm fails so this method calculates the optical flow using the Thirion's formula:

$$\hat{u} = ((m-s)\nabla s)/|\nabla s|^2+(m-s)^2 \tag{15}$$

where $\hat{u} = (u_x, u_y, u_z)$ represents the displacement m is the intensity of the moving image, s is the intensity of the static image and ∇s is the gradient of the static image and it represents the internal force emanating from the static image.

The displacement is calculated by carrying out a number of iterations. Each iteration involves the calculation of optical flow, followed by the application of Gaussian filter to regularize it. In the proposed approach registration is made efficient by assuming diffusion taking place in both directions; thereby static image diffuses into the moving image in addition to the moving image diffusing into the static image. When both the images are treated in the same manner there is a considerable improvement in the accuracy of the registration. There are two forces which contribute to the gradient information of the image; active force contributes to the moving image and passive force to the static image. Coarse to fine multiresolution estimation is also done to minimize the large deformation which results in large registration errors.

Curvature Registration

The main idea of this approach which is described in [23] is to find the transformation of the template T with respect to the reference R. So there is a need to determine the displacement field u which is nothing but mapping between two images and it is done by minimizing the cost function given by:

$$C(u) = 1/2\int_{\Omega} (T(x - u(x)) - R(x))^2 \; dx \qquad (16)$$

There is a joint constraint which should be minimized by the mapping u given by:

$$J(u) = \alpha S(u) + C(u) \tag{17}$$

There is a smoothing function introduced in this approach to eliminate the solutions below optimal and it is given by:

$$S^{\text{curv}}[u] = \sum_{l=1}^{d} \int_{\Omega} (\Delta u_t)^2 \, dx \tag{18}$$

This represents an approximation to the curvature of the *l*-th component of the displacement field. The mapping u should satisfy Euler Lagrange criterion for S = S^{curv}

$$f(x, u(x)) + \alpha \Delta^2 u(x) = 0 \quad \mathit{for} \quad x \varepsilon \Omega \tag{19}$$

subject to boundary conditions.

CHAPTER THREE: VISUAL TRACKING

As already mentioned visual tracking is an active research in computer vision which comes with innumerable applications. Broadly it can be divided into three types:

- Feature based methods which involves a prior knowledge of the features in the image. This method has a limitation of being sensitive to intensity changes.
- Optical flow methods involve in calculating the motion vectors which ultimately helps to calculate the position and the velocity of the object.
- Contour based methods models the outline or the borders of the object as curves and tracking is achieved by minimizing the energy function calculated in the vicinity of contour. This method has the advantage of handling occlusions which is missing in the aforementioned two methods.

Feature Based Methods

This approach extracts features from images such as corners, edges and tries to track from frame to frame. It is done in two steps namely Feature extraction and tracking. If the extraction of the features is done well the problem of excessive processing of the scene is eliminated ultimately leading to better matching thereby making tracking easier. There are many algorithms reported in the literature for feature selection but let us consider KLT tracker and IPAN tracker which have been used often. KLT tracker finds features in the first image and tries to match those in the subsequent images whereas IPAN tracker finds features in each static image and tries to track it in the sequence.

KLT Tracker

Kanade-Lucas-Tomasi Tracker involves in detecting a feature such as a corner which has high intensity variation in the x and y directions. If the intensity function is denoted by I(x, y) then the local intensity variation matrix is given by

$$D = \begin{bmatrix} I_x I_x & I_x I_y \\ I_x I_y & I_y I_y \end{bmatrix} \tag{20}$$

A window of variable size such as 5*5, 10*10, 25*25 can be used and the best candidate is chosen when the eigenvalues $-_1$ and $-_2$ of D in the center of the window exceeds a threshold -, given by min($-_1$, $-_2$) > -. The number of features to be used for tracking can be given by the user. KLT corrects the change of appearance between the current and the previous frame by measuring the dissimilarity occurring between them thereby allowing for affine changes. Although KLT does not provide the uniform velocity flow the spatial distribution of the flow tends to be regular.

IPAN Tracker

The IPAN Tracker is procedure that compares between three consecutive images and maximizes the smoothness of the trajectory. Here the path is extended in each and every frame from the already existing partial trajectories. Here two constraints are taken into consideration namely intensity maximum and mean intensity maximum.

A pixel is said to have a local maximum intensity if its brightness level exceeds that of any of its 8 neighbors in a 3*3 neighborhood given by:

$$I(x, y) > I(i, j) \text{ for all } (i, j) \neq (x, y): |i-x| \leq 1, |j-y| \leq 1 \quad (21)$$

A pixel is said to have a mean intensity maximum if the average gray value of that exceeds any of its 8 neighbors given by:

$$\hat{I}(x, y) > \hat{I}(i, j) \text{ for all } (i, j) \neq (x, y): |i-x| \leq 1, |j-y| \leq 1 \quad (22)$$

where I^(x, y) is the mean intensity maximum given by:

$$\hat{I}(x, y) = 1/9 \sum_{p=x-1}^{x+1} \sum_{q=y-1}^{y+1} I(p,q) \quad (23)$$

Each feature point is assigned a feature value and it's been normalized to represent the strength of the feature. The IPAN tracker considers 3 frames simultaneously and tries to minimize the error function by applying changes in the feature value, direction and the magnitude of the velocity vector. The post processing of the velocity vector field is done by coherence filtering. Consider a feature point f_c with the velocity v_c and the features lying with in a distance D from f_c be denoted by f_i, i = 1, 2…n and their velocities are denoted by v_i. Since these vectors form a cluster mean cumulative difference is taken between the velocity vectors given by:

$$\Delta_i = \Sigma_{j \neq i} v_i - v_j / (n - 1) \quad (24)$$

The median velocity vector is given by:

$$\Delta_{med} = \text{arg.min } \Delta_i \quad (25)$$

The difference between the median velocity vector and the outlier vector should be significant and should satisfy:

$$\| v_c - v_{med} \| > \Delta_{med} \quad (26)$$

Optical Flow Methods

Some of the methods used to determine optical flow are Horn and Schunck, Barnard and Thompson method. Optical flow represents the motion as vectors originating or terminating at any pixel in a sequence of images.

Horn and Schunck Method

This method considers an image sequence i(x, y, t) where x, y represent the pixel positions and t represents the time. Now assume that the object moves to a new position x + dx, y+dy at time t+dt given by:

$$i(x, y, t) = i(x+dx\ , y+dy, t+dt) \tag{27}$$

By applying Taylor's series and simplifying the equation we end up with

$$i_x\, dx + i_y\, dy + i_t\, dt = 0 \tag{28}$$

where $i_x = \delta f/\delta x$, $i_y = \delta f/\delta y$ and $i_t = \delta f/\delta t$

Now divide the equation (28) by dt

$$i_x\, u + i_y\, v + i_t = 0 \tag{29}$$

where u = dx/dt, v = dy/dt represents the optical flow.

The authors estimated the optical flow by minimizing the error function given by:

$$e(x, y) = (\, i_x\, u + i_y\, v + i_t\,)^2\ + \varepsilon(u_x^{\,2} + u_y^{\,2} + v_x^{\,2} + v_y^{\,2}) \tag{30}$$

Differentiating e with respect to zero and representing the sum of u along x and y direction by a term $\Delta^2 u$ and sum of v along x and y as $\Delta^2 v$ we get

$$(i_x u + i_y v + i_t)\, i_x\ + \varepsilon(\Delta^2 u) = 0$$

$$(i_x u + i_y v + i_t)\, i_y\ + \varepsilon(\Delta^2 v) = 0$$

If we represent the average value of u and v as u_{av} and v_{av} and $\Delta^2 u = u - u_{av}$,

$\Delta^2 v = v - v_{av}$ Solve for u and v we get:

$$u = u_{av} - i_x(i_x u_{av} + i_y v_{av} + i_t)/(\varepsilon + i_x^2 + i_y^2) \tag{31}$$

$$v = v_{av} - i_y(i_x u_{av} + i_y v_{av} + i_t)/(\varepsilon + i_x^2 + i_y^2) \tag{32}$$

Token Based Method

In stereo motion the problem of correspondence is simplified by the use of tokens since the possible matches can occur only along the epipolar line. If we have N points to be matched in 2 frames the number of possible mappings is N! and this is kind of complex when the number of points increases. So there are constraints such as maximum velocity, smoothness of motion, rigidity etc. introduced to reduce the complexity. If the velocity bound is known beforehand then the velocity constraint limits the search of the match in the next frame in a small neighborhood. There is a basic assumption in this approach that the direction and speed does not change to a larger extent resulting in a smoother motion. The objects are said to be rigid if the Euclidean distance between the two consecutive frames does not change. Smoothness and Rigidity together tries to give solutions close to the optimal solution.

Prediction based methods

Kalman filters comes under this category since it estimates or predicts the state of the system, compares it with the actual value and gives a new estimate by minimizing the error function, that is, the difference between the actual value and the predicted value. An alternative of the extension of Kalman filter, the Particle Filter is also widely used for tracking applications.

Kalman Filters

The Kalman Filter is a set of mathematical equations that provides an efficient means to estimate the state of a process that minimizes the mean squared error. According to [17] Kalman filter estimates a state at time t_i given by

$$\hat{s}(t_i) = \tilde{s}(t_i) + G(t_i) .[y(t_i) - H(t_i) . \tilde{s}(t_i)] \tag{33}$$

and the prediction term is given by

$$\tilde{s}(t_i) = B(t_i). \hat{s}(t_{i-1})$$

where $B(t_i)$ – System matrix

$H(t_i)$ – Measurement matrix

$y(t_i)$ – System Input

Matrix $G(t_i)$, the Kalman gain matrix is obtained from the error covariance matrix. The Kalman filter is a recursive predictor, that is, it includes all the past values without storing the measured values. The steps followed by Kalman Filter in tracking any object at a time instant t are:

Initialization : This is at time instant t = 0. The whole image is looked for the object since there is no prior information about the position of the object.

Prediction : This takes place at time t > 0 . The relative position of the object is found out as s^(t_{i-1}) and it forms the search center.

Correction : Here the object is located which is supposed to lie in the neighborhood of s^(t_{i-1}) and the real position of the object is used to correct the state.

Particle Filter

It is also known as Sequential Monte Carlo methods (SMC). They are used to estimate the Bayesian models based on the observation data. Particle filters have an edge over Kalman Filter in that it can be used to represent even multimodal distribution while Kalman is limited to Gaussian probability distribution. Reference [24] deals with tracking multiple targets using a Particle Filter wherein Gibbs Sampler is used to estimate the stochastic vector of assignation and [27] explains about tracking and recognition of human faces using Particle filters. It involves propagating the point estimate using MAP estimation and its velocity of motion is calculated using the first order approximation and including an additive noise.

CHAPTER FOUR: METHOD OF APPROACH

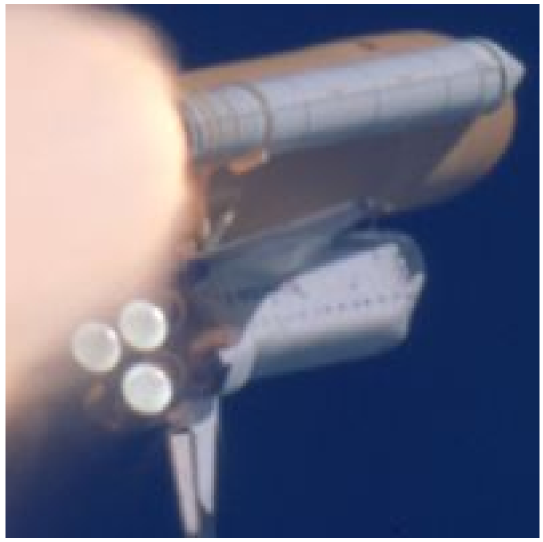
Why Image Registration?

The transformation parameters namely translation and rotation between the consecutive frames of the video sequence used in this work is very small (not more than 3 pixels). Although the shift is small, image differencing between the consecutive frames results in some residual noise. In order to avoid it and also to align to near accuracy there is the need for registration. The method used here involves phase correlation which was discussed earlier but it will be dealt in detail with respect to the work in the following section.

There are two instances of image sequence considered in this work:

1) When the Space shuttle is getting ready to take off and moves a little bit
2) When the Space shuttle is in flight.

Frame # 14



Frame # 15

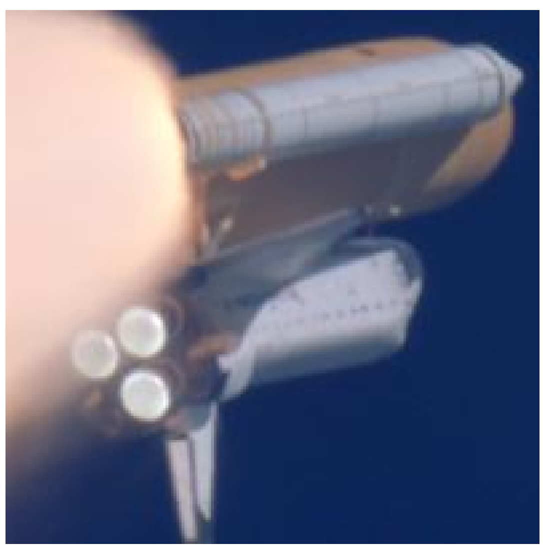


Figure 4.1 Consecutive Frames of the Shuttle Onboard

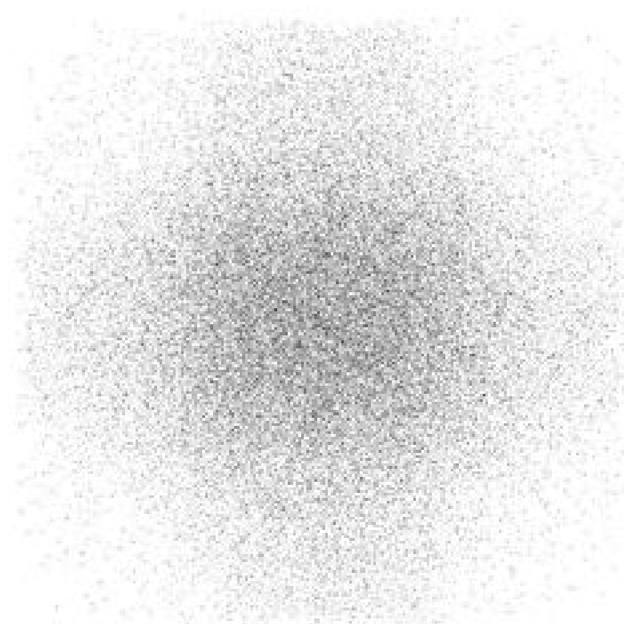


Figure 4.2 Absolute Magnitude of the frame # 14 in the Fourier domain

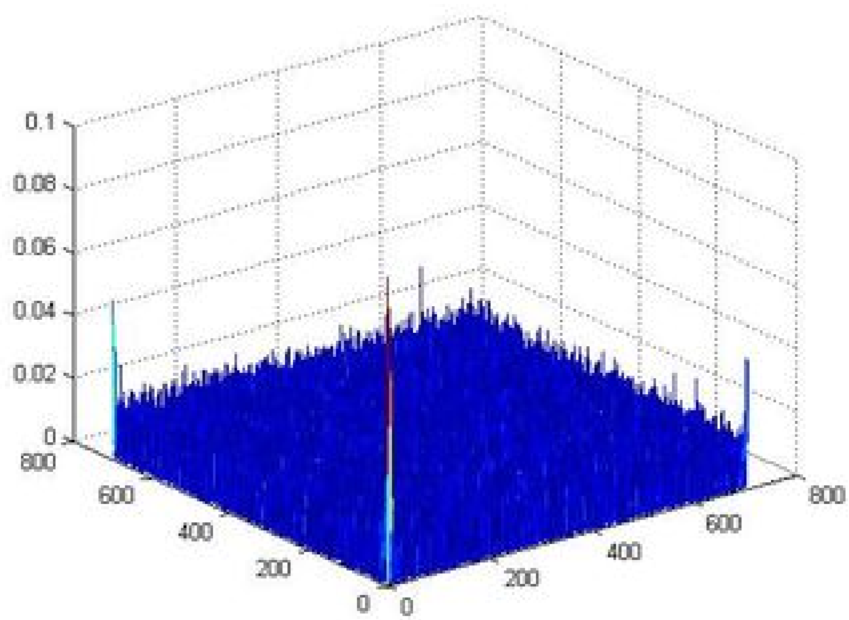


Figure 4.3 Normalized Cross Power Spectrum between the two frames

Frame # 77

Frame # 78

Figure 4.4 Consecutive Frames of the Shuttle on Ground

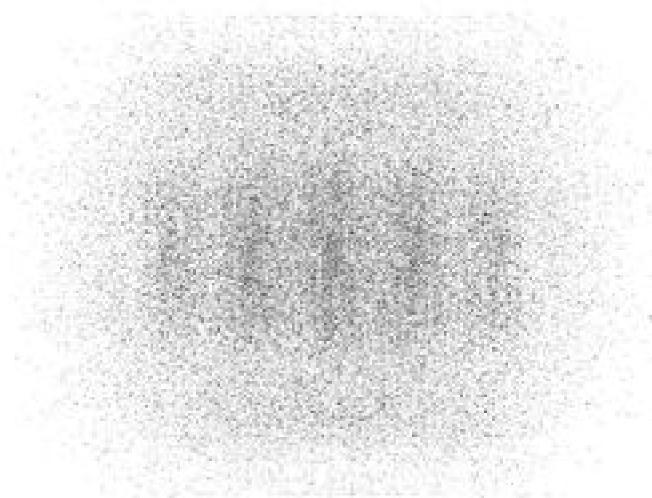


Figure 4.5 Absolute Magnitude of the frame # 77 in the Fourier domain

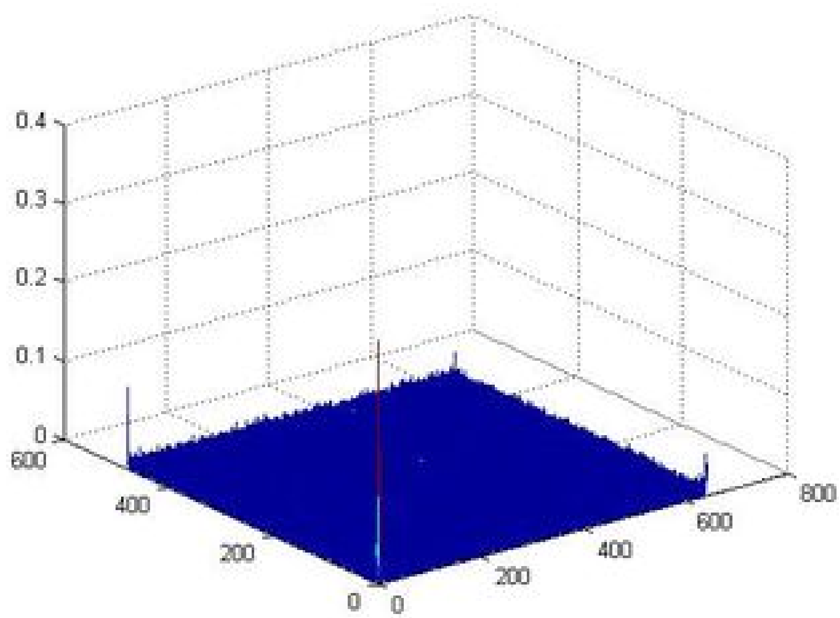


Figure 4.6 Normalized cross power spectrum between the two frames

Figure 4.1 represents two consecutive frames of the Space shuttle in flight with the falling of debris (foam) visible in one of the frames. Figure 4.2 represents the absolute magnitude of the image in the Fourier domain. As it can be seen most of the information lies in the center of the image. Figure 4.3 shows the cross power spectrum of the two images with the impulse function being maximum at that shift position in the x and y directions. Similarly Figure 4.4 shows the two frames when the shuttle takes off with a lot of debris and Figures 4.5 and 4.6 showing the absolute magnitude and cross power spectrum respectively. Both these sample images were obtained with a film camera. In both the cases space shuttle moves a little distance between the consecutive frames and so the background is not stationary. The first step would be to align the images so that further processing can be done. Here, registration is a kind of preprocessing step where in the template image is transformed with respect to the original image.

If there are two images $f_1(x, y)$ and $f_2(x, y)$ displaced by dx and dy given by

$$f_2(x, y) = f_1(x + dx, y + dy)$$

Then according to the Fourier Shift property their Fourier transforms are related by

$$F_2(\varepsilon, \eta) = F_1(\varepsilon, \eta)\, e^{-j2\pi(\varepsilon\, dx + \eta\, dy)}$$

The normalized cross power spectrum is given by:

$$\mathrm{Corr}(\varepsilon, \eta) = \frac{F_1(\varepsilon, \eta)\, F_2(\varepsilon, \eta)^*}{\left| F_1(\varepsilon, \eta)\, F_2(\varepsilon, \eta)^* \right|} = e^{j2\pi(\varepsilon\ dx + \eta\ dy)} \tag{34}$$

To solve for dx and dy take the inverse Fourier Transform of the normalized cross power spectrum given by:

$$\mathrm{Corr}(x, y) = F^{-1}(e^{j2\pi(\varepsilon\ dx + \eta\ dy)}) = \delta(x - dx, y - dy) \tag{35}$$

Where Equation 35 represents the dirac delta function centered at (dx, dy).

Therefore the translation values are given by the maximum value of the correlation given by:

$$(x_0, y_0) = \arg\{ \max \{ \mathrm{Corr}(x', y') \} \} \tag{36}$$

If the template image is a rotated version of the reference image phase correlation can be effectively implemented in polar coordinates to show rotation as a shift. When rotation and translation occur at the same time it poses a real challenge. When such cases arise rotation is found out first followed by translation. The phase of the cross power spectrum is given by:

$$\Phi(i, j) = \tan^{-1}(\mathrm{Corr}(\varepsilon(i, j), \eta(i, j))) \tag{37}$$

where φ represents the rotation angle. Although it offers better sensitivity to noise and computationally efficient, it can be applied to well defined geometric transformation namely translation and rotation. The algorithm can be represented as follows:

- Calculate the NCC Corr(x, y) from the equation (34)
- Calculate the phase shifts dx and dy after transforming it into the Fourier domain

- Move the template image by dx and dy to match with the original image.
- Repeat the steps till you get the maximum correlation(ideal condition is 1)

Need for Background Estimation and Foreground detection

After the consecutive frames are aligned the need for detecting moving object arises. Although the images have been registered there is no assurance that you get 100 % matching. So based on the information from the registration step the images are made to pass through Kalman Filter stage which involves in separating the foreground and background.

If the intensity of a pixel is given by f(x, y), state of the system is given by $\hat{f}$ (x, y, t_i) with respect to the estimated pixel background intensity and the estimated deviation is given by f^*(x, y, t_i) at t_i then the estimation is given by

$$\begin{bmatrix} \hat{f}(x,y,t_i) \\ f^*(x,y,t_i) \end{bmatrix} = \begin{bmatrix} f'(x,y,t_i) \\ f''(x,y,t_i) \end{bmatrix} + G(x,y,t_i).(f(x,y,t_i) - H(x,y,t_i)\begin{bmatrix} f'(x,y,t_i) \\ f''(x,y,t_i) \end{bmatrix}) \quad (38)$$

The prediction term is given by

$$\begin{bmatrix} f'(x,y,t_i) \\ f''(x,y,t_i) \end{bmatrix} = B.\begin{bmatrix} f'(x,y,t_{i-1}) \\ f''(x,y,t_{i-1}) \end{bmatrix} \quad (39)$$

With B = $\begin{bmatrix} 1 & b_{1,2} \\ 0 & b_{2,2} \end{bmatrix}$

In this work the background dynamics $b_{1,2} = b_{2,2} = 0.7$

The measurement matrix used is also constant given by

$$H = [1 \quad 0]$$

The Kalman gain is given by:

$$G(x,y,t_i) = \begin{bmatrix} g_1(x,y,t_i) \\ g_2(x,y,t_i) \end{bmatrix} \tag{40}$$

$$g_1(x,y,t_i) = g_2(x,y,t_i) \qquad = \alpha.m(x,y,t_{i-1}) + \beta.[1 - m(x,y,t_{i-1}] \tag{41}$$

Here, $m(x,y,t_{i-1})$ is a black and white image which represents the segmented images. Here the regions having intensity 1 forms the foreground while 0 represents the background. α and β represents the lower and higher threshold value of gain factor.

$$m(x,y,t_{i-1}) = \begin{cases} 1, if\ d(x,y,t_{i-1}) \geq thresh(x,y,t_{i-1}) \\ \\ 0, otherwise \end{cases} \tag{42}$$

Here, $thresh(x,y,t_{i-1})$ is chosen in such a way that intensity changes are adapted to represent the background in contrast to the foreground objects and:

$$d(x,y,t_{i-1}) = \left| f(x,y,t_{i-1}) - \hat{f}(x,y,t_{i-1}) \right| \tag{43}$$

There is a need for foreground adaptation in this approach since slow moving or non-continuously moving objects tends to occupy the background longer. This results in treating those objects as background if they stop for a while or there are chances of detecting them twice if they move afterwards. Therefore in this case Equation (41) becomes:

$$g_1(x, y, t_i) = g_2(x, y, t_i)$$

$$= \begin{cases} \alpha, if & [d'(x, y, t_i) \geq thresh(x, y, t_i)] \vee [d'(x, y, t_i) < thresh(x, y, t_i)) \wedge (d''(x, y, t_i) \geq thresh(x, y, t_i))] \\ \beta, if & [d'(x, y, t_i) < thresh(x, y, t_i)] \wedge [d''(x, y, t_i) < thresh(x, y, t_i)] \end{cases} \tag{44}$$

where:

$$d'(x, y, t_i) = | f(x, y, t_i) - f'(x, y, t_i) | \tag{45}$$

$$d''(x, y, t_i) = | f(x, y, t_i) - \hat{f}'(x, y, t_i) | \tag{46}$$

$$\hat{f}'(x, y, t_i) = f'(x, y, t_i) + \beta.[f(x, y, t_i) - f'(x, y, t_i)] \tag{47}$$

Equation (45) gives the difference between the actual measured value and the predicted value, (46) gives the difference between the measurement and the pre-estimation $\hat{f}'(x, y, t_i)$ which is calculated as given in Equation (47). If the difference between the actual intensity and the predicted value is greater than a threshold it belongs to the foreground and it is calculated using foreground gain α and if the difference is smaller it belongs to the background and a pre-estimation as given in Equation (47) is calculated using background gain β. If the difference between the pre-estimation and the measurement is greater than threshold it still belongs to the foreground and the calculation is done by α, otherwise the estimation becomes pre-estimation.

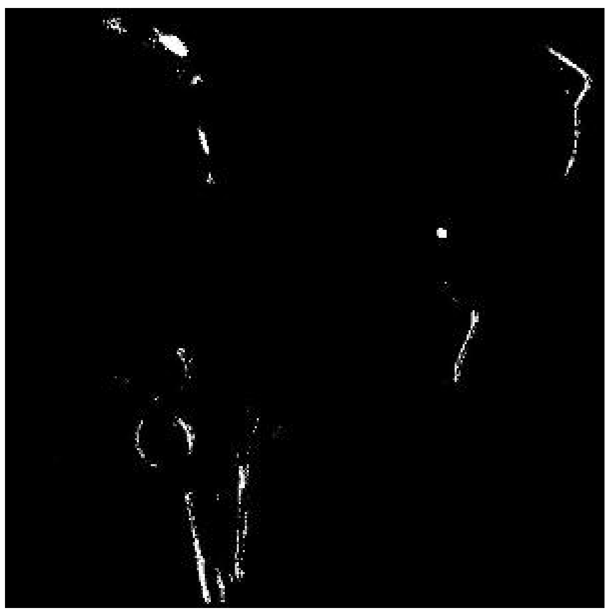


Figure 4.7 Mask Image of the Shuttle Onboard $(\alpha = 0.55 \quad and \quad \beta = 0.95)$

Figure 4.7 shows the mask image obtained after applying Kalman Filter to the registered image. The debris is clearly visible in the mask image but there is also error due to misregistration and also due to the parameters namely the gain parameters and the threshold value used in the Kalman filter implementation

Figure 4.8 Mask Image of the Shuttle on Ground $(\alpha = 0.55 \quad \textit{and} \quad \beta = 0.95)$

Figure 4.8 is for the second case where the algorithm worked well resulting in reducing registration errors due to proper selection of parameters of the Kalman filter.

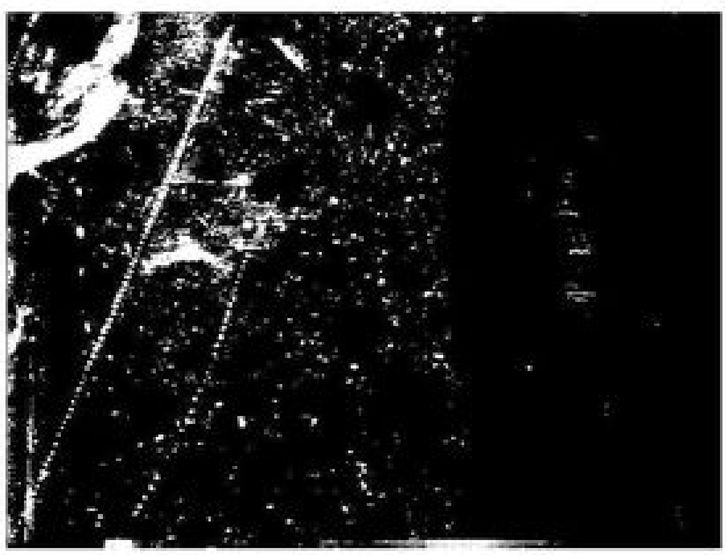


Figure 4.9 Accumulation of all the frames to depict the trajectory

Figure 4.9 involves in integrating all the frames to denote the path followed by the debris for the case when the shuttle is on Ground. It also gives a rough idea to remove those portions which does not follow a specific path as randomly occurring objects(which may appear for one or two frame).

Figure 4.10 Comparison of the frames with the different α *and* β values.

Figure 4.10 shows the same mask frame obtained by using two different α *and* β values. Left image shows the mask frame obtained with the best α *and* β value which is 0.55 and 0.95. The second frame is obtained with $\alpha = 0.25 \quad and \quad \beta = 0.75$ .As it is clearly seen there is some residual noise in the right image compared with the optimal one. This explains the importance for selecting a proper gain factors.

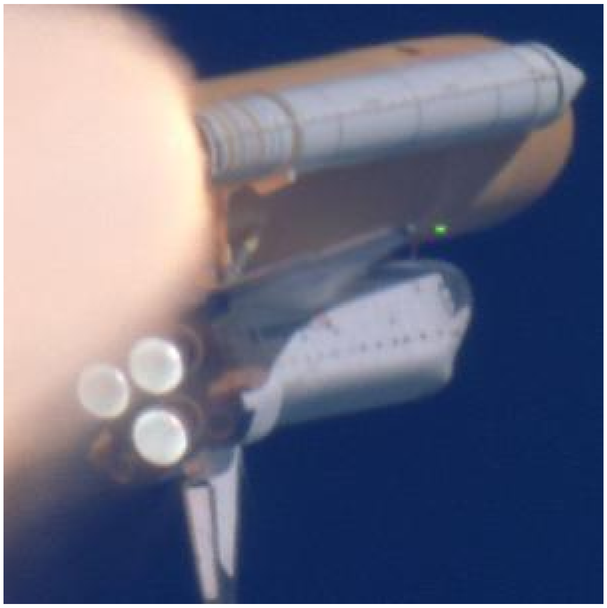
Representation of the moving regions with markers

This section forms the concluding step and it involves in marking the moving object based on the information obtained from the previous Kalman filtering stage. In this approach local threshold is used for detection of moving objects.

The difference image obtained from Kalman filter is divided into blocks of 5*5. Then the average intensity of each block is computed and it is compared with each pixel. If it exceeds a certain threshold, then that pixel is assigned a flag, otherwise it's made zero. This flag is highlighted for the pixels surpassing a predefined small threshold at each instant of the sequence. In simple the local thresholding technique can be shown as follows:

- Obtain the mask image from the Kalman filter
- Divide the image into 3*3 or 5*5 or 7*7 blocks
- Find the average intensity of each block
- Assign a small threshold and check if the average intensity of the block is above that threshold.
- If it is higher apply a flag for that block in that frame

Frame # 14



Frame # 20

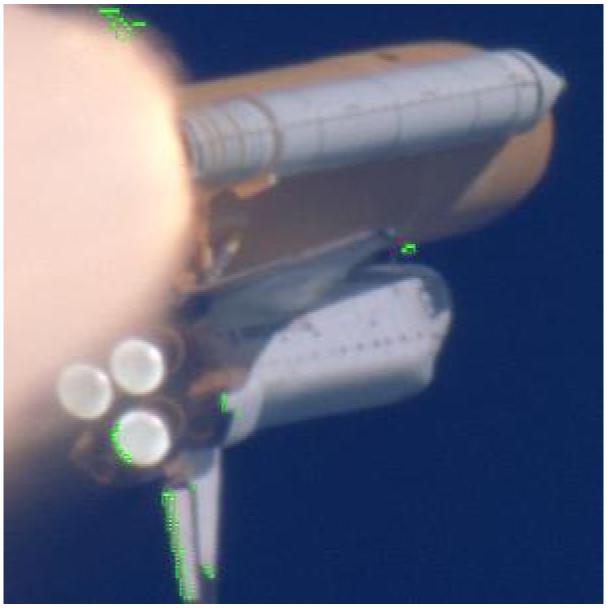
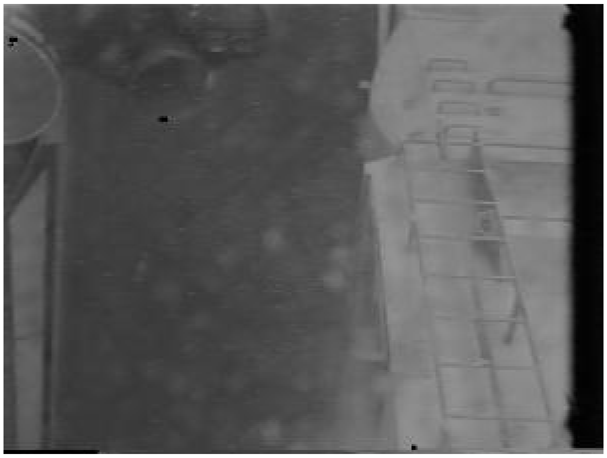


Figure 4.11 Marker Placed on Debris When the Shuttle is On Board

Frame # 144



Frame # 145

Figure 4.12 Marker for debris when the Shuttle is on Ground

CHAPTER FIVE: SUMMARY AND CONCLUSIONS

This work presented a methodology to detect any potentially dangerous debris occurring Space shuttle flights. The method used in this work is pretty straightforward with little complications. This algorithm may be optimized by testing it on many video sequences, keeping in mind this algorithm is developed for videos obtained during spacecraft launches. The effect of the foreground gain and the background gain and the difference between them to the system's performance depends on the initial grey value difference. Low resolution and low quality images are sufficient for the detection of the moving foreground objects. Slow as well as non-continuously moving objects can be detected with much ease.

A GUI demo has been developed during the course of this work which gives an option for the user to select a particular video and perform tracking. The main objective of developing this interface is to make the software user friendly as well as to reduce the amount of user interaction. Currently the GUI works for the three videos which we have been given, efforts are carried out to make it generic i.e. the user can input any video which follows certain constraints and perform registration, stabilization and tracking operation.

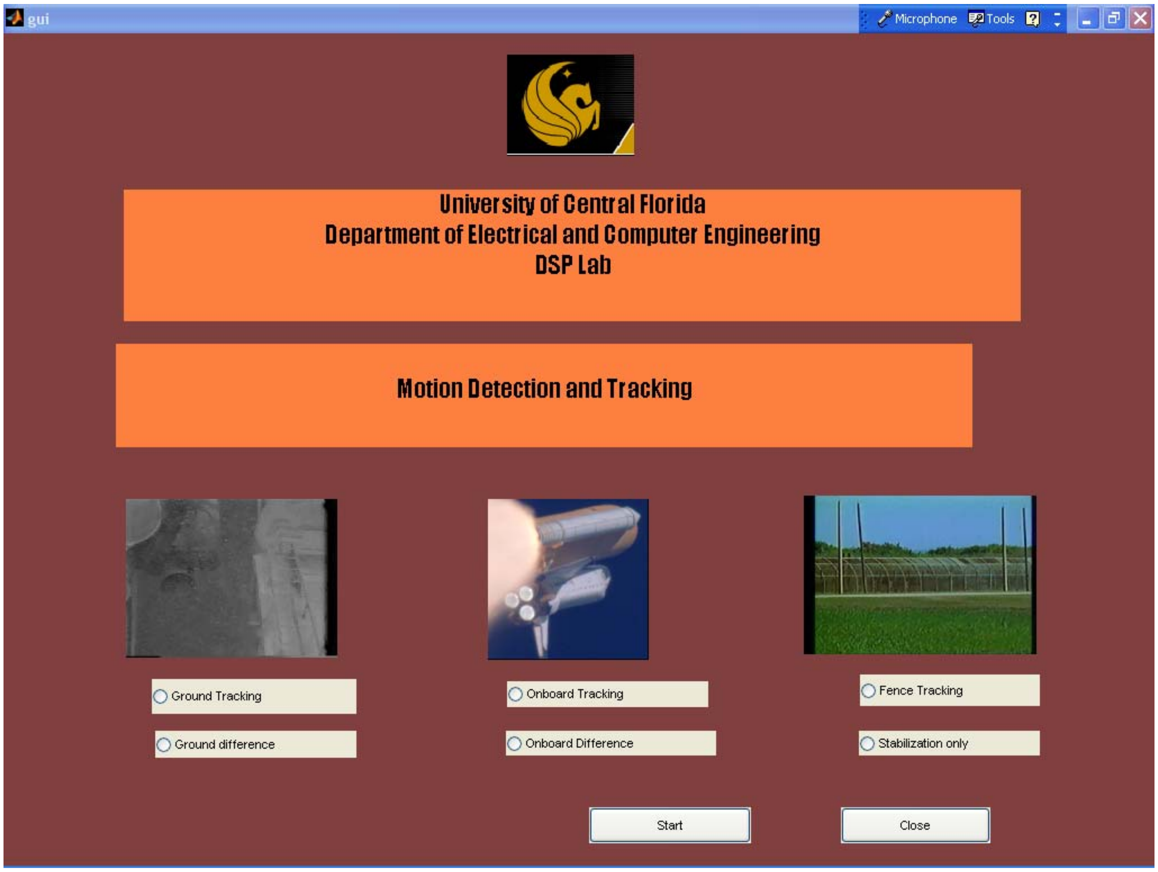


Figure 5.1 Screen shot of GUI Demo

This figure represents the screenshot of the GUI demo developed. The user has the choice of selecting any of the three operation namely registration, stabilization and tracking for a particular video. The input can be AVI, MPEG or image sequences and the user has the option of selecting the format of the output file.

CHAPTER SIX: FUTURE WORK

Currently the method can be applied only for images undergoing geometric transformation, so the extension of this work would be to make the algorithm applicable for problems involving 3-D projection, elastic body matching. A limitation of this method is that false alarms can be created around high intensity regions. A possible solution for this problem can be running this algorithm by considering more than 2 frames at a time. Another problem that needs to be dealt with is motion blur compensation. Motion blur is an effect which occurs when the objects are moving and it is clearly visible for fast moving objects in the scene or when they are exposed for a long time. Future work should also concentrate on taking camera motion into account so that it can be used for tracking objects obtained from a mobile camera. Kalman filter stage can be refined to calculate a variable and accurate threshold values for all the pixels.

LIST OF REFERENCES

[1] Lisa Gottesfeld Brown "A Survey of Image Registration Techniques", Course Notes, Department of Computer Science, Columbia University

[2] Katja Nummiaro, Esther Koller- Meier and Luc Van Gool "Color Features for tracking Non- Rigid Objects", *Special Issue on Visual Surveillance, ACTA Automatica Sinica* (Chinese Journal of Automation) 2003

[3] Paul J.Besl and Neil D.Mckay " A method for Registration of 3-D Shapes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No.2, Feb 1992

[4] J.B.Antoine Maintz and Max.A.Viergever "An Overview of Medical Registration Methods", *Medical Image Analysis*

[5] R. P. Woods, S. R. Cherry, and J. C. Mazziotta, "Rapid automated algorithm for aligning and reslicing PET images," *Journal of Computer Assisted Tomography*, Vol. 16, No. 4, pp. 620–633, 1992.

[6] D. L. G. Hill, D. J. Hawkes, N. A. Harrison, and C. F. Ruff, “A strategy for automated multimodality image registration incorporating anatomical knowledge and imager characteristics,” in *Information Processing in Medical Imaging*, H. H. Barrett and A. F. Gmitro, Eds. 1993, vol. 687 of *Lecture Notes in Computer Science*, pp. 182–196, Springer-Verlag, Berlin.

[7] Josien P. W. Pluim, J. B. Antoine Maintz and Max A. Viergever, ”Mutual information based registration of medical images: a survey”

[8] June-Sic Kim, Jong-Min Lee, Yong-Hee Lee, Jae-Seok Kim, In-Young Kim and Sun I.Kim, ” Intensity based affine registration including feature similarity for spatial normalization”, Comput Biol Med. 2002 Sep;32(5):389-402

[9] Carlos Ó. S. Sorzano*, Philippe Thévenaz, and Michael Unser, “Elastic Registration of Biological Images Using Vector-Spline Regularization”, *IEEE Transactions on Biomedical Engineering*, Vol. 52, No. 4, April 2005

[10] Martin Held, Werner Weiser, Franz Wilhelmstotter, “Fully Automatic Elastic Registration of MR Images With Statistical Feature Extraction”, *Journal of WSCG, Vol.12, No.1-3., ISSN 1213-6972* WSCG 2004, February 2-6, 2004, Plzen, Czech Republic.

[11] Morten Bro-Nielsen, claus Gramkow, “Fast fluid registration of medical images”, *Visualization in Biomedical Computing*. 4th International Conference, VBC '96 Proceedings

[12] H. Shekarforoush, M. Berthod, and J. Zerubia, “Subpixel image registration by estimating the polyphase decomposition of cross power spectrum,” in *Computer. Vision. Pattern Recognition.’96*, June 1994, pp. 532–537.

[13] Minghui Xia, Bede Liu, “Image Registration by Super-Curves”, *IEEE Transactions on Image Processing* 13 (2004) 720-732

[14] Y.C. Hsieh, D.M. McKeown, F.P. Perlant, “Performance evaluation of scene registration and stereo matching for cartographic feature extraction”, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14 (1992) pp. 214–237.

[15] Zerubia, Josiane B.; Foroosh, Hassam; “Extension of Phase Correlation to Subpixel Registration.” *IEEE Transactions on Image Processing*, Vol 11. No. 3, March 2002, pp. 188 – 202

[16] Christof Ridder, Olaf Munkelt and Harald Kirchner,”Adaptive Background Estimation and Foreground Detection using Kalman-Filtering”, *Proceedings of International Conference on recent Advances in Mechatronics, ICRAM'95*, UNESCO Chair on Mechatronics, pp. 193-199, 1995.

[17] Greg Welch and Gary Bishop,” An Introduction to the Kalman Filter” Course Notes, Department of Computer Science, University of North Carolina at Chapel Hill

[18] Hsieh, J.W., Liao, H.Y.M., Fan, K.C., Ko, M.T., Hung, Y.P.,” Image Registration Using a New Edge-Based Approach”, *CVIU(67)*, No. 2, August 1997, pp. 112-130

[19] Judit Verestóy, Dmitry Chetverikov and Marcell Nagy," Digital PIV: a Challenge for Feature Based Tracking" Machine Graphics and Vision, vol. 8, 1999, pp. 553-569

[20] Mubarak Shah, "Fundamentals of Computer Vision" ,Course Notes for CAP 5415,Fall 2004

[21] A. Averbuch, Y. Keller," FFT Based Image Registration" IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2002, Orlando, USA, May 2002.

[22] He Wang, Lei Dong, Jennifer O'Daniel, Radhe Mohan, Adam S Garden, K Kian Ang, Deborah A Kuban, Mark Bonnen, Joe Y Chang and Rex Cheung,"Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy", Phys. Med. Biol. 50 (2005) 2887–2905

[23] Bernd Fischer and Jan Modersitzki," Fast Curvature Based Registration of - MR-Mammography Images" "Bildverarbeitung für die Medizin",Springer, 139-143, 2002.

[24] Hue, C.; Le Cadre, J.-P.; Perez, P**,"** A particle filter to track multiple objects", Multi-Object Tracking, 2001. Proceedings. 2001 IEEE Workshop on 8 July 2001 Page(s):61 – 68

[25] DePiero, F,**"** Fast landmark-based registration via deterministic and efficient processing, some preliminary results", 3D Data Processing Visualization and Transmission, 2002. Proceedings. First International Symposium on 19-21 June 2002 Page(s):544 – 548

[26] Jeremie Bigot, "Landmark-based registration of curves via the continuous wavelet transform", http://www.lsp.ups-tlse.fr/Fp/Bigot/Papers/AlignFANOVAfinal.pdf

[27] Shaohua Zhou; Chellappa, R.; Moghaddam, B ,"Adaptive Visual Tracking and Recognition using particle filters", Multimedia and Expo, 2003. ICME '03. Proceedings. 2003 International Conference on Volume 2, 6-9 July 2003 Page(s):II - 349-52 vol.2