# Cooperative Infrared and Visible Band Tracking

V. Deodeshmukh[1] S. Chaudhuri[2] S. Dutta Roy[2]
[1]BJM School of Bioscience & Bioengineering [2]Department of Electrical Engineering
Indian Institute of Technology Bombay, Mumbai 400076
{vivek@cc, sc@ee, sumantra@ee}.iitb.ac.in

## Abstract

*Trackers based on cameras operating in the visible band do not work well in low lighting conditions. Infrared (IR) cameras typically have a low frame rate, hence making tracking in the IR band difficult. This paper presents a novel approach for cooperative tracking between two cameras operating in the IR and visible bands. We represent a framework for fusing results of two such kalman trackers, using an estimated geometric relationship between the two cameras.*

## 1 Introduction

Tracking a human being in low light conditions using a visible band camera is a difficult task due to low image contrast. An IR camera captures IR radiation emitted by the human body. The human body temperature of $37°C$ is usually above the environmental temperature - techniques used in an IR camera to suppress the background [10] are sufficient to get a good contrast in an IR image. Since the image obtained from an IR camera is independent of lighting conditions, an IR tracker is used to track moving objects in poor lighting conditions. However, if a person is moving rapidly, an IR tracker fails because it has a lower frame rate as compared to a visible band camera. In this paper, we propose to use both IR and visible band cameras to track a moving person and fuse the results obtained from both trackers. The fusion techniques give good results when either tracker fails.

Pfinder [15] is a real-time system for tracking a person which uses a multi-class statistical model of color and shape to segment a person from a background scene. Haritaoglu *et al.* [8] model background scene pixels by minimum, maximum and maximum difference, and then use thresholding and statistical procedures to segment the object being tracked. Mammen *et al.* [12] use $Cb$ and $Cr$ values of skin colour to track a moving hand by using an algorithm which is trained for skin color values which can track only specified color values. Gupta *et al.* extend the concept of EigenTracking to include a predictive component, for faster and more reliable tracking [7]. They also propose a framework for cooperative tracking - enhancing any tracker with additional shape information. This is however, only for restricted affine motion. Even with a predictive framework however, EigenTracking is inherently slow due to the iterative non-linear framework. In all prediction-based trackers, the measurement is made with respect to the presence of the object of interest within the predicted region. Whether this is skin colour-based for a human hand for example, or motion-based - such techniques do not perform robustly under low lighting conditions. Noise further affects this process adversely. Therefore, a tracker operating in the visible band of electromagnetic radiation fails. In such cases, inputs from a tracker operating in another region of the spectrum (such as IR) may prove useful. Consider a case of a rapidly moving person. Most IR cameras have a low frame rate as compared to visible band cameras. The large displacement of the person between two successive frame leads to the failure of an IR tracker. In this paper, we propose a symbiotic framework for fusing information from two such trackers, for conditions adverse to any one particular tracker. We propose a Kalman Filter-based technique, along with state vector fusion and measurement fusion, for robust tracking.

## 2 Cooperative Tracking

In this section, we discuss models used to track both IR and visible band images. We further discuss how to map the motion window from IR image to visible band image and vice-versa, and fusion of results obtained from IR and visible band trackers.

### 2.1 Tracker Framework

We use a Kalman filter-based scheme is used for both visible band and IR images. The Kalman filter is based on a representation of the system using the *state-space approach*, in which a dynamical system is described by a set of variables called the state [9]. The system is described in terms of the following two equations where $\mathbf{x}(n)$ is the state at time $n$ and $\mathbf{y}(n)$is the measurement or observation.The *process*

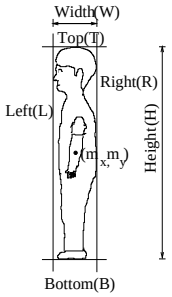



Figure 1: Illustration of choice of states

*equation* and *measurement equation* are given as:

$$\mathbf{x}(n+1) = \Phi(n+1,n)\mathbf{x}(n) + \mathbf{v_1}(n) \quad (1)$$

$$\mathbf{y}(n) = C(n)\mathbf{x}(n) + \mathbf{v_2}(n) \quad (2)$$

The vectors $\mathbf{v_1}(n)$ representing the *process noise* and $\mathbf{v_2}(n)$ representing the *measurement noise* are modeled as zero-mean, white-noise processes whose correlation matrices are $\mathbf{Q_1}$ and $\mathbf{Q_2}$ respectively. $\mathbf{y}(n)$ represents the measurement. Here the process equation (Equation 1) describes the state dynamics *i.e.*, how the states change with time. The measurement equation (Equation 2) shows the relationship between the observation *i.e.*, the measured value and the state. The state transition matrix $\Phi(n+1,n)$ and the measurement matrix $C(n)$ are both assumed to be known along with $\mathbf{Q}_1(n)$ and $\mathbf{Q}_2(n)$. The filtering problem is to find the minimum mean-square estimates of the components of the state $\mathbf{x}(n)$ by using the observed data.

Any tracker needs three states - a state vector, a state dynamics, and a measurement vector. We select the center coordinates $(m_x, m_y)$ of the rectangular window bounding the person, and its height $(H)$ and width $(W)$, as elements of the state vector $\mathbf{x(n)}$ (Figure 1). Thus the state changes allow the window to move, expand or shrink. The elements of the measurement vector $\mathbf{y(n)}$ are the top, bottom, left and right edges of the motion window. For such a situation, we observe that the the noise affecting $m_x$ and $W$ is independent of noise affecting $m_y$ and $H$. The first pair depends on the vertical measurements, whereas the second depends on the horizontal measurements. Thus, instead of a single Kalman filter in a 4-dimensional state space, we implement the motion tracker as two Kalman filters in 2-dimensional state space. The states for the two Kalman filters are $\mathbf{x_1}(n) = [m_x \quad W]^T$ and $\mathbf{x_2}(n) = [m_y \quad H]^T$ and their corresponding measurements are $\mathbf{y_1}(n) = [T \quad B]^T$ and $\mathbf{y_2}(n) = [L \quad R]^T$ respectively.

For the measurement, we have a very simple technique to detect motion. To avoid spurious motion detection due to noise, we normalize and threshold the absolute difference image. The threshold is based on the distribution of grey levels in the difference image. The minimum bounding rectangle, surrounding the all the pixels above threshold, is used for tracking. For more accurate motion estimation, one may use dominant motion extraction [11], for instance. For the state dynamics, we have experimented with two models: constant position, and constant velocity.

### 2.1.1 Constant Position Model

In this case, one may think of the states as undergoing a transition due to the effect of white noise. This implies that the change in state from one frame to another is uncorrelated. The motivation behind such a model is a common observation that states do not change very rapidly across consecutive frames and hence the change can be modeled by noise. In our case we have a time-invariant state transition matrix $\mathbf{\Phi}(n) = \mathbf{I}$ and measurement matrix $C(n) = C = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & -1 \end{bmatrix}$ respectively. Thus, we may rewrite the process equation (Equation 1) and the measurement equation (Equation 2) as follows.

$$\mathbf{x}(n+1) = \mathbf{x}(n) + \mathbf{v_1}(n) \quad (3)$$

$$\mathbf{y}(n) = C\mathbf{x}(n) + \mathbf{v_2}(n) \quad (4)$$

### 2.1.2 Constant Velocity Model

The *process equation* and the *measurement equation* of the constant velocity model is given as [6], [2]

$$\mathbf{x}(n+1) = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \mathbf{x}(n) + \begin{bmatrix} \frac{1}{2} \\ 1 \end{bmatrix} \mathbf{v_1}(n) \quad (5)$$

$$\mathbf{y}(n) = \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{x}(n) + \mathbf{v_2}(n) \quad (6)$$

Where $T$ is sampling time. In this case too, we have a vertical tracker, and a horizontal tracker. For our experimental setup, the person's height does not change abruptly and the distance between camera and object is large. We use a constant position model for the vertical tracker (Section 2.1.1), while in horizontal tracker the horizontal lines of minimum bounding rectangle left and right ($Y$ Coordinates, Figure. 1) are tracked independently. The state of the horizontal model is $\mathbf{x}(n) = [Position \ velocity]^T$ ($Position$ is either the left or right line of the minimum bounding rectangle)

## 2.2 Fusion and Updating

Figure 2 shows a block diagram of the proposed cooperative tracking mechanism. A tracker (IR, or visual band) fails when its entries of the covariance matrices exceeds the pre-determined thresholds. The IR and visible band camera look at approximately the same part of the scene. For such a setup, we show in [5] that image points in the two images are related by a 2-D *affine* transformation:

$$\begin{bmatrix} v_x \\ v_y \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} i_x \\ i_x \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (7)$$

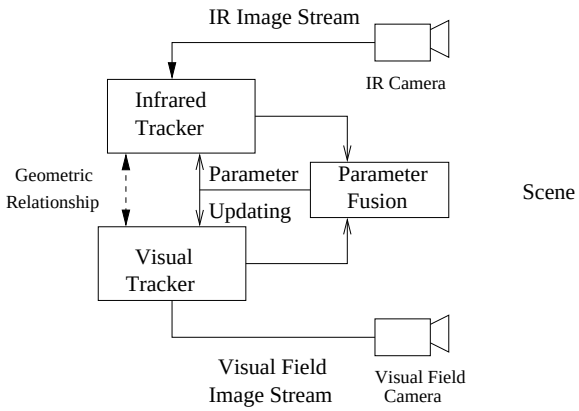



Figure 2: Cooperative IR and Visible Band tracking: Flow diagram

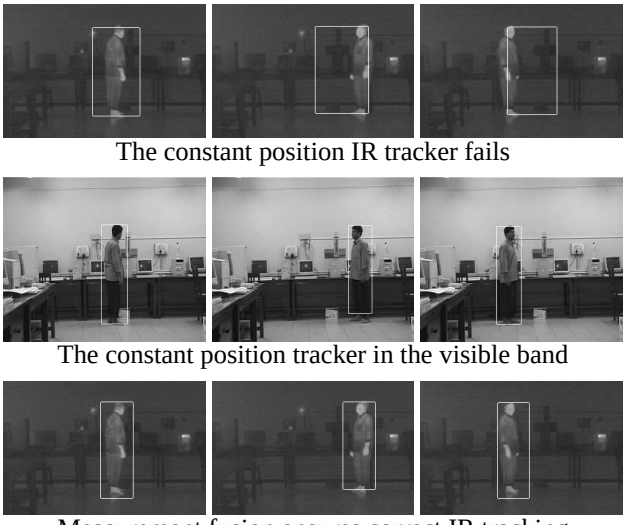



Figure 3: Constant Position Model: The effect of measurement fusion

where $[v_x\ v_y]^T$ and $[i_x\ i_y]^T$ are corresponding points in the visible band and IR images, respectively, and $a$, $b$, $c$, $d$, $t_x$ and $t_y$ are the 6 affine parameters relating the two. We estimate the 6 affine parameters from point correspondences in a set of training images. The system has the IR and visible band cameras temporally synchronized with each other.

The cooperation between the trackers can be through either measurement fusion, or state vector fusion. We assume the measurement noise to be independent, for the two cameras. In measurement fusion [14], [13], the measurement vectors $\mathbf{y_i}(n)$ and $\mathbf{y_v}(n)$ are fused to obtain the minimum square estimate $\bar{\mathbf{y}}$.

$$\bar{\mathbf{y}} = \mathbf{y_i} + \mathbf{R_i}(\mathbf{R_i} + \mathbf{R_v})^{-1}(\mathbf{y_v} - \mathbf{y_i}) \quad (8)$$

$$\bar{\mathbf{R}} = [(\mathbf{R_i})^{-1} + (\mathbf{R_v})^{-1}]^{-1} \quad (9)$$

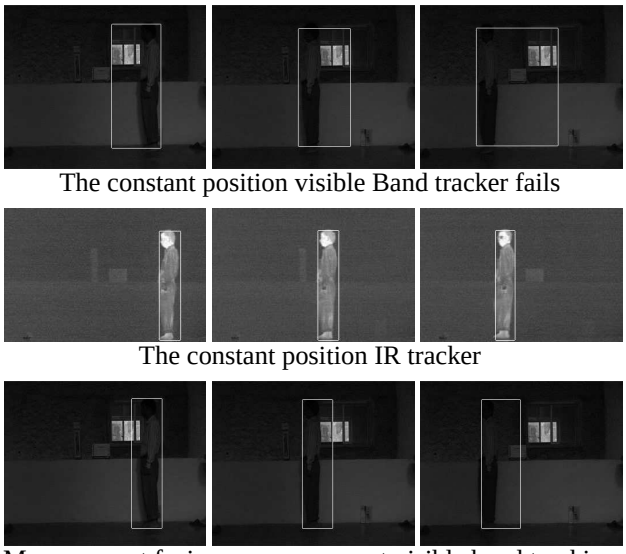



Figure 4: Constant position model: The effect of measurement fusion

Here, $\bar{\mathbf{R}}$ is the covariance matrix of the fused measurement vector $\bar{\mathbf{y}}$. $\mathbf{R_v}$ and $\mathbf{R_i}$ are covariance matrices of the measurement vectors $\mathbf{y_i}$ and $\mathbf{y_v}$, respectively.

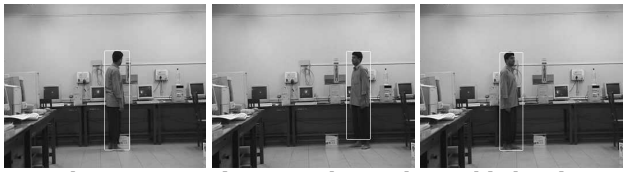
In our experimental setup, the frame rate of the visible band camera is approximately 8 times than that of the IR camera. First, we show results with using a constant position model (Section 2.1.1). In Figure 3, the IR tracker fails due to rapid displacement of the moving object, between any two consecutive frames. measurement fusion using results from the visible tracker enables correct tracking, as shown. In Figure 4, the visible band tracker fails due to poor lighting. Measurment fusion using results from the IR tracker mitigates this problem.

Let $\hat{\mathbf{x}}_\mathbf{i}(n/n)$ and $\hat{\mathbf{x}}_\mathbf{v}(n/n)$ denote the state estimates of the IR and visible band trackers, respectively. state vector fusion [6], [2], [3], performs a Maximum likelihood fusion in the following manner ([6], [4], [1]):

$$\bar{\mathbf{x}} = \hat{\mathbf{x}}_\mathbf{i} + (\mathbf{P_i} - \mathbf{P_{iv}})(\mathbf{P_i} + \mathbf{P_v} - \mathbf{P_{iv}} - \mathbf{P_{vi}})^{-1}(\hat{\mathbf{x}}_\mathbf{v} - \hat{\mathbf{x}}_\mathbf{i})$$
$$\bar{\mathbf{P}} = \mathbf{P_i} - (\mathbf{P_i} - \mathbf{P_{iv}})(\mathbf{P_i} + \mathbf{P_v} - \mathbf{P_{iv}} - \mathbf{P_{vi}})^{-1}(\mathbf{P_i} - \mathbf{P_{vi}})$$

Here, $\bar{\mathbf{x}}$ and $\bar{\mathbf{P}}$ are the fused estimates and the corresponding covariance matrices; and $\mathbf{P_i}$ and $\mathbf{P_v}$ are the covariance matrices of $\hat{\mathbf{x}}_\mathbf{i}(n/n)$ and $\hat{\mathbf{x}}_\mathbf{v}(n/n)$ respectively. In both measurement and state vector fusion, the fusion and updating is done when the corresponding IR and visible band image pairs are available. Otherwise, the visible band tracker continues to track independently. In Figures 5 and 6, we use the constant velocity model, and additionally show results using state vector fusion. In both cases, the fusion enables more reliable tracking of the moving object, across frames.

The constant velocity IR tracker fails



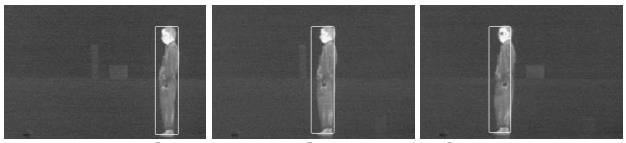
The constant velocity tracker in the visible band

State vector fusion: successful tracking

Measurement fusion: successful tracking

Figure 5: Constant velocity model: the effect of state and measurement fusion

The constant velocity visible band tracker fails



The constant velocity IR tracker

State vector fusion: successful tracking

Measurement fusion: successful tracking

Figure 6: Constant velocity model: effect of state and measurement vector fusion

# 3 Conclusion

This paper presents a symbiotic framework for combining results of two trackers using state vector and measurement fusion; Even when one tracker fails in adverse conditions, the symbiotic information fusion enables robust tracking as shown by our experimental results.

# References


[1] Y. Bar-Shalom. On the Track-to-Track Correlation Problem. *IEEE Transactions on Automatic Control*, AC-26(2), April 1981.

[2] Y. Bar-Shalom and L. Campo. The Effect of the Common Process Noise on the Two-Sensor Fused-Track Covariance. *IEEE Transactions on Aerospace and Electronic Systems*, AES-22(6):803 – 805, 1986.

[3] Y. Bar-Shalom and X. Li. *Estimation and Tracking: Principles Techniques and Software*. Dedham MA Artech House, 1993.

[4] Y. Bar-Shalom and X. Li. *Multitarget-Multisensor Tracking:Principles and Techniques*. Storrs, CT: YBS Publishing, 1995.

[5] U. Bhosle, S. Chaudhuri, and S. Dutta Roy. Multispectral Mosaicing. In *International Conference on Advances in Pattern Recognition*, 2003. (Submitted for review).

[6] K. Chang, R. Saha, and Y. Bar-Shalom. On Optimal Track-to-Track Fusion. *IEEE Transactions on Aerospace and Electronic System*, 33(4):1271 – 1275, 1997.

[7] N. Gupta, P. Mittal, S. Dutta Roy, S. Chaudhury, and S. Banerjee. A Predictive scheme for Appearence-Based Hand Tracking. In *Proc. Indian Conference on Computer Vision, Graphics and Image Processing*, pages 49 – 54, 2002.

[8] I. Haritaoglu, D. Harwood, and L. Davis. A Real Time System for Detecting and Tracking People. *Third International Conference on Automatic Face and Gesture*, April 1998.

[9] S. Haykin. *Adaptive Filter Theory*. Englewood Cliffs, 1986.

[10] R. D. Hudson. *Infrared System Engineering*, pages 235 – 263. John Wiley & Sons, Inc, 1969.

[11] M. Irani, B. Rousso, and S. Peleg. Computing Occluding and Transparent Motions. *International Journal of Computer Vision*, 12(1):5 – 16, January 1994.

[12] J. Mammen, S. Chaudhuri, and T. Agarwal. Simultaneous Tracking of Both Hands by Estimation of Erroneous Observations. In *Proc. British Machine Vision Conference (BMVC)*, 2001.

[13] J. A. Roecker and C. McGillem. Correspondence. *IEEE Transactions on Aerospace and Electronic System*, 24(4):447 – 449, 1988.

[14] D. Willner, C. B. Chang, and K. Dunn. Kalman Filter Algorithm for Multi-Sensor system. *IEEE Conference on Decision and Control*, pages 570 – 574, December 1976.

[15] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-Time Tracking of the Human Body. In *SPIE Conference on Integration Issues in Large Commercial Media Delivery System*, October 1995.