

META-OPTIMIZATION OF THE EXTENDED KALMAN FILTER'S PARAMETERS FOR IMPROVED FEATURE EXTRACTION ON HYPER-TEMPORAL IMAGES

^{†‡}*B.P. Salmon*, ^{†‡}*W. Kleynhans*, [‡]*F. van den Bergh*, [†]*J.C. Olivier*, ^{*}*W.J. Marais* and [‡]*K.J. Wessels*

[†]Department of Electrical, Electronic and Computer Engineering, University of Pretoria, South Africa

[‡]Remote Sensing Research Unit, Meraka Institute, CSIR, Pretoria, South Africa
bsalmon@csir.co.za

^{*}Space Science and Engineering Center, University of Wisconsin-Madison, Wisconsin, United States

ABSTRACT

Time series derived from the first two spectral bands of the MODerate-resolution Imaging Spectroradiometer (MODIS) land surface reflectance product can be modelled as a pair of triply (mean, phase and amplitude) modulated cosine functions. This paper proposes a meta-optimization approach for setting the parameters of the non-linear Extended Kalman Filter to rapidly and efficiently estimate the features for the pair of triply modulated cosine functions. The approach is based on an unsupervised search algorithm over an appropriately defined manifold using spatial and temporal information. Performance of the new method is compared to other applicable methods and is tested on the Gauteng province which is South Africa's province with the fastest growing economy.

Index Terms— Hellinger distance, Kalman Filter, Time series analysis, Unsupervised learning, Spatial information

1. INTRODUCTION

Reliable surveying of land cover and transformation has always been a key area of interest to the remote sensing community. The increase in human population is one of the major contributions to settlement expansion within a geographical area [1]. Several studies have investigated the effects that anthropogenic activities have on the environment and it is estimated that more than a third of the Earth's land surface has been transformed by human activities [2].

The Gauteng province is of interest as it is the fastest growing province in South Africa, housing more than 10.5 million people in the year 2010. Proper knowledge of land cover is a critical tool in the effective allocation and management of the environmental resources. Digital classification of land cover consist mainly of spatial and spectral analysis. Few methods exploit the temporal sampling rate which coarse resolution satellites provide that captures the dynamics of different land cover types.

The Extended Kalman Filter (EKF) has previously been shown as a feature extraction method to model a NDVI time series for a given pixel as a triply modulated cosine function to improve land cover separation [3]. The objective of this paper is to propose an extension to [3], in which each of the first two spectral bands be modelled separately as a triply modulated cosine function and thereafter describe an appropriately defined manifold using a spatio-temporal window to search for improved parameters to be used within the EKF. The method uses an unsupervised search algorithm to search for improve parameters within manifold. The performance of this new method is compared to a least squares algorithm and an unbiased EKF.

The paper is organized as follows. Section 2 discusses the study area and data set. In section 3 we present the new method for improved feature extraction and section 4 present the experimental results by comparing several approaches to the new proposed method. Section 5 presents the conclusions.

2. STUDY AREA AND DATA DESCRIPTION

The study area is the Gauteng province which is located in the Highveld of South Africa. The province is the most urbanized province in the country and contributes more than a third of South Africa's economy. Large areas of natural vegetation still exist within the province. The method was applied to a validated study area that corresponds to a total area of approximately 285.5 km². The study area's land cover is predominantly natural vegetation and human settlements. The time series in the validated study area were verified using visual interpretation of SPOT images to map areas of no change in land cover type during the study period for the temporal component of the analysis. The proposed method was then applied to the entire Gauteng province, covering an area of 19676 km².

This research was funded by the CSIR Strategic Research Panel.

2.1. MODIS time series data

MODIS spectral bands 1 and 2 time series data were extracted from the 8-day composite MCD43A4 bidirectional reflectance distribution function (BRDF) adjusted MODIS land surface reflectance product, with a spatial resolution of 500 m for the time period January 2001 to January 2009 [4].

3. METHODOLOGY

The EKF has been used as a feature extraction method to model a NDVI time series for a given pixel as a triply modulated cosine function to improve land cover separation [3]. This paper proposes an extension to [3], that each of the first two spectral bands be modelled separately as a triply modulated cosine function and is expressed as

$$y_{i,k,b} = \mu_{i,k,b} + \alpha_{i,k,b} \cos(\omega k + \phi_{i,k,b}) + v_{i,k,b}, \quad (1)$$

where $y_{i,k,b}$ denotes the observed value of the b -th spectral band's time series, $b \in \{1, 2\}$, of the i -th pixel at time k . The noise sample of the i -th pixel at time k for each spectral band is presented by $v_{i,k,b}$. The noise is additive but with an unknown distribution on both spectral bands. The cosine function model was separately fitted on each of the two spectral bands and is based on several different features; the frequency ω which is the same over both spectral bands, the nonzero mean $\mu_{i,k,b}$, the amplitude $\alpha_{i,k,b}$ and the phase $\phi_{i,k,b}$. The frequency can be explicitly calculated as $\omega = 2\pi f$ where f is based on the annual vegetation growth cycle and the sampling rate of the MODIS sensor. Given the 8 daily composite MCD43 MODIS data, f was calculated to be $8/365$. The values of $\mu_{i,k,b}$, $\alpha_{i,k,b}$ and the phase $\phi_{i,k,b}$ are functions of time and must be estimated for each pixel, $i, \forall i \in [1, I_{\max}]$, given both the spectral band observations $y_{i,k,b}$ for $k, \forall k \in [1, N]$, and $b \in \{1, 2\}$. The total number of pixels is denoted by I_{\max} and the total number of observations is defined by N .

A state vector is estimated by the EKF at each time increment k for each spectral band and contains all the features, and is denoted by

$$\mathbf{x}_{i,k,b} = [\mu_{i,k,b} \ \alpha_{i,k,b} \ \phi_{i,k,b}]^T. \quad (2)$$

For the present case, it was assumed that the state vector $\mathbf{x}_{i,k,b}$ does not change significantly through time; hence, the process model is linear. The measurement model, however, contains the cosine term and, as such, is evaluated via the standard Jacobian formulation, thereby linearizing the nonlinear measurement model around the current state vector [5]. The state vector $\mathbf{x}_{i,k,b}$ is related to the observation vector $y_{i,k,b}$ via a non-linear measurement function \mathbf{h}_b . Both these models are possibly non-perfect, so the addition of process $\mathbf{w}_{i,k,b}$ and measurement $v_{i,k,b}$ noise is required [5]. This is expressed as

$$\mathbf{x}_{i,k,b} = \mathbf{x}_{i,(k-1),b} + \mathbf{w}_{i,k,b}, \quad (3)$$

and

$$\hat{y}_{i,k,b} = \mathbf{h}_b(\mathbf{x}_{i,k,b}) + v_{i,k,b}. \quad (4)$$

Both state vectors features may be estimated over time k by recursive iteration [5] based on the observation data $y_{i,k,b}$ up to time k . The predicted measurement for the b -th spectral band is denoted by $\hat{y}_{i,k,b}$ in (4). Function \mathbf{h}_b is used to compute a measurement given the predicted state, and, $v_{i,k,b}$ is the observation noise vector. The estimated values for $\mathbf{x}_{i,k,b}$ over time k effectively results in a time series of feature vectors for each of the I_{\max} pixels for both spectral bands.

The general approach to estimating and initializing the state parameters, as well as the observation noise and process covariance matrix for the EKF, is to determine these parameters off-line on known training data sets. Let these input parameters to the EKF be denoted by $\mathcal{S}_{j,b}$ and \mathcal{S} denote the multidimensional space of all possible values of $\mathcal{S}_{j,b}$. The EKF will then produce features $\mathbf{x}_{i,k,b}$ if it is initialized with $\mathcal{S}_{j,b}$. Proper estimation of $\mathcal{S}_{j,b}$ through various techniques leads to good features from the EKF, while improper estimation of $\mathcal{S}_{j,b}$ will cause system instability and leads to delayed tracking in the system. A method is proposed that will use the spatial information of the geographical area to design a parameter space within the manifold where desirable system behaviour is present. This is accomplished by observing the dependencies between the initial state vectors, the process covariance matrix and the observation noise.

The analysis of the parameters in $\mathcal{S}_{j,b}$ requires three conditions that operate on a spatio-temporal window that extracts a short period of data samples from a given time series of length N . Without loss of generality it is assumed that the spatio-temporal window starts at the first time sample in the time series and that a suitable geographical size is selected. The EKF is conditioned and verified on three separate sets of parameters with the spatio-temporal window that will ensure the following three conditions holds when the separate parameters sets are presented to the EKF. The conditions are determined as

$$\sigma_{v,b} = \min_{\mathcal{S}_{j,b} \in \mathcal{S}} \left(\sum_{i=1}^{I_{\max}} \sum_{k=1}^Q \|\hat{y}_{i,k,b} - y_{i,k,b}\| \right), \quad (5)$$

$$\sigma_{\mu,b} = \min_{\mathcal{S}_{j,b} \in \mathcal{S}} \left(\sum_{i=1}^{I_{\max}} \sum_{k=1}^Q \left\| \hat{\mu}_{i,k,b} - \left(\frac{1}{Q} \sum_{k=1}^Q \mu_{i,k,b} \right) \right\| \right), \quad (6)$$

$$\sigma_{\alpha,b} = \min_{\mathcal{S}_{j,b} \in \mathcal{S}} \left(\sum_{i=1}^{I_{\max}} \sum_{k=1}^Q \left\| \hat{\alpha}_{i,k,b} - \left(\frac{1}{Q} \sum_{k=1}^Q \alpha_{i,k,b} \right) \right\| \right). \quad (7)$$

In (5), the $\sigma_{v,b}$ denotes the estimated deviation of the actual observations $y_{i,k,b}$ compared to the EKF estimated observations $\hat{y}_{i,k,b}$ of the b -th spectral band. In (6)–(7), the deviations of the estimated features are compared to the average EKF estimated features of the b -th spectral band.

Three separate probability density functions can now be derived spatially from the set of pixels $\{i\}$ for conditions stated in (5)–(7) at time k . Let $\delta_{(v,k,b)}$ denote the probability density function derived from (5), $\delta_{(\mu,k,b)}$ denote the probability density function derived from (6) and $\delta_{(\alpha,k,b)}$ denote the probability density function derived from (7).

The proposed method evaluates the fit of the candidate manifold to the three conditions in (5)–(7) using a divergence metric. In probability theory, several different f-divergence metrics exist that measure the difference between two probability density functions. The modified Hellinger distance is proposed to quantify the similarity in the range $[0, 1]$ within the manifold [6] and is computed for each condition as

$$\mathcal{H}_{v,b}(\mathcal{S}_{j,b}) = 1 - \sqrt{1 - \sqrt{\hat{q}_{v,b}(\mathcal{S}_{j,b}) \delta_{(v,k,b)}}}, \quad (8)$$

$$\mathcal{H}_{\mu,b}(\mathcal{S}_{j,b}) = 1 - \sqrt{1 - \sqrt{\hat{q}_{\mu,b}(\mathcal{S}_{j,b}) \delta_{(\mu,k,b)}}}, \quad (9)$$

$$\mathcal{H}_{\alpha,b}(\mathcal{S}_{j,b}) = 1 - \sqrt{1 - \sqrt{\hat{q}_{\alpha,b}(\mathcal{S}_{j,b}) \delta_{(\alpha,k,b)}}}. \quad (10)$$

Where in (8), $\mathcal{H}_{v,b}(\mathcal{S}_{j,b})$ denotes the modified Hellinger distance of the b -th spectral band for testing similarity between the candidate EKF and the conditioned EKF given in (5). $\hat{q}_{v,b}(\mathcal{S}_{j,b})$ denotes the probability density function of the deviation in the observations of the candidate EKF. Similarly, (9) and (10) denotes the modified Hellinger distance of the b -th spectral band for conditions (6) and (7) respectively.

A global metric that encompasses all 3 distance metrics for each b -th spectral band will be denoted as the minimum modified Hellinger distance Γ_b and is defined as

$$\Gamma_b = \min \left(\mathcal{H}_{v,b}, \mathcal{H}_{\mu,b}, \mathcal{H}_{\alpha,b} \right). \quad (11)$$

Using (11) we can iteratively search through the manifold space to find state parameters that will maximize Γ_b . The state parameters that create the candidate manifold is expressed as

$$\mathcal{S}_{\text{opt},b} = \operatorname{argmax}_{\mathcal{S}_{j,b} \in \mathcal{S}} \{\Gamma_b\}, \quad (12)$$

where $\mathcal{S}_{\text{opt},b}$ denotes the optimal search vector within the manifold. A suitable search algorithm was applied to find an optimal candidate manifold.

4. RESULTS

A spatio-temporal window was used to extract a set of sub-sequences from the time series data set. The temporal window length Q was set to 46 and the spatial window included $I_{\text{max}}=1142$ pixels into the spatio-temporal window. Three

Table 1. Standard deviation on all parameters obtained through empirical simulations on the validated data set for the first two spectral bands.

Conditions	Least Squares	EKF $\mathcal{S}_{0,b}$	EKF $\mathcal{S}_{\text{opt},b}$
$\sigma_{v,1}$	74.158	119.304	78.210
$\sigma_{\mu,1}$	184.599	190.973	176.623
$\sigma_{\alpha,1}$	66.609	117.322	86.532
$\sigma_{v,2}$	136.225	274.405	124.095
$\sigma_{\mu,2}$	229.608	256.107	211.410
$\sigma_{\alpha,2}$	167.821	181.748	97.779

different algorithms were then used to fit a pair of cosine functions to both spectral bands separately and used the mean and amplitude variables produced as features to be clustered.

The first algorithm was a linear least squares method that was used to fit a pair of cosine functions to provide a performance bound that can be expected from the given validated data set. The second algorithm was an EKF that was used to fit the cosine functions which was unbiased by setting its covariance matrix set to unity. This unbiased EKF is iteratively reprocessed until the internal state of the EKF stabilizes and is denoted by $\mathcal{S}_{0,b}$ in the experiments. The last algorithm uses the EKF again but with the proposed method derived in section 3 and is denoted by $\mathcal{S}_{\text{opt},b}$ in the experiments.

A clustering method was required to process the features produced. These features extracted were analyzed with a K-means clustering technique [7]. The K-means clustering technique is a partitional clustering technique and is usually used as a benchmark for other algorithms, and was used on all feature extraction methods evaluated in this section.

4.1. Validation data set

The standard deviations on all the parameters extracted using the algorithms are given in Table 1. It was shown that improved class separability is not solely based on a decrease in the parameter's standard deviation [3]. The clustering accuracy on the validated data set is reported in Table 2. The unbiased EKF improves on the class separability from the performance observed using the least squares method. A significant improvement is observed in the clustering accuracy when the manifold space is improved using $\mathcal{S}_{\text{opt},b}$ to 88%.

Table 2. Clustering accuracy on validated data set obtained through empirical simulations.

Conditions	Least Squares	EKF $\mathcal{S}_{0,b}$	EKF $\mathcal{S}_{\text{opt},b}$
Vegetation	76.443%	81.116%	88.042%
Settlement	91.198%	88.159%	91.374%

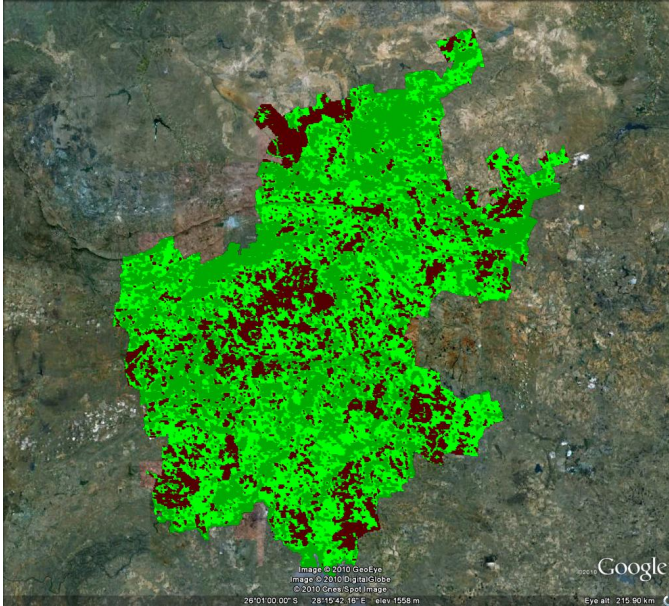


Fig. 1. Clustering results of $S_{opt,b}$ for the entire Gauteng province of March 2006. The human settlement are coded in red and natural vegetation are coded in several shades of green (courtesy of GoogleTMEarth).

4.2. Gauteng province – case study

Next the method was applied to the entire Gauteng province which covers an area of 19676 km². A silhouette graph was used to determine the optimal number of clusters for partitional clustering and resulted in five clusters being the preferred choice [7]. The clusters were evaluated and grouped into human settlement and natural vegetation (Figure 1). The clusters allocated to the human settlements had a total land coverage of 23.16% throughout the entire Gauteng province.

5. CONCLUSIONS

This paper demonstrated that improved features can be obtained by using the information within a spatio-temporal window. The proposed unsupervised feature extraction method is not dependent on acquiring a labelled training data set. It was shown that with proper selection of the initial state parameters, observation noise and process covariance matrix, the cluster separation is improved with the EKF (section 4).

The EKF that was initialized using $S_{opt,b}$, has accurately clustered the validated data set at 88% overall accuracy (Table 2). When the state parameters $S_{opt,b}$ were estimated and used over the entire Gauteng province, a total land cover of 23.16% (4556.75 km²) of human settlements were detected.

The algorithm described in this paper can be optimised by adjusting the temporal length of the spatio-temporal window, to ensure it takes cognisance of short-term inter-annual climate variability and adapts to longer-term trends in climate. The features extracted from the EKF that is initialized using $S_{opt,b}$ can also be applied in combination with a variety of other machine learning algorithms.

6. REFERENCES

- [1] G. Daily and P. Ehrlich, “Population, sustainability, and earth’s carrying capacity,” *Bioscience*, vol. 42, no. 10, pp. 761–771, November 1992.
- [2] P.M. Vitousek, H.A. Mooney, J. Lubchenco, and J.M. Melillo, “Human domination of earths ecosystems,” *Science*, vol. 277, pp. 494–499, July 1997.
- [3] W. Kleynhans, J.C. Olivier, K.J. Wessels, F. van den Bergh, B.P. Salmon, and K.C. Steenkamp, “Improving land cover class separation using an Extended Kalman Filter on MODIS NDVI Time-Series Data,” *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 2, pp. 381–385, April 2010.
- [4] C. Schaaf et al., “First operational BRDF, albedo nadir reflectance product from MODIS,” *Remote Sensing of Environment*, vol. 83, no. 1/2, pp. 135–148, November 2002.
- [5] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman filter: Particle Filters for Tracking Applications*, Artech House, London, 1 edition, 2004.
- [6] R.J. Beran, “Minimum Hellinger distance estimates for parametric models,” *Annals of Statistics*, vol. 5, no. 3, pp. 445–463, 1977.
- [7] L. Kaufman and P.J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley-Interscience, New Jersey, ninth edition, 1990.