Vehicle Tracking Through the Exploitation of Remote Sensing and LWIR Polarization Science

Hamilton Scott Clouse[a] and Hamid Krim[a] and Wesam Sakla[b] and Olga Mendoza-Schrock[b]

[a]North Carolina State University, Campus Box 7911, Raleigh, North Carolina;
[b]Air Force Research Laboratory, 2241 Avionics Circle, Wright-Patterson AFB, Ohio

ABSTRACT

Vehicle tracking is an integral component in layered sensing exploitation applications. The utilization of a combination of sensing modalities and processing techniques provides better insight about a situation than can be achieved with a single sensing modality. In this work, several robust features are explored for vehicle tracking using data captured in a remote sensing setting. Particularly, a video sequence is evaluated in the layered sensing framework. The target area is surveyed by two sensors operating across two different modalities: a LWIR polarized sensor and a LWIR sensor. We here extend our previous work ([1]) to experimental analysis on several feature sets including three classic features (Stokes images, DoLP, the Degree of Linear Polarization, and AoP, the Angle of Polarization) and two experimental features (ρ, an extension of the DoLP, and δ, the phase difference between two orthogonal projections of an observed electromagnetic wave).[1]

Keywords: layered sensing, distributed sensing, polarimetric, infrared, tracking, feature-aided, fusion, multi-sensor

1. INTRODUCTION

Polarimetry information allows a more complete description of observed waveforms. In the thermal or long-wave infrared (LWIR) the polarization information makes an already robust observation mode even more so. LWIR sensors, while dependent upon material properties of the objects in the observed scene, are independent of the scene illumination. The geometry afforded by the polarimetry gains a further removal from the confusers in a scene by generating intensity invariant feature sets.

These features are inter-related and therefore can be represented as functions of each other. Feature selection is a process by which the most useful, in some sense, can be singled out and utilized alone. The process followed in this work was:

1. Feature selection

2. Fusion

3. Tracking

Using polarimetric infrared data as an example, feature selection was used to generate feature sets from the geometric properties of the polarization ellipse. The generated feature sets were fused with two different techniques. Tracking was performed with the input being both the fused feature sets and the features independently. For the tracks resulting from independent elements of the feature sets, the tracks were fused via the method described in [1] and reiterated in section 4.3.2. The automatically generated tracks were compared to truth data generated by human-in-the-loop techniques for accuracy evaluation.

In Section 2 we discuss, in general, the polarization phenomena of electromagnetic waves and the associated mathematical representations as well as the methodologies of feature selection, data fusion and tracking via Kalman filters.

We continue with the details of our particular implementations in Section 4. The organization of the paper is as follows: Section 2-Background, Section 3-Data Description, Section 4-Methods, Section 5-Results, and Section 6-Conclusions.

2. BACKGROUND

2.1 Polarization of EM Radiation

The sensing modality utilized in this work, polarimetric infrared (PolarIR), is the same as in our earlier effort.[1] Hence, the physical nature of the phenomena observed is the same. More detailed information can be found in [2], [3], and [4].

Essentially, this effort focused on the geometric diversity afforded by the addition of the polarimetry to the long wave infrared (LWIR) video collection. As such, a sketch of the mathematical formalism employed follows.
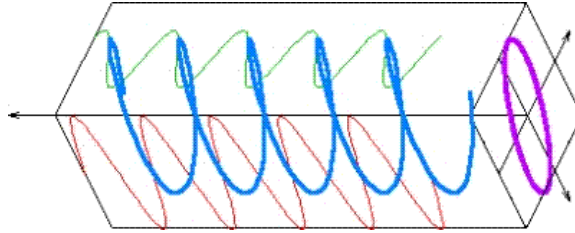

Figure 1. In general, the E-field component of an EM wave rotates about the transverse axis and thus traces an ellipse in the plane of rotation and can be decomposed into orthogonal components E_{0_X} and E_{0_Y}.

The electric wave, as it propagates will, in general, have some rotation about the z-axis and will thus trace out an ellipse in the plane of rotation, as shown in Fig.1. The ellipse that is traced out may be described by way of of the electric wave properties as indicated in Fig.2.
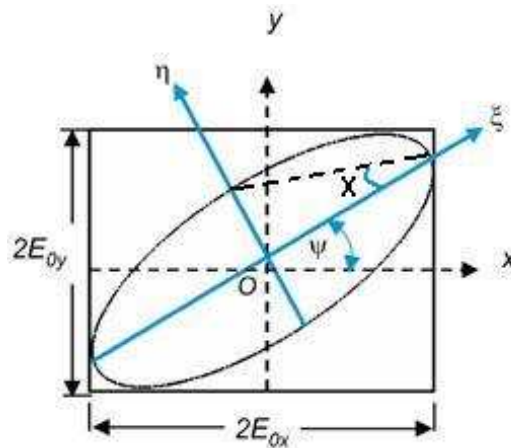

Figure 2. The polarization ellipse with the labelled parameters. From [5]

The instrument by which the data used in the work was captured produces video sequences of the Stokes[6] vector of a scene. The video is in three channels:

- S_0 - intensity

- S_1 - horizontal preference

- S_2 - $+45^o$ preference

Further details about the post-processing that produced such data sequences can be found in [7].

The geometric relations implied by Figure 2 are further describe by the relations between the components of the Stokes vector in Eqn. 1. In both representations ψ is the angle of polarization (AoP), χ is the angle of ellipticity, a is the radiant intensity, E_{0_X} and E_{0_Y} are the amplitudes of the orthogonal components of the observed E-field wave and δ is the phase difference between them.

$$S = \begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} = \begin{bmatrix} a^2 \\ a^2 cos(2\chi)cos(2\psi) \\ a^2 cos(2\chi)sin(2\psi) \\ a^2 sin(2\chi) \end{bmatrix} = \begin{bmatrix} E_{0_X}^2 + E_{0_Y}^2 \\ E_{0_X}^2 - E_{0_Y}^2 \\ 2E_{0_X}E_{0_Y}cos(\delta) \\ 2E_{0_X}E_{0_Y}sin(\delta) \end{bmatrix} \tag{1}$$

Historically, two variables/features have been used to characterize and explore the polarization phenomenon. One, the angle of polarization, ψ (Eqn. 2), is directly observed on the polarization ellipse. The other, the degree of linear polarization (DoLP) is a derived feature (Eqn. 3). It is simply the ratio of the lengths of the minor and major axes of the polarization ellipse: $|cos(2\chi)|$ where $|\cdot|$, represents the absolute value operator. The advantage of these two features is that they are invariant to the illumination a of the observed scene. Two other commonly used features are the normalized versions of S_1 and S_2, $S1p = \frac{S_1}{S_0}$ and $S2p = \frac{S_2}{S_0}$, respectively. In this work additional geometric features were considered, as listed in Table 1.

$$AoP = \psi = \frac{1}{2}tan^{-1}\left(\frac{S_2}{S_1}\right) \tag{2}$$

$$DoLP = \frac{\sqrt{S_1^2 + S_2^2}}{S_0} = |cos(2\chi)| \tag{3}$$

Variable/Feature	Description	Function
ρ	$cos(2\chi)$, DoLP with sign information	$\frac{S_2}{S_0 sin(2\psi)}$
δ	phase difference between orthogonal projections E_X and E_Y	$\delta_x - \delta_y = cos^{-1}\left(\frac{S_2}{\sqrt{S_0 - S_1}\sqrt{S_0 + S_1}}\right)$
I_{max}	length of major axis of the polarization ellipse	$\frac{1}{2}\left(S_0 + \sqrt{S_1^2 + S_2^2}\right)$
I_{min}	length of minor axis of the polarization ellipse	$\frac{1}{2}\left(S_0 - \sqrt{S_1^2 + S_2^2}\right)$
R	area of polarization ellipse	$I_{min}I_{max}\frac{\pi}{4}$

Table 1. The additional geometric features considered in this work.

2.2 Feature Selection

Large numbers of variables, potential features, are inherent to some data types. This can prove cumbersome and even crippling when computations must be performed for higher-level tasks such as classification and tracking. Feature selection is a key step in mitigating the downfalls of such a problem. As is the case with the dataset utilized for this work, feature selection can also be used to prevent redundancy, both in information and in noise. It is not always true that utilizing all of the variables available will improve the produced results. Indeed noise in the variables may interfere constructively, just as the signals could. To select the correct feature subset is key in these instances.

An apt summary of such techniques is provided in [8]. Feature selection methods can be classified as either filters or wrappers. Filters are methods that select feature subsets independent of any classification algorithms or information. These methods are fast and provide the most unbiased selection for any given dataset. Examples

of such methods include correlation measures. One could choose the least correlated features to maximize the diversity maintained in the feature subset. However, if classification was the purpose for the selected feature set these methods would not necessarily be best.

Wrapper methods iteratively add or remove features from the subset, calculating for each transition how the classification results or the target optimization function would be affected. Entropy is a good example. To include as much information as possible in the subset, while maintaining as few variables as possible, one could attempt to optimize by increasing the entropy of the feature subset at each addition/subtraction of a variable while at the same time minimizing the number of variables in the feature set.

The majority of the methods utilized in this are filters, except for the Minimum Redundancy, Maximum Relevancy (mRMR) method.[9] Since this effort was not specifically focused on classification of the observed phenomena but, rather in completely describing it, it followed that filter methods were more suitable.

2.3 Fusion

As in previous work[1] fusion techniques were used in this effort to combine features and decisions to, hopefully, formulate a more robust feature/system response. In [10] fusion is described as being viable at three different levels in sensor interaction:

1. data-level where signal information is combined

2. feature-level after the data has been processed and signatures created

3. decision-level where algorithms utilize features to produce some results and those results are combined.

In this work fusion was performed at both the feature-level and decision-level via a toy tracking example. The advantages of utilizing fusion techniques include increased robustness of systems and possible generation of new knowledge.[11] More discussion can be found in resources such as [12] and [13]. Details about the particular methods used in this effort are in Sect. 4.3.

2.4 Tracking with Kalman Filters

The tracking portion of this work relied on a standard technique, the use of a Kalman filter. A detailed description of Kalman Filtering can be found in [14]. However, we also provide a short explaination.

The basic idea of the Kalman filter as a tracker is to determine the current location of an object via the current and prior measurements. The Kalman filter does this recursively and assumes that the system and measurement noise are uncorrelated. In this work we employ the discrete linear Kalman Filter which assumes the current state to be a linear function of the prior state as well as a linear measurement model. This is illustrated in Equation 4 and Equation 5 for the system and measurement models, respectively.

$$x_k = A_k x_{k-1} + w_k \tag{4}$$

$$z_k = H_k x_k + v_k \tag{5}$$

Where, A is the state-transition model, H is the observation model, x_k represents the state at time k, z_k represents the measurement vector at time k, and w and v are the system and measurement noise, respectively, and are normally distributed about zero with covariance Q_k and R_k.

The Kalman filter is implemented in two phases: predict and update. The predict phase is governed by the system estimate (Eqn. 4) and the covariance of that estimate:

$$P_k = A_k P_{k-1} A_k^T + Q_k \tag{6}$$

The update phase involves other system information, an observation as well as estimate calculations. An observation z_k is made and the residual of that measurement, $\tilde{y}$, is calculated as the deviation of the estimate from the true observation:

$$\tilde{y}_k = z_k - H_k \hat{x}_k \tag{7}$$

The optimal Kalman gain K_k needed to utilize $\tilde{y}$ in estimating $\hat{x}_{k+1}$ is calculated as shown in Eqn. 8.

$$K_k = P_k H_k (H_k P_k H_k^T + R_k)^{-1} \tag{8}$$

These values are used to update the state estimate $\hat{x}_{k+1}$:

$$\hat{x}_{k+1} = \hat{x}_k + K_k \tilde{y}_k \tag{9}$$

.

To implement the Kalman Filter, one must:

1. Initialize

   (a) Initialize the state estimate x_0. This should be an accurate estimate of the initial state, before the tracking process begins, of the entity to be tracked.

   (b) Calculate the initial state estimate covariance P_0. This is a measure of the accuracy of the initial state estimate x_0.

   (c) Define the system model A. Typically, for linear kinematic systems, this is a matrix of coefficients defining the relationship between elements of the state vector and the physical constraints of the system.

   (d) Define the observation model H that relates the measured values to the dynamics of the system.

   (e) Initialize the system and observation noise covariances: Q_k and R_k.

   (f) Make the first observation.

2. Iterate between:

   - Predict Phase (no observation available):
     (a) Calculate the state estimate x_{k+1}. (Eqn. 4)
     (b) Calculate the state estimate covariance P_{k+1}. (Eqn. 6)
   - Update Phase (observation available)
     (a) Record an observation z_k.
     (b) Calculate the measurement residual $\tilde{y}_k$. (Eqn. 7)
     (c) Compute the new gain value K_k. (Eqn. 8)
     (d) Update the state estimate x_{k+1}. (Eqn. 9)
     (e) Update the estimate covariance P_{k+1}. (Eqn. 6)

3. DATA DESCRIPTION

As described briefly in Sect. 2.1, the data utilized in this work is the result of a post-processing method applied to output from a cooled micro-polarimeter array. The resulting video sequence has three channels, one for each Stokes component. Each frame is 471×641 pixels in size and the video is captured at twenty-four frames per second. The target wavelength of the sensor is the LWIR spectrum, i.e. the thermal spectrum. The different sub-bands of the Infrared spectrum band are described in Table 2. The definition of these sub-bands is described more completely in [15].

The observed scene is composed of a series of intersecting streets with one intersection as the focus. In the scene there are several parked automobiles, trees, and street lamps. In our area of focus only one moving object is captured: an automobile traveling along one of the roads. An example image from the data set, with the moving automobile visible, is shown in Fig.4.

Sub-bands of the Infrared (IR) Spectrum	
Band Label	**Wavelength (λ) Range**
Near infrared (NIR)	$0.7 - 1\mu m$
Short-wave infrared (SWIR)	$1 - 3\mu m$
Mid-wave infrared (MWIR)	$3 - 5\mu m$
Long-wave infrared (LWIR)	$8 - 14\mu m$

Table 2. Position of the sub-bands within the IR spectrum band.
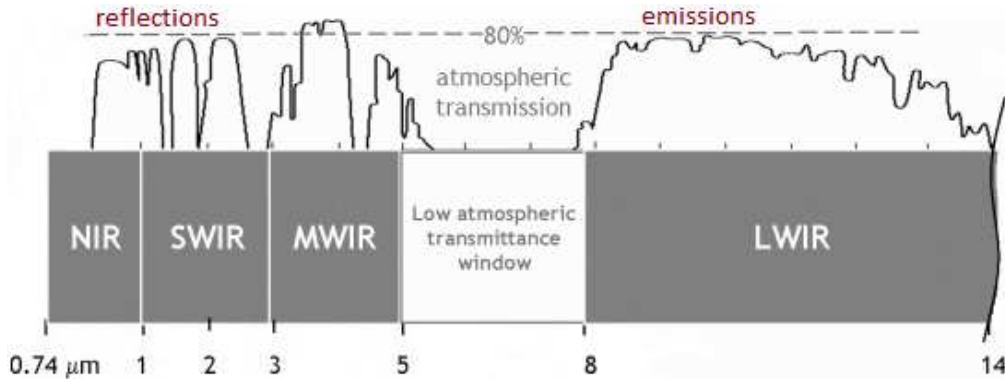

Figure 3. The Infrared band of the EM spectrum with radiance curves for reflections and thermal emissions. Clearly emissions dominate the spectrum in the LWIR. From `http://mivim.gel.ulaval.ca` accessed August 30, 2010.

Figure 4. A cleaned sample image from the scenario observed for this work. Notice the parked cars along the street and the truck in the foreground travelling along the road. This truck was the focus of our tracking effort.

4. METHODS

4.1 Kalman Filter Implementation

In this work, the centroid of the truck moving through the scene was the point being tracked. The Kalman filter was implemented as in [1]. It was initialized with an estimate of the centroid location at the center of the frame and having no velocity.

4.1.1 Object Detection

The detection scheme employed in this work was simple. The signal, or calculated values of each variable/feature was formulated as an image sequences I (Eqn. 10), a discrete function of three variables: two spatial (i, j) and

one temporal (t). This formal representation facilitated simple change detection and was thus differentiated with respect to time (Eqn 11).

$$I = f(i, j, t) \tag{10}$$

$$I' = \Delta_t f(i, j, t) = f(i, j, t + \Delta t) - f(i, j, t) \tag{11}$$

where Δt is the time between subsequent frames. In the resulting image sequence I' a value of zero indicated that no change occurred in the scene between the capture of frames at time t and time $t + \Delta t$. Since our sensor was stationary, as was the majority of the scene, and the spectrum of interest (LWIR) reflects the relatively slow rate of thermal changes it was assumed that large values in I' were due to changes in the variable I caused by motion in the scene. Thus, only the magnitude, $|I'|$, was necessary for detecting objects.

To maintain consistency regardless of the input variable a normalization step was included, described by Eqn. 12. Hard thresholding, coupled with the knowledge that the scene contained one mover, provided the segmentation of the objects from the scene.

$$\|I'(i, j, t)\| = \frac{|I'(i, j, t)| - arg_{(i,j)}min\{|I'(i, j, t)|\}}{arg_{(i,j)}max\{|I'(i, j, t)|\} - arg_{(i,j)}min\{|I'(i, j, t)|\}} \tag{12}$$

4.1.2 Tracker Performance

Manual truthing supplied the basis by which the accuracy of the automatically produced tracks were compared. An individual viewed each frame independently and, with the aid of a set of simple morphological operations, indicated the spatial location of the centroid of the mover for the entire video sequence. As in [1], this method of "truthing" provides benefits such as being independent of tracking algorithms entirely and representing the "true" track as it would be perceived by a human. The covariance between the observations along the two orthogonal spatial orientations was computed and this was used as the initialization for the observation covariance matrix in the Kalman filter.

To compare the tracks produced by the Kalman filter to the truthed tracks the average root mean square deviation (RMSD) (Eqn. 13) for each tracker was calculated over the whole scene. This measure is what was ultimately used to evaluate the trackers' performance.

$$Err = E\left[\sqrt{\frac{(\hat{\theta}_i - \theta_i)^2 + (\hat{\theta}_j - \theta_j)^2}{2}}\right] \tag{13}$$

In Eqn. 13, $\hat{\theta}$ is the vector of estimated centroid locations from the tracker, θ is the vector of the true centroid locations, $E[\cdot]$ is the expectation operator, and c and r represent the column and row coordinates (in the image plane) of the centroid, respectively.

4.1.3 Tracking in the Image Plane

A simple Kalman filter tracker, as described in Section 2.4, was used to track the moving object in the image plane. A 4×1 vector $x = \begin{bmatrix} x_i & x_j & x'_i & x'_j \end{bmatrix}^T$ was used to describe the location (x_i, x_j) and velocity (x'_i, x'_j) of the moving object in the two orthogonal spatial orientations i and j. The kinematic system model described by Eqn. 4 was used with the following form, for each value of k:

$$x_k = A_k x_{k-1} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} x_{k-1} \tag{14}$$

$$H_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{15}$$

The measurement vector, z_k, for each update was taken to be the centroid of the largest moving object located within the area predicted by the filter. Since the initial position of the mover in the scene was known the a priori state estimate covariance matrix P was set to identity, i.e. $P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

4.2 Feature Selection

4.2.1 Statistical Methods

Several data-driven techniques were employed to determine whether or not there existed direct relationships between the various features/variables. These methods are based on observing trends in the features and comparing them.

- Covariance - a measure of similarity between changes in variables. For two features $I_1(i, j, t)$ and $I_2(i, j, t)$ the covariance was computed by representing each feature as a $n \times 1$ vector and performing the following calculation:

$$Cov(I_1, I_2) = \mathrm{E}\left[(I_1 - \mathrm{E}[I_1])(I_2 - \mathrm{E}[I_2])^T\right] \tag{16}$$

  This provided a measure between zero and one of similarity in change between the two features I_1 and I_2.

- Pearson Correlation - a measure of the strength of a linear relationship between two variables. This measure is computed by simply dividing the covariance of two variables by their independent variance measures:

$$\rho_{I_1, I_2} = \frac{Cov(I_1, I_2)}{var(I_1) \cdot var(I_2)} \tag{17}$$

- Rank Correlations - a measure of strength of relationship between two variables' changes (need not be linear).

  - Kendall Correlation - once the features are ordered independently they are paired and the Kendall correlation is a measure of how the ordering of the paired feature samples coincides.

$$\tau_{I_1, I_2} = \frac{(\# \text{ of concordant pairs})(\# \text{ of discordant pairs})}{\frac{1}{2}n(n-1)} \tag{18}$$

  - Spearman Correlation - a measure of how well the relationship between two variables can be represented by a monotonic function.

$$\rho_{I_1, I_2} = \frac{\sum_k (I_{1_k} - E[I_1])(I_{2_k} - E[I_2])}{\sqrt{\sum_k (I_{1_k} - E[I_1])^2 \sum_k (I_{2_k} - E[I_2])^2}} \tag{19}$$

The results of these feature selection methods is summarized in Table 9.

4.2.2 Principal Feature Analysis

One of the methods used for critical variable selection was the Principal Feature Analysis (PFA) method as described in.[16] The idea behind PFA is to utilize the structure of the principal components from principal component analysis (PCA) to choose features which retain essential information. PFA seeks to eliminate interdependencies amongst selected features while minimizing the reconstruction error. There are five steps to this algorithm summarized below and in detail in:[16]

1. Compute the sample covariance matrix.

2. Utilize PCA to determine the principal components and eigenvalues of the covariance matrix

3. Choose the subspace dimension q and construct a matrix of the first q eigenvectors

4. Cluster the data into p clusters where, $p \geqslant q$, using the K-Means algorithm with the Euclidean distance as the distance measure

5. For each cluster, find the vector which is closest to the mean of the cluster. The corresponding feature is hence chosen as a principal feature.

4.2.3 Minimum Redundancy, Maximum Relevancy (mRMR)

In [9] the idea of feature selection based on some Mutual Information criteria was described and utilized in [17]. As the name implies this method focuses on two ideas: minimum redundancy and maximum relevancy. In the mutual information sense, minimum redundancy is the solution to the optimization for m features:

$$arg_{i,j} min \frac{1}{m^2} \sum_i \sum_j \Im(i;j) \qquad (20)$$

Here, i and j are two features and the operator $\Im(\cdot)$ is the mutual information between them given by:

$$\Im(x;y) = \sum_i \sum_j p(x_i, y_j) log \left(\frac{p(x_i, y_j)}{p(x_i)p(y_j)} \right) \qquad (21)$$

where $p(x,y)$ is the joint probability of x and y and $p(x)$ is the marginal probability. This optimization is a search for the two features i and j which are the most different, i.e. have the least mutual information.

The maximum relevancy optimization is similar in that it utilizes mutual information however, it also incorporates some precursors to classification in the form of supervised class samples. From every feature a set of samples from all classes present in the data is constructed. In this work the scene was roughly categorized into five classes: vegetation, metal, glass, asphalt and rubber. For each of these classes a sample was defined and every feature was calculated for those samples; this formed the vector $h = \{h_1, h_2, h_3, h_4, h_5\}$, one vector of every feature for each class. The optimization formulation for this measure is:

$$arg_i max \frac{1}{m} \sum_i \Im(h;i) \qquad (22)$$

Intuitively, this optimization results in the feature i that has the highest mutual information with the desired classification strata.

The mRMR technique is a combination of Eqn. 20 and Eqn. 22 into a single optimization producing a feature i that has both the most mutual information with the desired classification scheme and the most unique information when compare to the other features. This combination, as described in [9] can be done with either a differencing or a quotient calculation. In this effort both were computed. Tables 3, 4, 5 and 6 summarize the results of the calculations. The resulting feature pairs were the same as those indicated by the Kendall and Spearman Correlation measures.

Order	Feature	Score
All Features: MaxRel Scores		
1	R	5.680
2	I_{max}	3.870
3	I_{min}	3.525
4	ψ	2.800
5	$\frac{S_1}{S_0}$	2.715
6	S_1	2.709
7	ρ	2.592
8	S_2	2.431
9	δ	2.382
10	$\frac{S_2}{S_0}$	2.335

Table 3. Maximum relevancy scores for features based on training set. The results were the same for both Difference and Quotient methods.

All Features: mRMR Scores					
MI Method: Difference			MI Method: Quotient		
Order	Feature	Score	Order	Feature	Score
1	R	5.680	1	R	5.680
2	S_1	0.049	2	S_2	1.019
3	I_{max}	0.534	3	I_{max}	1.192
4	I_{min}	0.353	4	I_{min}	1.138
5	S_2	-0.115	5	ρ	1.004
6	DoLP	-0.197	6	S_1	0.934
7	ψ	-0.213	7	ψ	0.915
8	ρ	-0.527	8	$\frac{S_1}{S_0}$	0.804
9	$\frac{S_2}{S_0}$	-0.567	9	$\frac{S_2}{S_0}$	0.791
10	$\frac{S_1}{S_0}$	-0.568	10	DoLP	0.802
11	δ	-0.815	11	δ	0.745

Table 4. Minimum redundancy, maximum relevancy scores for features based on training set.

Illumination Invariant Features: MaxRel Scores		
Order	Feature	Score
1	ρ	4.964
2	$\frac{S_1}{S_0}$	3.273
3	ψ	2.631
4	δ	2.331

Table 5. Maximum relevancy scores for illumination invariant features based on training set. The results were the same for both Difference and Quotient methods.

Illumination Invariant Features: mRMR Scores					
MI Method: Difference			MI Method: Quotient		
Order	Feature	Score	Order	Feature	Score
1	ρ	4.964	1	ρ	4.964
2	$\frac{S_2}{S_0}$	-0.214	2	δ	0.914
3	$\frac{S_1}{S_0}$	0.108	3	$\frac{S_1}{S_0}$	1.018
4	ψ	-0.594	4	ψ	0.809
5	δ	-1.227	5	$\frac{S_2}{S_0}$	0.646

Table 6. Minimum redundancy, maximum relevancy scores for illumination invariant features based on training set.

4.3 Fusion

4.3.1 Feature-Level Fusion

Feature-level fusion implies, as described in 2.3, the combination of information from multiple measurements. Two methods were employed in this effort: simple feature-level fusion and feature-level fusion via discrete wavelet transform (DWT) and the associated inverse DWT ($IDWT$). In both methods the same two fusion rules were used:

$$min\{\cdot\} \text{ OR } max\{\cdot\} \tag{23}$$

The simple feature-level fusion consisted of defining the new feature as in Eqn. 24. The function $g\{\cdot\}$ is the decision rule used from Eqn. 23.

$$feat_{new}(i,j,t) = arg_{(i,j)}g\{feat_1(i,j,t), feat_2(i,j,t)\} \tag{24}$$

The feature-level fusion method relying on the DWT is an employment of the techniques outlined in [18]. The DWT feature-level fusion method first represents each feature as a set of coefficients for some set of basis

functions:

$$\psi_{kl}(i,j) : k, l \in \mathbb{Z} \tag{25}$$

These bases are wavelets.

The coefficients are then compared as in Eqn. 24 where $feat_1$ and $feat_2$ are wavelet coefficients of the original features instead of the feature values themselves. The wavelet set utilized for this work was the Daubechies wavelet Daubechies 4. A framework for the procedure is illustrated as follows:

$$feat_{new}(i,j,t) = IDWT \left[arg_{(i,j)} g\{DWT \left[feat_1(i,j,t)\right], DWT \left[feat_2(i,j,t)\right]\} \right] \tag{26}$$

Once the fusion of the wavelet coefficients was complete, the $IDWT$ was applied to the new fused coefficient set to produce the final fused feature for the input feature set supplied. The function g is defined in the same manner as in Eqn. 23.

The features produced from this fusion method were utilized as input to the tracker scheme, just as the individual features, for analysis and, ultimately, accuracy comparison.

4.3.2 Decision-Level Fusion

As in [1] the state estimate covariance matrix, P, was used to calculate a confidence measure (Eqn. 27) for each tracking result.

$$c = tr\left(PC\right) \tag{27}$$

In Eqn. 27, $C = (c_{ij})$, $i, j \in [1, n]$, $c_{ij} = 0$ for $i \neq j$ and $c_{ii} = 1$ for $i \in [1, k]$ where n is the number of state variables in the tracker, k is the number of state variables to consider in the confidence calculation and $tr(\cdot)$ is the matrix trace function. In our case, k=2 reflects our interest in the accuracy of the centroid position estimate.

Therefore, our decision-level fused tracker is simply a decision rule (Eqn. 28), evaluated at each node. We use $\hat{\theta}$ to represent the estimate with the subscript "0" indicating the current node and "m" indicating all of the neighboring nodes.

$$\hat{\theta}_0 = \begin{cases} \hat{\theta_m} & \text{if } c_0 > c_m \\ \hat{\theta}_0 & \text{else} \end{cases} \tag{28}$$

5. RESULTS

S_0	S_1	S_2	I_{max}	I_{min}	R
2	4	9	4	5	2

Table 7. This is the tabulated Average RMSD (in pixels) of the tracking results from true, when tracking was performed with each feature independently. This table indicates the features dependent on scene illumination.

DoLP	ψ	δ	ρ	$\frac{S_1}{S_0}$	$\frac{S_2}{S_0}$
7	17	3	4	4	17

Table 8. This is the tabulated Average RMSD (in pixels) of the tracking results from true, when tracking was performed with each feature independently. This table indicates the features which are independent of scene illumination.

Feature Set	Fusion Decision Level	Fusion Feat. Lvl. (Min.)	Fusion Feat. Lvl. (Max.)	Fusion Feat. Lvl. DWT(Min.)	Fusion Feat. Lvl. DWT(Max.)
Covariance $(\text{DoLP}+\rho)$	4	4	7	32	28
Pearson Correlation $(\psi + \frac{S_2}{S_0})$	17	20	7	30	17
Kendall Correlation and mRMR:MIQ $(\rho + \delta)$	8	4	19	9	24
Spearman Correlation and mRMR:MID $(\rho + \frac{S_2}{S_0})$	4	4	3	37	28
Principal Feature Analysis $(\psi + \rho)$	4	21	10	26	17
All	9	N/A	N/A	N/A	N/A

Table 9. This is the tabulated Average RMSD (in pixels) of the tracking results from true, when tracking was performed with feature sets and fusion techniques.

6. CONCLUSIONS

From the experiments performed in this effort it has been shown that scene illumination/intensity information greatly improves tracking accuracy when the mover is easily segmented from the background using only intensity information. This is shown by R, the area of the ellipse, a function of the intensity observed, providing the most accurate tracking result when used independently. However, because of the possible variations in illumination conditions and emission strength it is more desirable to utilize features not reliant on intensity information; of such features δ and ρ, in that order, provide the most accurate tracking result from the employed scheme.

The performance of the paired features fused at the feature level were all comparable except for the $\psi + \frac{S_2}{S_0}$ pair produced by the Pearson Correlation method. The distribution of ψ can be modeled as a random variable uniformly distributed over the appropriate range. This characteristic could have skewed the correlation and information based selection methods as it would be both highly uncorrelated from the other variables and have the highest entropy. To test this, a uniform noise signal with the same data size parameters was processing with the described tracking scheme and the resulting average RMSD was fifty pixels. This, albeit coarse, allows some intuition about the accuracy provided by the selected features, each of which is less than 35 pixels. Overall the feature-level fusion of pairs performed via the discrete wavelet transform were less accurate than direct feature-level fusion. This may indicate that the min/max rules for DWT fusion is naive for this data type. Future work will include a more precise evaluation of the selection and fusion rules to improve robustness of the output feature set. Also a spectral and stability analysis of the described scheme will be performed to determine possible methods to, again, improve the robustness of the output decision.

ACKNOWLEDGMENTS

REFERENCES

[1] Clouse, H. S., Krim, H., and Mendoza-Schrock, O., "A scaled, performance driven evaluation of the layered sensing framework utilizing polarimetric infrared imagery," in [*Proceedings of SPIE*], *Polarization Science and Remote Sensing V* (2011).

[2] Brosseau, C., [*Fundamentals of Polarized Light: A Statistical Optics Approach*], John Wiley & Sons, Inc. (1998).

[3] Goldstein, D., [*Polarized Light*], Marcel Dekker, Inc., 2nd ed. (2003).

[4] Sandus, O., "A review of emission polarization," *Applied Optics* **4**, 1634--1642 (December 1965).

[5] Collett, E., [*Field Guide to Polarization*], SPIE Press (2005).

[6] Stokes, G. G., "On the composition and resolution of streams of polarized light from different sources," *Transactions of the Cambridge Philosophical Society* **9**, 399--416 (1852).

[7] LeMaster, D. A. and Cain, S. C., "Multichannel blind deconvolution of polarimetric imagery," *Journal of Optical Society of America* **25**(9), 2170--2176 (2008).

[8] Guyon, I. and Elisseeff, A., "An introduction to variable and feature selection," *Journal of Machine Learning Research* (2003).

[9] Peng, H., Long, F., and Ding, C., "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2005).

[10] Varshney, P. K., "Multisensor data fusion," *Electronics & Communication Engineering Journal* , 245--253 (December 1997).

[11] Hall, D. L. and Llinas, J., "An introduction to multisensor data fusion," in [*Proceedings of the IEEE*], **85**, 6--23 (January 1997).

[12] Lefebvre, E., ed., [*Advances and Challenges in Multisensor Data and Information Processing*], vol. 8 of *Sub-Series D: Information and Communication Security*, IOS Press (2007).

[13] Gan, Q. and Harris, C. J., "Comparison of two measurement fusion methods for Kalman-filter-based multisensor data fusion," *IEEE Transaction on Aerospace and Electronics Systems* **37**, 273--280 (January 2001).

[14] Grewal, M. S. and Andrews, A. P., [*Kalman Filtering*], John Wiley & Sons, Inc., 3rd ed. (2008).

[15] Miller, J. L., [*Principles of Infrared Technology: A Practical Guide to the State of the Art*], Van Nostrand Reinhold (1994).

[16] Cohen, I., Xiang, Q. T., Zhou, S., Sean, X., Thomas, Z., and Huang, T. S., "Feature selection using principal feature analysis," (2002).

[17] Ding, C. and Peng, H., "Minimum redundancy feature selection from microarray gene expression data," *Journal of Bioinformatics and Computational Biology* **3**(2), 185--205 (2005).

[18] de Zeeuw, P. M., "Wavelet and image fusion," tech. rep., Centrum Wiskunde & Informatica (1998).