Target Tracking and Range Estimation using an Image Sequence

Raj Talluri† and W. Clay Choate

Texas Instruments,
Central Research Laboratories,
Dallas, Texas 75265.

Abstract

This paper describes a new method of tracking targets across a sequence of images and estimating the range of these tracked objects from the camera. The motion of the camera between the successive frames is assumed to be known. The distance from the sensor to the objects is assumed to be much larger than the physical dimensions of the object. The approach presented makes use of the known sensor motion to generate Expected Images which are then used in establishing a reliable correspondence and in tracking the the targets across the image sequence. A mean normalized area correlation method is used to establish the inter-frame correspondence. Once this correspondence is established, two new methods of estimating the range of the target from the sensor are also developed. The techniques developed are tested with good success rate on a sequence of real aerial images.

1 Introduction

Passive ranging is an important area of application of computer vision research. The need to estimate the range of the objects from the on-board sensor is particularly important in applications where it is not desirable to use active sensors due to power limitations or if covertness of operation is desired. This is one of the central issues in the *nap-of-the-earth* helicopter guidance [5, 6] and autonomous planetary missions.

Using a sequence of visual images to estimate the structure and motion of the sensor has been considered by a large number of researchers in computer vision [1, 7]. By using a sequence of images it is not only possible to estimate the range of the objects from the sensor but also to compute the motion of the sensor. Broadly the techniques in motion analysis can be divided into two categories : 1) optical flow based methods 2) feature based methods. Anandan [2] reviews the work done by various researches on both these approaches. The optical flow based approaches compute the *optic flow* or the two dimensional field of instantaneous velocities of the gray values in the image plane. This optic flow is then used to along with additional constraints regarding the scene to compute the structure of the scene and the motion of the sensor. The feature based methods extract a relatively sparse set of features of interest such as edges, lines, regions etc. using a 2-d feature detector from the images. Inter-frame correspondence is then established between these features. Using constraints such as rigid body motion, and these established correspondences the 3-d structure of the scene and the motion of the sensor can be computed. With the advent of accurate position determination systems such as Inertial Navigation Systems (INS) and Global Positioning Systems (GPS) it is now possible to compute quite precisely the motion of the sensor in the environment. So, the passive ranging methods use the known motion and only concentrate on computing the range of the objects of interest from the sensor [5, 6].

In all the approaches studied so far, the fundamental drawback is that if the motion of the sensor is known *a priori*, this is not used effectively in the feature extraction and the correspondence process. The feature-based methods extract features that essentially depend only on the local grey scale changes. As a result when the sensor motion is considerable, the object shape may change significantly and the features extracted for matching could be significantly different from those before and thereby the procedures used to establish correspondence will not be reliable and hence the estimated range will be erroneous. Dutta and Snyder [3] show that small errors in correspondence can lead to large errors in the inferred motion and large errors in inferred depth. Even in the case when the motion is known *a priori*, small errors in image displacement can lead to large errors in depth for points more that a few multiples of the baseline from the sensor. Similar problems plague the intensity based methods in that the grey value distributions of the object in the second image in the sequence may be considerably different from the first one.

In this paper we argue that in most passive ranging applications the motion of the sensor is known *a priori* to a reasonable degree of accuracy. So, if this known sensor motion is used effectively in establishing the correspondence, the range estimated will be more accurate and reliable. We discuss a novel method of generating motion compensated images called *Ex-*

† Presently with The Computer and Vision Research Center, The Department of ECE, ENS 523, The University of Texas at Austin, Austin, Texas 78712.

0-8186-2840-5/92 $03.00 © 1992 IEEE

pected Images (EIs) from the original image sequence using the known sensor motion and an assumed range to the object. These EIs are used in a mean-normalized area correlation method in establishing a correspondence between the objects to be tracked across the image sequence. This technique of explicitly making use of the sensor motion in the feature selection and correspondence process so as to achieve a more reliable and robust correspondence is one of the original contributions of this paper. Also presented are two iterative methods of estimating the range to the tracked objects. The first method uses the established inter-frame correspondence across a number of frames to iteratively refine the assumed range to the objects being tracked till it converges to the true value. The second method uses the correlation co-efficient value in refining the assumed range so as to arrive at the true range. The methods of range estimation presented are simple, efficient and are shown to give accurate results when tested on a sequence of real aerial images taken from a helicopter. Our work relates the work in passive ranging and target tracking in that, it uses the known sensor motion and an assumed range in establishing an accurate correspondence and also uses this correspondence in refining the assumed range so as to finally arrive at both a reliable correspondence and an accurate range estimate.

Section 2 describes briefly the approach used to track the objects and the estimate the range. Section 3 describes in detail the generation of the *Expected Image* used in the correspondence. Section 4 describes the mean normalized correlator used to establish the correspondence. Section 5 details two methods to determine the range to the tracked objects. Section 6 presents the results of the techniques developed in this report when tested on a sequence of aerial images taken from a helicopter.

2 The Approach

In this research we attempt to rectify the drawbacks of the traditional approaches by using the known sensor motion to predict and regenerate the *Expected Image* (EI) of the object and use this EI in establishing the correspondence. This method draws from the work presented by Merchant [4] in the context of image registration. Merchant derives the transformations relating two different views of a two or three dimensional object scene as a function of the displacement of the camera. He applies these transformation to the task of image registration.

Consider an image sequence $I_1, I_2, ...I_n$ across which we are interested in tracking the positions of a set of objects. Let the sensor locations corresponding to these images be represented by $S_1, S_2, ...S_n$. Consider a correlation based technique. We extract a template T_1 centered around the object O_1 from the image I_1. In the traditional approach we correlate T_1 with I_k to establish its position at the location S_k. Then using the known camera geometry and the known sensor motion between S_1 and S_k, represented by a rotation matrix R_{1k} and a translation vector T_{1k}, we can estimate the range to the object O_1.

In our approach we propose to use I_k and the known sensor motion given by R_{1k} and T_{1k} to generate the *Expected Image* (EI_{1k}) that would be seen at I_1. The Expected Image in thus a motion compensated, warped version of the original image. We then correlate T_1 with EI_{1k} to determine the position of O_1. Once this location is established, we can estimate the range of O_1. The central theme of this approach is that since we use R_{1k} and T_{1k} to generate EI_{1k}, EI_{1k} is more *similar* to I_1 than I_k. Thus, correlating T_1 with EI_{1k} gives a more reliable estimate of the location of O_1 in the image I_k. This procedure can be used across the image sequence i.e., using I_n and (R_{1n}, T_{1n}) to generate E_{1n} and use E_{1n} instead of I_n in establishing the correspondence with I_1. Figure 3 shows an example of the image I_1. Figure 4 shows the image at I_7. Figure 5 shows the expected image generated using I_7.

2.1 Image generation

It is not possible to exactly reconstruct the image EI_{1n}, given just I_1, I_n and sensor motion. We need to know the entire scene structure and the lighting conditions also. So, using the hypothesis that the range of the object O_1 is large compared to its dimensions, we make the simplifying assumption that all the points on the object O_1 lie on a plane Ω. The orientation of the plane is assumed to be normal to the line of sight (LOS) at the sensor location S_1 where I_1 was imaged. This assumption models the reality quite well for a certain length k of the sequence. After this, if the sensor motion is significant we begin to see parts of the object that were not visible before in the original image I_1. We then update the template to be T_k (extracted form the image I_k) and the image to establish correspondence with to be I_k and continue the process. Even though our method may require template updates like the previous methods, the life of the template is however much longer. In the tests we ran the life of the template is typically increased three fold.

In the above method of image reconstruction we also need to know the range of the plane Ω from the sensor position S_1. This is essentially the range to the object O_1 which we are interested in computing after establishing the correspondence. So we propose to use a iterative procedure wherein we hypothesize a certain range to the plane Ω from the sensor, use this to establish the correspondence, estimate the range using the correspondence and thus refine the range estimate iteratively. This procedure finally converges to the best range estimate. Also note that since we use the range of one object O_1 only in defining the plane Ω, in the EI generated only this object's motion is corrected for

exactly. Thus if we desire to track multiple objects, it is necessary to repeat this procedure for each of the objects.

2.2 Estimating the range

Once we establish the location in EI corresponding to the template T_1, we can then transform this to the image I_k (using the inverse of the motion transform given by R_{1k} and T_{1k}) to get the corresponding location. Thus we have two images from two known sensor positions S_1 and S_k and two corresponding locations in the image planes. These can used to estimate the range to the object corresponding to the image plane locations. In this research we have also developed a simple but effective method to estimate the range using the image correspondences. The basic idea of the approach is to project two rays from the two sensor locations through the image plane points and determine the *point of closest approach* of these two rays. The range of this point gives the range of the object. When we have multiple sightings of the object, (corresponding points at various sensor locations), we develop a weighted least square technique to combine this data into a more reliable range estimate. The details are given in Section 5.

It is observed that the correlation coefficient of the result of correlating the template with the EI peaks when the assumed range to the object is the true range. This suggests the use of the correlation coefficient in the range estimation process. We have also developed a method to estimate the range to the object using the correlation coefficient. This is also described in Section 5.

3 Generation of the Expected Image

Consider a right handed world co-ordinate system (WCS) $OX_wY_wZ_w$ with X_w pointing North, Y_w pointing East and Z_w pointing towards the center of the Earth. Let $OX_{S_i}Y_{S_i}Z_{S_i}$ be the sensor co-ordinate system (SCS_i) at the sensor location S_i. Let the transformation that takes the WCS into SCS_i be given by a translation vector $\vec{T}_{WS_i}$ followed by a rotation matrix $[R_{WS_i}]$.

Thus if $\vec{p'} = (x_{S_i}, y_{S_i}, z_{S_i})$ is a point in SCS_i and $\vec{p} = (x_W, y_W, z_W)$ is its image in WCS then

$$\vec{p'} = [R_{WS_i}](\vec{p} + \vec{T}_{WS_i}) \tag{1}$$

or

$$\vec{p} = [R_{S_iW}]\vec{p'} + \vec{T}_{S_iW} \tag{2}$$

where, $\vec{T}_{S_iW} = -\vec{T}_{WS_i}$ and $[R_{S_iW}] = [R_{WS_i}]^{-1} = [R_{WS_i}]^T$. The matrix $[R_{WS_i}]$ is defined in terms of Euler angles. They are the heading angle ψ, attitude angle θ and the bank angle ϕ. The rotation sequence is ψ about Z-axis, θ about Y-axis followed by ϕ about X-axis, where all the rotations are are considered positive in the right-handed sense. So, given the sensor position S_i in the WCS as (X_{sw}, Y_{sw}, Z_{sw}) and the attitude of the sensor by its three Euler angles $(\psi_{sw}, \theta_{sw}, \phi_{sw})$ we can compute the rotation matrix $[R_{WS_i}]$ and the translation vector as $\vec{T}_{WS_i} = (-X_{sw}, -Y_{sw}, -Z_{sw})^T$.

Let the image plane be oriented along the rows and columns of the sensor array. Let u point along the rows and v along the columns. The image plane axes originate at the image center, given by (u_0, v_0), where the optical axis meets the image plane. The pixel co-ordinates n_u and n_v originate at the upper left hand corner and run in the same direction as u and v. If the aspect ratio is given by AR and the focal length of the lens expressed in pixels is given by f, then, the co-ordinates of a point (n_u, n_v) in the sensor co-ordinate system are given by

$$\begin{aligned} x_{S_i} &= f \\ y_{S_i} &= (n_u - u_0)AR \\ z_{S_i} &= (n_v - v_0) \end{aligned} \tag{3}$$

Now, given two sensor positions S_1, S_k and the images I_1 and I_k at these locations, we detail below the procedure to generate the *Expected Image*, EI_{1k} from I_k. Figure 1 illustrates the image generation process.

Let O'' be the origin of the co-ordinate system at S_k. Consider a point $p(n_u, n_v)$ in the image plane I_k. We form its co-ordinates in the SCS_k using Equation (3) as P''. Considering the line $L_{S_k} = \overline{O''P''}$, and transforming this into the WCS using Equation (2) we have the line L_W. See Figure 1. Now let Ω_{S_1} represent a plane in SCS_k at an assumed range d_x from the sensor location S_1 and normal to the LOS at S_1. This plane is assumed to contain the object points. We can represent this plane in the point-normal form and transform it into the WCS using Equation (2) as Ω_W. We find the intersection of L_W and the plane Ω_W. Let this intersection be Q_W. We now transform Q_W into the SCS_1 using Equation (1) as $Q_{S_1}(x_{S_1}, y_{S_1}, z_{S_1})$. Using the perspective projection the image of this point in the image plane at S_1 is given as $q(n_u, n_v)$ where, $n_u = u_0 + \frac{y_{S_1}}{x_{S_1}} fAR$, and $n_v = v_0 + \frac{z_{S_1}}{x_{S_1}} f$.

We repeat this procedure for all the pixels $p(n_u, n_v)$ in the image plane I_k and the image formed at S_k is the required motion compensated image EI_{1k}. Using this procedure all the pixels in the image EI_{1k} may not filled by the image from I_k. To keep the EI_{1k} the same size as I_1 we set all the unfilled pixels to black. This can be seen in Figure 5. Note that we have used the sensor motion between the two frames I_1 and I_k as the given by the sensor position and orientation in the WCS at the two locations S_1 and S_k.

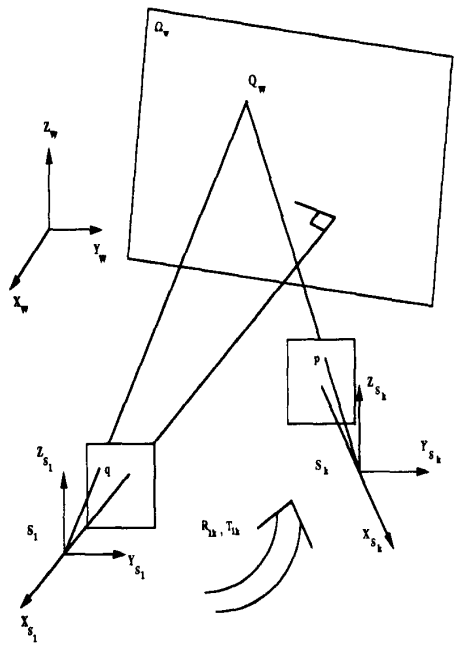

Figure 1: The geometry of the image generation

4 Correlation

Once we generate EI_{1k} from I_k we use an area correlation technique to establish the location of the object in it. Let B represent the template containing the object O_1 in the image I_1. Let A be the expected image EI_{1k}. The size of the template (M, N) is taken to be odd. B is moved all over A and the value of the mean normalized correlation between A and B, ρ, is computed [6]. A value of 1.0 indicates a perfect match between the two regions. The normalized correlation coefficient has the important property that the correlation value is independent of the linear brightness changes from one region to the other. We take the pixel location of the central pixel of the template corresponding to the highest value ρ to be the best estimate of the position of the object O_1 in the expected image EI_{1k}. However to make the approach more robust we need sub-pixel accuracy in the pixel location. To achieve this, after locating the pixel position corresponding to the maximum value of ρ we fit a a second order surface to the correlation coefficient values in a 3x3 window about this pixel. We then evaluate the peak of this surface by evaluating the partial derivatives with respect to i and j and setting them to zero. The location of the peak of this correlation surface is then taken to be the best estimate of the object location on the expected image. In out tests we observe that this technique of achieving sub-pixel accuracy by interpolating the correlation surface has the effect of making the tracker track the object more closely between the frames and hence makes the iterative range estimation procedure described in Section 5 converge smoothly to the final value.

5 Range Estimation

In this section we discuss two different methods of estimating the range of the object from the sensor. The first method uses the established correspondence between the image frames to estimate the range. The second method uses the correlation coefficient and searches for the true range. In our experiments with a sequence of helicopter images we find that the range estimated using both these methods are of comparable accuracy. However the former method is less computationally intensive.

5.1 Range estimation using the *point of closest approach* method

Given a pair of images I_1 and I_k at sensor locations S_1 and S_k respectively and the known motion transformation between the two sensor locations by R_{1k}, T_{1k}, we compute the Expected Image at S_1 as EI_{1k} using I_k and an assumed range to the object r. We then correlate the template T_1, centered around the pixel location p_1, and containing the object O with this EI_{1k} and consider the pixel location corresponding to the correlation peak as the location of the object in EI_{1k}. Once this location is identified, using the inverse of the motion transformation between S_1 and S_k, we can compute the pixel location p_k in I_k corresponding to the object O. Thus, we have a correspondence between two pixel locations p_1 and p_k in the images I_1 and I_k at the two known sensor locations S_1 and S_k. Repeating this procedure for a sequence consisting of n images $I_1, I_2, ..., I_n$, we have a correspondence between n pixel locations p_1, $p_2, ..., p_n$, at the sensor locations $S_1, S_2, ..., S_n$.

One way to find the range of the object O from the sensor is to consider n lines drawn from the sensor locations $S_1, S_2, ..., S_n$ through the pixel locations $p_1, p_2, ..., p_n$ and find their intersection. To account for the inaccuracies in the procedure used for establishing the correspondence we compute the *point of closest approach* of these n lines. Let the range estimated by this procedure be denoted by r_{est}. Figure 2 illustrates this. The details of the derivation of the point of closest approach of two 3-d lines is given below. We now modify the assumed range r to be the estimated range r_{est} and iterate till it converges. Section 6 discusses the convergence properties of the algorithm when tested on a real image sequence.

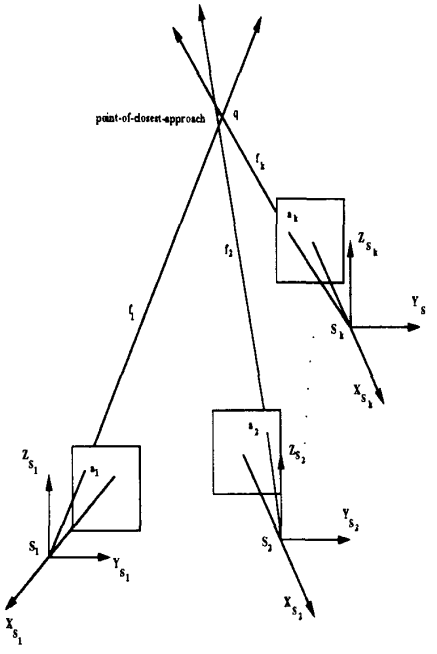

Figure 2: Range estimation by point of closest approach

5.1.1 Point of closest approach of n lines

Let the n lines be represented in the point normal form as $\mathbf{L_i} : \vec{a_i} + \alpha_i \vec{f_i}$ $(i = 1, \ldots, n)$.

Let $\vec{q} = [q_x, q_y, q_z]^T$ be the point of closest approach of the n lines. If the perpendicular distance from the line $\mathbf{L_1}$ to $\vec{q}$ is denoted by δ then

$$\delta^2 = \|\vec{q} - \vec{a_1} - \alpha \vec{f_1}\|^2. \tag{4}$$

If $\hat{\alpha_1}$ is the α_1 that minimizes δ^2 then

$$\frac{\partial \delta^2}{\partial \alpha_1} = -2\vec{f_1} \cdot (\vec{q} - \vec{a_1} - \hat{\alpha_1}\vec{f_1}) = 0. \tag{5}$$

Thus $\hat{\alpha_1} = \vec{f_1} \cdot (\vec{q} - \vec{a_1})$ and $\mathbf{L_1} : \vec{a_1} + \vec{f_1} \cdot (\vec{q} - \vec{a_1})\vec{f_1}$ or

$$\begin{aligned} \hat{\delta^2} &= \|\vec{q} - \vec{a_1} - \vec{f_1} \cdot (\vec{q} - \vec{a_1})\vec{f_1}\|^2 & (6) \\ &= \|\vec{q} - \vec{a_1}\|^2 - \langle \vec{q} - \vec{a_1}, \vec{f_1} \rangle & (7) \end{aligned}$$

where we have used $\langle \vec{x}, \vec{y} \rangle$ to denote $\vec{x} \cdot \vec{y}$.

Now if we have n lines, we attempt to seek $\hat{\vec{q}}$ such that

$$J = \frac{1}{n} \sum_{i=1}^{n} \left[\|\vec{q} - \vec{a_i}\|^2 - \langle \vec{q} - \vec{a_i}, \vec{f_i} \rangle^2 \right] \tag{8}$$

is a minimum. Thus

$$\nabla_{\vec{q}} J|_{\vec{q}=\hat{\vec{q}}} = \vec{0}. \tag{9}$$

Evaluating the gradient, we get

$$\nabla_{\vec{q}} J = \frac{2}{n} \sum_{i=1}^{n} ([I] - \vec{f_i}\vec{f_i}^T)(\vec{q} - \vec{a_i}) \tag{10}$$

where $[I]$ is the identity matrix. Let

$$\begin{aligned} [F_i] &= [I] - \vec{f_i}\vec{f_i}^T & (11) \\ [\bar{F}] &= \frac{1}{n} \sum_{i=1}^{n} [F_i] & (12) \\ \bar{\vec{r}} &= \frac{1}{n} \sum_{i=1}^{n} [F_i]\vec{a_i} & (13) \end{aligned}$$

then setting the gradient equal to the null vector, $\vec{0}$, we get

$$[\bar{F}]\vec{q} = \bar{\vec{r}} \tag{14}$$

which is a linear system and can be solved for $\vec{q}$, the point of closest approach. Note that the normalization by n is optional provided $[\bar{F}]$ and $\bar{\vec{r}}$ are treated consistently.

As the sensor moves further away from the initial position and closer to the object, the assumption that all the object points lie in a plane starts to break down. We begin to see new structure of the object which was not visible in previous frames. Hence the EI generated will start to be less *similar* to the original image. As a result the correlator output begins to drop and the correspondence established using this method becomes less reliable. To account for these effects we argue for the use of a weighted least squares estimation procedure in the range estimation. Since the value of the correlation coefficient is a good measure of the mismatch between the images, this value is used in the weight generation. From our experiments we find that using a weight $w_i = \rho_i^3$, where ρ_i is the correlation coefficient associated with the ith image proves to be quite effective.

To form the weighted least squares we modify Equation (8) to be

$$J = \frac{1}{n} \sum_{i=1}^{n} w_i \left[\|\vec{q} - \vec{a_i}\|^2 - \langle \vec{q} - \vec{a_i}, \vec{f_i} \rangle^2 \right]. \tag{15}$$

Propagating the weights through, we have to modify $[F_i]$ in Equation (12) to be

$$[F_i] = w_i([I] - \vec{f_i}\vec{f_i}^T). \tag{16}$$

Also if the sensor positions corresponding to the lines used in the range estimation are close to each other the

Figure 3: Image at S_1

Figure 4: Image at S_7

range estimated by the point of closest approach is not very reliable. To account for these in our experiments we only considered those sensor positions which are a certain threshold distance away from the initial position.

5.2 Correlation based method

Recall that in the generation of the EI we make use of the assumed range to the object r. It is observed from our experiments that the EI generated is most *similar* to the original image when the assumed range is closest to the true range of the object. This is quantified by the value of the of the correlation coefficient when the template containing the object is correlated with the EI. This suggests the use of the correlation coefficient as a guide in searching for the true range of the object. The procedure we adopted was to assume a certain range r of the object, generate the EI evaluate the correlation coefficient and update the assumed range such that the correlation coefficient peaks. The direction and the increments to the assumed range can be determined by observing the change in the correlation coefficient as the assumed range *sweeps* through different values. Figure 8 shows a plot of the correlation coefficient versus the error in the assumed range. As can be seen the correlation coefficient peaks when the assumed range is closest to the true range value. However, this method of range estimation may be computationally more expensive than the previous method.

6 Results and Observations

The methods of tracking and range estimation detailed in the previous sections have been tested on real image sequences obtained from NASA. The data consists of a sequence of images taken from a helicopter flying at a low altitude over a set of stationary trucks on an airport runway. The data set consists of ten images each separated in time by 0.3 seconds.

Figure 5: The expected image EI_{17}

In this image sequence the helicopter is flying in approximately along a line towards the trucks. Figure 3-4 show two of the images of the sequence taken by the helicopter at locations S_1 and S_7 respectively. The location of the helicopter and the sensor at these locations is known *a priori*. A 32 x 32 pixel square region enclosing the right most truck in Figure 3 is used as the correlation template, T_1. This truck is at a true range of 272.6 feet from the sensor. The Figures 5 shows the motion compensated images (EIs) using a plane at the true range of this truck along the LOS. As can be seen from the images the motion compensation transforms the truck back to its right size, shape and orientation at S_1. The black areas in the images arise due to the fact that since not all pixels in the EIs are filled by the image from S_k, to keep the EIs to be of the same size as the original image, the pixels that are not filled are set to black.

Figure 6 illustrates the effects of the motion compensation on the correlation co-efficient. Here, the template is correlated with the various images in the sequence and the value of the correlation coefficient is plotted. The plot also shows the values of the correlation co-efficient when correlated with the motion com-

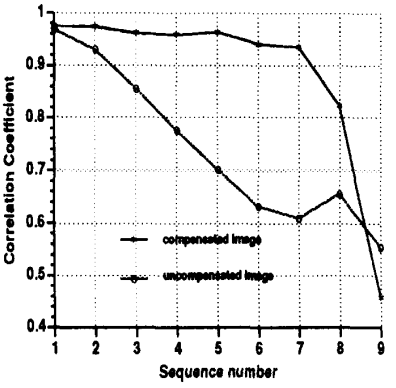

Figure 6: The correlation coefficient for different images in the sequence

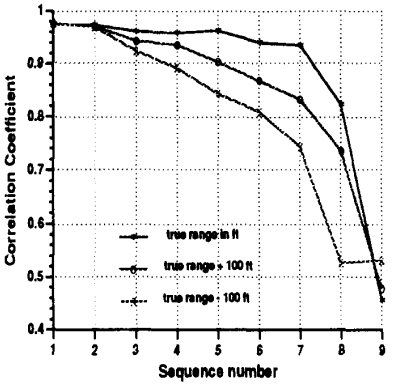

Figure 7: The correlation coefficient for different images in the sequence

pensated images. As can be seen from the plot, the coefficient is much higher for the compensated images and remains higher for a greater part of the sequence. Towards the tail of the sequence the truck is out of the field of view of the sensor and hence the coefficient is low in both cases. This curve illustrates the fact that this method of motion compensation does prolong the life of the template significantly and results in a more reliable tracking.

Figure 7 shows a plot of the correlation coefficient versus sequence number with the range of the plane Ω as a parameter. As can be seen the correlation coefficient is higher when the plane is assumed to be at the true range to the object of 272.6 feet. This suggests he use of the correlation coefficient in the range estimation process.

Figure 8 shows the plot of the correlation coefficient versus the error in the assumed range of the plane. Here the EI at S_1 is generated using the image at S_7.

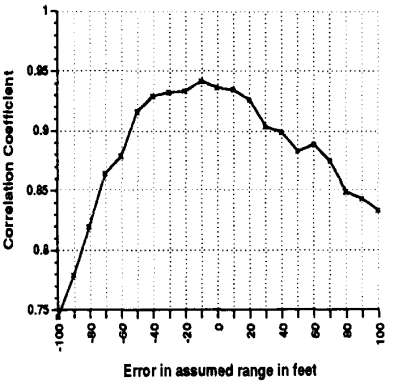

Figure 8: A plot of the correlation coefficient vs the error in assumed range

The figure illustrates that the correlation coefficient peaks when the error in the assumed range is close to zero. This suggests that it is possible to estimate the true range to the object in an iterative fashion by assuming a range for the plane, compensating the image, computing the correlation coefficient and searching along the assumed ranges of the plane till the correlation coefficient peaks. The figure also illustrates that the correlation coefficient falls rapidly as the assumed range gets much smaller compared to the true range. The rate of fall of the coefficient as the assumed range gets larger than the true range is not as steep. This suggests that it is better to initially overestimate the range and then converge to the true range iteratively.

Figure 9 shows a plot of the estimated range versus the sequence number using the *point of closest approach* method applied to two images. As can be observed the range estimate gets more accurate as the sequence number grows. This is because, with increasing sequence number the distance between the current and the reference sensor position gets larger and hence the base line increases. The two projected rays thus, have a greater angular separation between them yielding a more reliable and accurate estimate.

When the *point of closest approach* method is applied using all the ten frames in the sequence in a weighted least squares we find the range estimate to be about 3 feet away from true value. The weights used were the cube of the correlation coefficient. Figure 10 shows the effects of the error in the assumed depth of the plane Ω on the range estimate. It can be observed that the range estimate is most accurate when the assumed depth is close to the true depth. This suggests a iterative estimation procedure where in, we assume a depth of the plane Ω, generate the motion compensated image, determine the location of the object using a correlation template, transform this location into the second image frame, use the *point of closest approach*

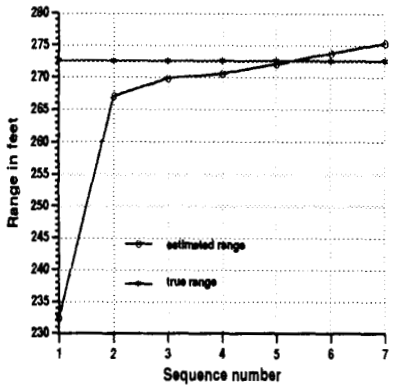

Figure 9: The range estimated using different images in the sequence

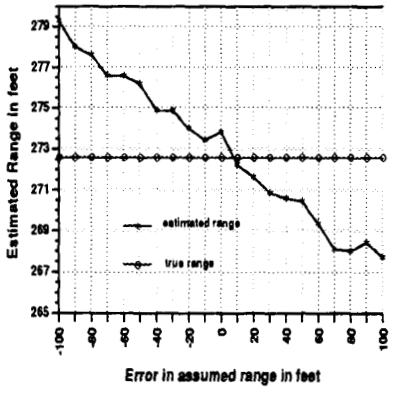

Figure 10: The effect of error in assumed range on the range estimation

to estimate the depth, then use this depth estimate to refine the assumed plane depth and iterate till it converges. From the shape of the curve it can be seen that such a depth estimation procedure converges to a range which is quite close to the true range. In the Figure 10 the true range to the object is 272.6 feet. The convergence of the iterative range estimation algorithm can be explained as follows: suppose we assume the object to be at a range of 350 feet, i.e., a range error of 77.4 feet, the estimated range from the curve is about 268 feet, which is off from the true range by less than 5 feet. This indicates the fast rate of convergence of the algorithm.

7 Conclusions

In this paper we have presented a new method for tracking objects across a sequence of images. The known motion of the sensor between the images is used to transform the images and generate *Expected Images*. These *Expected Images* are used to establish a reliable correspondence and track the objects of interest. A mean normalized area correlation method is used to establish the inter-frame correspondence. Once this correspondence is established, two methods of estimating the range to the objects are presented. The algorithms are tested on a sequence of real aerial images taken from a helicopter and the results of tracking and range estimation are presented. It is concluded from the results that this method of tracking the objects and estimating their range to the sensor is more reliable and robust than the existing correlation based methods.

8 Acknowledgments

We wish to thank NASA, Ames for providing the images to test the algorithms developed in this work.

References

[1] Aggarwal J. K., and Nandhakumar N., "On the computation of motion from a sequence of images - a review," *Proceeding of the IEEE* pp. 917-935, Aug 1988.

[2] Anandan P., " A review of motion and stereopsis research," *COINS Technical Report 85-52,* Dec 1985, The University of Mass. at Amherst.

[3] Dutta R. and Synder M. A., " Robustness of structure from binocular known motion," *IEEE Workshop on Visual Motion,* Princeton, NJ, October 1991.

[4] Merchant J., "Exact area registration of different views of a common object scene," *Optical Engineering,* May/June 1981, vol. 20, no. 3, pp. 424-436.

[5] Sridhar B. and Phatak A.V., "Simulation and Analysis of Image-based navigation System for Rotorcraft Low-Altitude Flight," *AHS National Specialist's meeting on Automation Applications of Rotorcraft,* Atlanta, GA., APRIL 4-6, 1988.

[6] Sridhar B., Cheng V. and Phatak A., "Kalman filter based range estimation for autonomous navigation using imaging sensors," *Automatic Control in Aerospace, IFAC Symposium,* Tsukuba, Japan, July 1989.

[7] Ullman S., "The interpretation of visual motion," Cambridge, MA: M.I.T. Press, 1979.