A new Approach to Segmentation of Remote Sensing Images with Hidden Markov Models

Josef Baumgartner[†1], Marcelo Scavuzzo[*2], Cristian Rodriguez Rivero[†3] and Julian Pucheta[†4]

[†]*LIMAC, Universidad Nacional de Córdoba*
Vélez Sarsfield 1611, Córdoba, Argentina
[1]jbaumgartner@efn.uncor.edu
[3]crodriguezrivero@efn.uncor.edu
[4]jpucheta@efn.uncor.edu
[*]*Comisión Nacional de Actividades Espaciales (CONAE)*
Ruta C-45 KM 8, Falda de Cañete, Córdoba, Argentina
[2]scavuzzo@conae.gov.ar

***Abstract*—In this work, we present a new segmentation algorithm for remote sensing images based on two-dimensional Hidden Markov Models (2D-HMM). In contrast to most 2D-HMM approaches, we do not use Viterbi Training, instead we propose to propagate the state probabilities through the image. Therefore, we denote our algorithm Complete Enumeration Propagation (CEP). To evaluate the performance of CEP, we compare it to the standard 2D-HMM approach called Path Constrained Viterbi Training (PCVT). As both algorithms are not restricted to a certain emission probability, we evaluate the performance of seven probability functions, namely Gamma, Generalized Extreme Value, inverse Gaussian, Logistic, Nakagami, Normal and Weibull. The experimental results show that our approach outperforms PCVT and other benchmark algorithms. Furthermore, we show that the choice of the probability distribution is crucial for many segmentation tasks.**

***Resumen*—En este trabajo se presenta un nuevo algoritmo de segmentación de imágenes de teledetección basado en modelos ocultos de Markov de dos dimensiones (2D-HMM). En contraste con la mayoría de los enfoques 2D-HMM, no se utiliza Viterbi Training, en cambio proponemos propagar las probabilidades de estado a través de la imagen. Por lo tanto, denotamos nuestro algoritmo Complete Enumeration Propagation (CEP). Para evaluar el desempeño del CEP, lo comparamos con el algoritmo estándar 2D-HMM llamado Path Constrained Viterbi Training (PCVT). Como ambos algoritmos no están restringidos a una cierta probabilidad de emisión, se evalúa el desempeño de siete funciones de probabilidad, a saber, Gamma, Valor Extremo Generalizado, Gauss inversa, Logística, Nakagami, Normal y Weibull. Los resultados experimentales muestran que nuestro enfoque supera PCVT y otros algoritmos de referencia. Además, se muestra que la elección de la distribución de probabilidad es crucial para muchas tareas de segmentación.**

I. Introduction

The segmentation of images plays an important role in many fields of computational intelligence [1], [2]. In this work, we try to segment satellite images of rural areas. On one hand, our goal is to determine the type and state of crops in agricultural regions in order to predict market prices of local stock exchange. On the other hand, we try to detect forest fires at an early stage [3].

To segment remote sensing images we present two algorithms based on the theory of Hidden Markov Models (HMM). In the last years HMM were successfully applied to one-dimensional problems like speech recognition [4] or the analysis of biometric data [5]. In contrast to that, the pixels of an image are a representation of two-dimensional data. Hence, we have to extend the classical HMM to two dimensions. Mathematically this extension is straightforward but the standard method of parameter estimation for one-dimensional HMM, the Baum-Welch algorithm [6], is not feasible for higher dimensions. So the main question for two-dimensional Hidden Markov Models (2D-HMM) is: How can the computational complexity be reduced in order to make the n-dimensional HMM feasible?

The firsts attempts to extend the HMM framework to two dimensions were the so called pseudo 2D-HMM. Pseudo 2D-HMM use superstates to represent the rows of an image, while the columns are connected by a simple Markov Chain [7]. Some years later, the Viterbi Algorithm was applied to obtain a feasible version of a true 2D-HMM [8]. This algorithm was successfully applied to segment man-made landscape like roads or cities from natural areas [9]. Still, this approach is only feasible for blocks of 8×8 pixels.

In order to segment whole images without dividing them into blocks, the Path Constrained Viterbi Training (PCVT) was developed on the basis of the Viterbi Algorithm [10]. Since then, the PCVT has been used by many researchers for different image processing tasks [11], [12] where it showed good results compared to benchmark algorithms [13].

Even so the PCVT is a widely accepted segmentation algorithm, it has a notable drawback. Before decoding a 2D-HMM it is necessary to pre-select a small number of possible states. In other words, we are throwing away most of the possible states of every pixel, before we actually use the 2D-HMM. To avoid this problem, we present a new, feasible approach to 2D-HMM that propagates the state probabilities through the image without discarding any state. We denominate our algorithm Complete Enumeration Propagation (CEP) [14].

This paper is organized as follows: In Section II we present the mathematical background of 2D-HMM and explain why further assumptions are necessary to make the 2D-HMM feasible. Thereafter, we present PCVT and CEP in detail and show how both algorithms can be used with an arbitrary probability density function as an emission prob-

978−1−4799−4269−5/14/$31.00

ability. In Section IV we present the experimental results of PCVT, CEP and two benchmark algorithm for three test images. Before we draw the conclusions in Section VI we discuss some important advantages of CEP over PCVT in Section V.

II. Extension of Hidden Markov Models to two dimensions

In this section, we describe the theoretical background of 2D-HMM. A Hidden Markov Model is a probabilistic model, that is used to analyze and describe correlated noisy data. Given an image I of size $M \times N$, the main goal of the segmentation process is to find the hidden state $s_{i,j} \in 1, 2, \ldots, S$ of every pixel (i, j) with $i \in 1, 2, \ldots, M$ and $j \in 1, 2, \ldots, N$.

So, how can we find the hidden state of a pixel? Clearly there are several factors that influence the choice of the hidden state. First of all, we will talk about contextual information. Let us suppose that the image is a Markov Random Field. This means, that, given the image, the hidden state of pixel (i, j) is conditionally independent of the pixels outside a certain neighborhood. Obviously this is a very reasonable assumption. For example the hidden state of a pixel in a crop field is definitely independent of a pixel in city that lies many kilometers away.

Hence, the first thing we have to do is to define a neighborhood for every pixel. Keep in mind, that one-dimensional hidden Markov models use "past" states as contextual information. In other words, the state s_t at time t depends on the previous state s_{t-1} according to a certain transition probability. But how can we introduce "past" for two-dimensional data like the pixels of an image?

For pixel (i, j) we define $(i^{'}, j^{'}) \prec (i, j)$ if $i^{'} < i$ or $i^{'} = i$ and $j' < j$. It can be shown, that under the Markov assumption this definition leads to a 2nd order Markov Mesh which specifies for state $s_{i,j}$:

$$P(s_{i,j}|s_{i',j'} : (i^{'}, j^{'}) \prec (i, j)) = P(s_{i,j}|s_{i,j-1}, s_{i-1,j}).$$

The two pixels $(i, j-1)$ and $(i-1, j)$ can be understood as the "past" of pixel (i, j) as shown in Figure 1. As a result we are moving from the top-left pixel to the bottom-right pixel. Along this way, we assume, that the transition probabilities from states $s_{i,j-1}$ and $s_{i-1,j}$ to state $s_{i,j}$ do not depend on the current pixel. Therefore, we can gather all transition probabilities in a matrix A where

$$a_{m,n,l} = P(s_{i,j} = l | s_{i,j-1} = m, s_{i-1,j} = n)$$

for $l, m, n \in 1, 2, \ldots, S$. The exact formula for the transition probabilities $a_{m,n,l}$ is given by

$$a_{m,n,l} = \frac{\sum_{i,j} H_{m,n,l}(i, j)}{\sum_{l'=1}^{S} \sum_{i,j} H_{m,n,l'}(i, j)}. \quad (1)$$

In equation (1), $H_{m,n,l}(i, j)$ is the probability of a transition from states m and n to state l at pixel (i, j) for any possible hidden state map. Without going into detail, note, that there are S^{MN} possible hidden state maps. Thus, it is impossible to calculate $H_{m,n,l}(i, j)$ even for small images with few states. In the next section we discuss this topic with more detail.

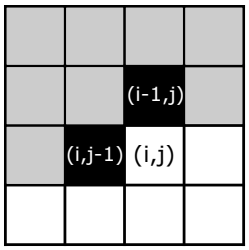

Fig. 1. Transitions among states in a 2nd order Markov Mesh. The gray and the black pixels fulfill $(i^{'}, j^{'}) \prec (i, j)$ but the two black pixels are sufficient statistics for pixel (i, j) under the Markov assumption.

Besides the contextual information, the state of pixel (i, j) depends obviously on the pixel intensity. We suppose that all hidden states are observed through a certain emission probability. Hereby we would like to point out, that the mathematical framework of 2D-HMM does not require a particular probability distribution to be used as an emission probability. Still, most researchers use the normal distribution because it is a well known distribution with two meaningful parameters [10], [15].

In the following, we show exemplary how to estimate the parameters of a normal distribution, but keep in mind that the estimation of other probability functions is straightforward, too. Let's suppose we have observed state $l \in 1, 2, \ldots, S$ through a normal distribution with mean μ_l and standard deviation σ_l. The emission probability of state l is than given as

$$b_l(x) = P(x|s_{i,j} = l) = \frac{1}{\sigma_l \sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left(\frac{x - \mu_l}{\sigma_l} \right)^2 \right\}.$$

The mean μ_l and the standard deviation σ_l of $b_l(x)$ can be estimated by

$$\mu_l = \frac{\sum_{i,j} L_l(i, j) I_{i,j}}{\sum_{i,j} L_l(i, j)} \quad (2)$$

$$\sigma_l = \frac{\sum_{i,j} L_l(i, j) \left(I_{i,j} - \mu_l\right) \left(I_{i,j} - \mu_l\right)^T}{\sum_{i,j} L_l(i, j)} \quad (3)$$

where $L_l(i, j)$ indicates the probability of pixel (i, j) being in state l for any possible hidden state map. And just like in case of $H_{m,n,l}(i, j)$ we are facing computational problems when calculating $L_l(i, j)$ because of the very high number of possible hidden state maps.

Note, that equations (1), (2) and (3) can be derived directly from the Expectation-Maximization algorithm (EM) [16]. Hence, there is a theoretical prove, that an iteration of the 2D-HMM parameters converges.

Before we describe in the next section how the parameters of the 2D-HMM can be estimated, we like to give a formal representation of the optimal segmentation $\mathbf{s}^*$:

$$\mathbf{s}^* = \arg\max_{\mathbf{s}} P(\mathbf{s}|I, \theta). \quad (4)$$

In equation (4), θ are the parameters of the 2D-HMM, such as the initial probabilities π_l, the transition probabilities $a_{m,n,l}$ and the emission parameters μ_l and σ_l of each state. Please keep in mind, that the $\mathbf{s}$ in equation (4) stands for all possible hidden state maps which is usually a infeasible high number.

III. Implementations of two-dimensional Hidden Markov Models

In the previous section we pointed out, that some parameters of the 2D-HMM are infeasible even for small images. In this section we investigate the complexity problems with more detail and show different ways of how the exact 2D-HMM can be approximated. Note, that by using approximations instead of exact formulas, we leave the EM framework and thus have no more a theoretical guarantee that our 2D-HMM converges. Still, it is our only choice if we want to come up with a feasible model that – at least – approximates the optimal hidden state map as defined by equation (4).

A. Parameter estimation

Before we show approximations of the exact 2D-HMM, we like to illustrate some computational problems when trying to calculate $H_{m,n,l}(i,j)$ and $L_l(i,j)$. Let's think of a really small image of size 10×10 with three states. This would mean that you have to sum up $3^{10*10} \approx 5.2 * 10^{47}$ elements to get $H_{m,n,l}(i,j)$. Even if every element of the sum would take only one millisecond to calculate, it would still take billions of millenniums until we found $H_{m,n,l}(i,j)$. The same is valid for the computation of $L_l(i,j)$.

In order to reduce the computational burden, we propose an approximation of the transition probabilities $a_{m,n,l}$ and of the emission parameters μ_l and σ_l. Instead of summing over all possible state maps we suggest to estimate all parameters by using only the state map of the current iteration. For $a_{m,n,l}$ in iteration step $p+1$ we get

$$a_{n,m,l}^{(p+1)} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M} I\left(s_{i-1,j}^{(p)}=n, s_{i,j-1}^{(p)}=m, s_{i,j}^{(p)}=l\right)}{\sum_{i=1}^{N}\sum_{j=1}^{M} I\left(s_{i-1,j}^{(p)}=n, s_{i,j-1}^{(p)}=m\right)}, \tag{5}$$

where $I(\cdot)$ is the indicator function. For the means and standard deviations of each state l we use

$$\mu_l^{(p+1)} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M} I\left(s_{i,j}^{(p)}=l\right) I_{i,j}}{\sum_{i=1}^{N}\sum_{j=1}^{M} I\left(s_{i,j}^{(p)}=l\right)} \tag{6}$$

$$\sigma_l^{(p+1)} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M} I\left(s_{i,j}^{(p)}=l\right) (I_{i,j}-\mu_l)(I_{i,j}-\mu_l)^T}{\sum_{i=1}^{N}\sum_{j=1}^{M} I\left(s_{i,j}^{(p)}=l\right)}. \tag{7}$$

One can think of these simplified formulas as "count instead of evaluate", because we are only taking into account our currently best guess of the hidden state map instead of evaluating all possible state maps. Therefore, the approximations presented in equations (5), (6) and (7) are computationally extremely simple. All we have to do, is to process the state and the observation of each pixel.

Note, that switching to another emission probability – like a Gamma or a Weibull distribution – is straightforward in this context. All we have to do, is to replace equations (6) and (7) by the corresponding parameter estimators. For maximum likelihood estimators of many common probability distributions please refer to [17].

B. Evaluation of two-dimensional Hidden Markov Models

After solving the problem of parameter estimation in the previous section, we are left with the problem of decoding the 2D-HMM in order to obtain a hidden state map. Therefore, remember our notion of "past" as shown in Figure 1. So we can think of each bottom-left to upper-right diagonal of the image as one step in time, starting with the top-left pixel. The diagonals $T_0, T_1, T_2, \ldots$ consist of

$$T_0 = (s_{0,0}); T_1 = (s_{1,0}, s_{0,1}); T_2 = (s_{2,0}, s_{1,1}, s_{0,2}); \ldots$$

Because we are dealing with a 2nd order Markov Mesh we can make the Markov assumption and get

$$\begin{aligned} P(s) =& P(T_0)P(T_1|T_0)\ldots P(T_{z+w-2}|T_{z+w-3}, \ldots, T_0) \\ =& P(T_0)P(T_1|T_0)\ldots P(T_{z+w-2}|T_{z+w-3}). \end{aligned} \tag{8}$$

Note in equation (8), that each diagonal operates as an "isolating element" between neighboring diagonals. Hence, we have transformed the complex two-dimensional model into a pseudo one-dimensional HMM.

At this point, we could think of running a one-dimensional HMM decoder like the Baum-Welch algorithm [6], but once again we are facing computational issues. Remember, that in a $M \times N$ image the diagonals consist of up to $min(M,N)$ pixels. If there are S states, a diagonal can be in one of $S^{min(M,N)}$ superstates. So we are dealing with a one-dimensional HMM with up to $S^{min(M,N)}$ superstates at each time step. Such a HMM is clearly not feasible.

Before we present two feasible 2D-HMM decoders we have to make the assumption that the emission probability of pixel (i,j) depends only on the current state and not on neighboring states. Even though this assumption is reasonable and not very restrictive, both approaches relay on it.

1) Path Constrained Viterbi Training: In the last years, several approximations were proposed to make the 2D-HMM feasible. One of the most promising approaches is to cut down the superstates of each diagonal to a certain number Z and then run the Viterbi algorithm [10]. This algorithm is named *Path Constrained Viterbi Training* because the number of superstates – and hence the number of paths through the image – is constrained.

The selection of superstates is done by maximum likelihood, which means, that superstates are discarded without taking into account neighboring superstates. Thus, there is no guarantee, that the chosen superstates are close to the optimal hidden state map. At least the Viterbi algorithm considers all dependencies among neighboring pixels when evaluating the independently chosen superstates.

A sketch of the PCVT as presented by [10] is shown in Algorithm 1. Note, that the computation of step 5) has order $\mathcal{O}((M+N-1)Z^2)$ for an image of size $M \times N$, whereas the other steps have negligible computational complexity.

2) Complete Enumeration Propagation: In the following we present a new 2D-HMM decoding algorithm, that, in contrast to the PCVT, does not discard any possible hidden state and still is feasible. Instead of grouping the states of a diagonal in a superstate, we use complete enumeration to calculate the state probabilities of each pixel. Then we propagate the state probabilities through the image until we

Algorithm 1: Path-Constrained Viterbi Training (PCVT)

1) Initialize parameters μ_l and σ_l for $l \in \mathcal{S}$.
2) Initialize state map using ML segmentation.
3) Calculate transition probabilities $a_{n,m,l}$ for every $n, m, l \in \mathcal{S}$ using equation (5).
4) Choose the Z best superstates for each diagonal using Maximum Likelihood.
5) Run Viterbi algorithm.
6) Update parameters $a_{n,m,l}$, μ_l and σ_l using equations (6), (7) and (5).
7) Iterate steps 4), 5) and 6) until convergence.

Algorithm 2: Complete Enumeration Propagation (CEP)

1) Initialize parameters μ_l and σ_l for $l \in \mathcal{S}$.
2) Initialize state map using ML segmentation.
3) Calculate transition probabilities $a_{n,m,l}$ for every $n, m, l \in \mathcal{S}$ using equation (5).
4) Find new state map using equations (9) and (10).
5) Update parameters $a_{n,m,l}$, μ_l and σ_l using equations (6), (7) and (5).
6) Iterate steps 4) and 5) until convergence.

reach the bottom-right pixel. Therefore, we call our approach *Complete Enumeration Propagation*.

Let's start with the state probability of pixel (i, j) which is given by

$$P(s_{i,j}|s_{i,j-1}, s_{i-1,j}, I_{i,j}) \propto P(s_{i,j}, s_{i,j-1}, s_{i-1,j}, I_{i,j}) = \\ = P(s_{i,j-1}, s_{i-1,j}) P(s_{i,j}|s_{i,j-1}, s_{i-1,j}) P(I_{i,j}|s_{i,j}).$$

If we now replace $P(s_{i,j}|s_{i,j-1}, s_{i-1,j})$ by the transition probability $a_{s_{i,j-1}, s_{i-1,j}, s_{i,j}}$ and consider two diagonal pixels to be independent we can write

$$P(s_{i,j}|s_{i,j-1}, s_{i-1,j}, I_{i,j}) \propto \\ P(s_{i,j-1}) P(s_{i-1,j}) a_{s_{i,j-1}, s_{i-1,j}, s_{i,j}} P(I_{i,j}|s_{i,j}). \tag{9}$$

This is the main formula to calculate the state probabilities of pixel (i, j) given the observation and the two past states. Note, that the main assumption here is to suppose independence of the two past pixels. This assumptions may seem restrictive but keep in mind, that the spatial relation between neighboring pixels is already incorporated in the transition probabilities $a_{m,n,l}$. Besides that, the only way to leave out the independence assumption would be to estimate the joint probability distribution of the pixels $s_{i,j-1}$ and $s_{i-1,j}$. This could improve the segmentation in some cases but especially in noise images the joint distribution could also lead to worse results. Furthermore, PCVT has to make the same assumption when searching for the Z superstates of each diagonal.

Now, the main idea of CEP is to use equation (9) to calculate $P(s_{i,j} = l)$ for $l = 1, 2, \ldots, S$ for all possible combinations of past states, i.e. $s_{i,j-1} = m$, $s_{i-1,j} = n$ for $m, n = 1, 2, \ldots, S$ according to equation (10).

$$P(s_{i,j} = l|I_{i,j}) \propto \sum_{m=1}^{S} \sum_{n=1}^{S} a_{s_{i,j-1}=m, s_{i-1,j}=n, s_{i,j}=l} \\ P(s_{i,j-1} = m) P(s_{i-1,j} = n) P(I_{i,j}|s_{i,j} = l). \tag{10}$$

This procedure is nothing else than complete enumeration of $P(s_{i,j} = l)$. Keep in mind, that before we can go on with the next pixel, it is necessary to normalize $P(s_{i,j} = l)$ such that $\sum_{l=1}^{S} P(s_{i,j} = l) = 1$. For a schematic description of CEP please see Algorithm (2).

When analyzing the computational complexity of CEP, only the calculation of step 4) is worth mentioning. This step is of order $\mathcal{O}((MN)S^3)$ for an image of size $M \times N$ with S states.

A problem arises for the pixels on the left and upper edge of the image because there are no past states $s_{i,j-1}$ or $s_{i-1,j}$. To solve this issue one can think of two possible solutions. First, copy the first row and the first column and use maximum likelihood to determine the probabilities of these auxiliary pixels. Second, suppose a uniform distribution for the nonexistent terms $P(s_{i,j-1} = m)$ and $P(s_{i-1,j} = n)$. This is equal to leaving out the corresponding terms in equation (9). We prefer the second option, because otherwise noisy observations on the edges are encouraged to stay in a maximum likelihood state instead of adapting themselves to their neighborhood.

Once we have calculated the probabilities of all the pixels we assign each pixel the most probable state. The result is a hidden state map which, for now, is our best guess of the optimal hidden state map $\mathbf{s}^*$ as shown in equation (4). From this point on we use the formulas (5), (6) and (7) to update the parameters of the emission probabilities. Then we iterate this procedure until convergence.

C. Emission probabilities

As we have shown in Section III-A, PCVT and CEP allow different probability distributions as emission probabilities. To evaluate the influence of different probability functions we run the two algorithms with the following seven distributions: *Gamma*, *Generalized Extreme Value*, *inverse Gaussian* (also known as *Wald distribution*), *Logistic*, *Nakagami*, *Normal* and *Weibull* as defined by [17]. By comparing the segmentation results we try to determine the optimal distribution for every remote sensing image.

IV. Segmentation results

In this section we evaluate the different segmentation algorithms. As mentioned in Section III, there is no theoretical guarantee that the two segmentation algorithms converge. Hence, we set a maximum of 200 iterations for all experiments, which, in few cases, stopped the segmentation process. Nevertheless, we observed almost no changes of $\widehat{\kappa}$ for more than 150 iterations.

As benchmark methods we use Maximum Likelihood (ML) as described in Algorithm 3 and Iterated Conditional Modes (ICM) [18]. ML is a classical non-contextual segmentation method and ICM is a well studied algorithm that has been used for image segmentation tasks for over two decades. Both approaches are used as benchmark algorithms in many works [13], [15].

A. Landsat 7 image

The first experiment is a Landsat 7 image of an agricultural area in the humid pampa of Argentina. It shows agricultural fields of different sizes and orientations and

Algorithm 3: Maximum Likelihood Classification

1) Initialize parameters μ_l and σ_l for $l \in \mathcal{S}$.
2) Calculate $P(s_{i,j} = l|I_{i,j}, \theta) = P(I_{i,j}|s_{i,j} = l, \theta)$ for each pixel (i, j) and for each state l.
3) Assign pixel (i, j) the label given by $s_{i,j} = \arg\max_{l \in \mathcal{S}} P(I_{i,j}|s_{i,j} = l, \theta)$.

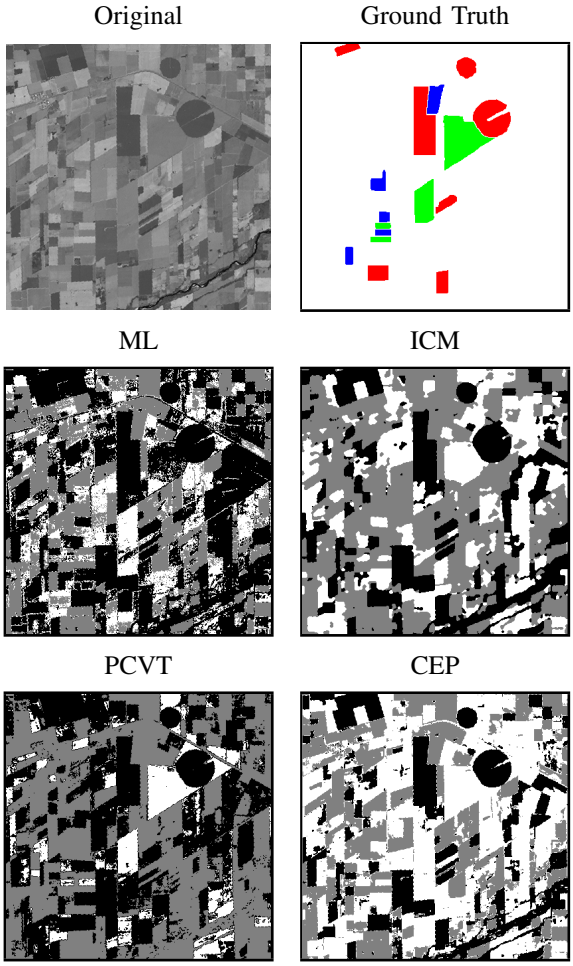

Fig. 2. Segmentation of band five of a Landsat 7 image using three states. While ML has problems of merging corps fields, ICM is leading to a segmentation with few details. The best result in terms of κ is obtained by CEP as shown in Figure 3.

two center-pivot irrigations. In this case, the performance is evaluated in the parts of the image that are shown in Figure 2, since we only have ground truth labels for these regions. We show some of the segmentation results in Figure 2 and in Figure 3 we plot Cohen's $\widehat{\kappa}$ coefficient [19] of all algorithms.

B. *Landsat* 8 *image*

In this experiment we try to identify burning forest in a Landsat 8 image. Therefore we use the thermal infrared band 10. In Figure 4 the segmentation results are shown. Unfortunately we have no exact ground truth so we can not calculate the $\widehat{\kappa}$ values.

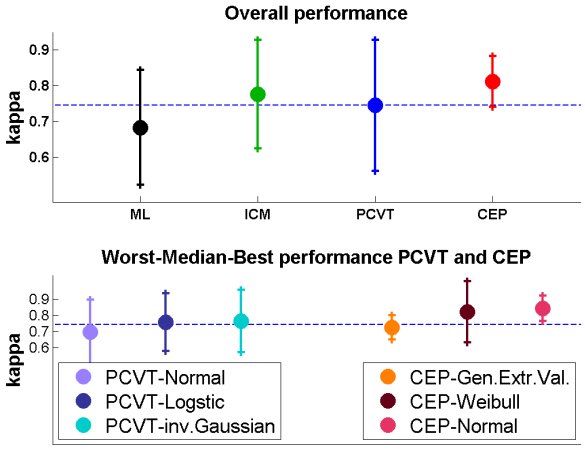

Fig. 3. Above: Overall comparison of ML, ICM, PCVT and CEP. Below: Performance of the worst, the median and the best emission probability plugged into PCVT and CEP. In total, we run 12 experiments that we generated by varying the initial values (four different initial values) and by pre-filtering the original image (three different images: no filter, 3×3 filter and 5×5 filter). The dots mark the mean $\widehat{\kappa}$ coefficients of the 12 experiments while the lines indicate the standard deviation.

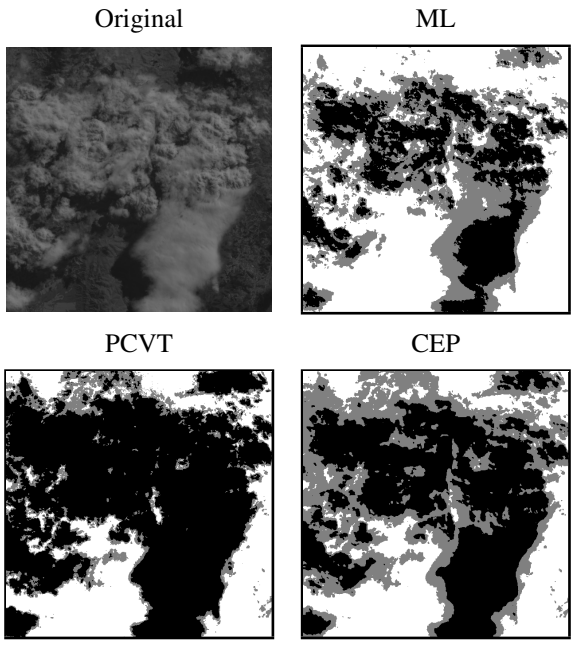

Fig. 4. Segmentation of burning forest in a Landsat 8 image. Note, that ML leads to a very pixelated segmentation, while PCVT is not distinguishing well between the burning region (black) and the smoky region (gray).

C. *Cosmo-SkyMed image*

Finally we use a Level $1B$ Cosmo-SkyMed image with two bands to test our algorithms. This image has a very low signal-to-noise ratio and the goal is to find two crop fields as well as a center-pivot irrigation. The ground truth and the experimental results are shown in Figure 5.

V. Discussion

In this work, we presented a new approach to segmentation of remote sensing images and compared it with standard

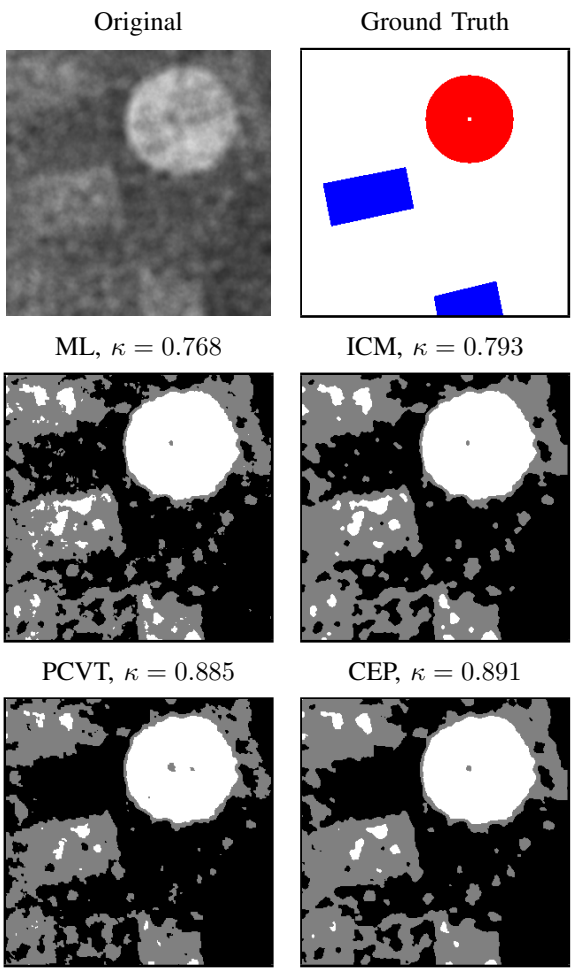

Fig. 5. Crops segmentation using a noisy Cosmo-SkyMed image. At first glance, all algorithms show good results but the κ coefficient reveals that CEP works better then the other methods.

algorithms. Especially in the Landsat 7 and the Landsat 8 experiment one can see that CEP preserves more details of the original image then PCVT and ICM and at the same time merges crop fields much better then ML. The κ coefficients of the Landsat 7 images as well as the Cosmo-SkyMed image support this observation.

On the other hand it has to be said, that CEP has the highest computational complexity. The main drawback is, that the order of CEP depends linearly on the number of pixels and to the power of three on the number of states. Hence the number of states is crucial for the run-time of CEP.

Finally, we like to point out, that CEP shows the best results when using it with the Normal probability distribution, whereas PCVT is preferably used with other probability distributions like the inverse Gaussian distribution. Note, that in contrast to PCVT and CEP, the ICM framework cannot be extended to other probability distributions without leaving its theoretical foundations.

VI. Conclusions

With regard to the experimental results, we can conclude that CEP has a very high capability of segmenting remote sensing images. Still, one has to keep in mind, that CEP requires more computational resources then its competitors. Hence, you can think of CEP not only as a challenge to other segmentation algorithms but also a complementary. In future works, it would be interesting to extend the 2D-CEP framework to higher dimensions.

Acknowledgment

This work has been partially supported by CONICET. The authors would like to thank CONAE and the UNC.

References

[1] L. Ding, P. Tang, and H. Li, "Subspace feature analysis of local manifold learning for hyperspectral remote sensing images classification," *Applied Mathematics and Information Sciences*, vol. 8, no. 4, pp. 1987 – 1995, 2014.

[2] R. Mehta, J. Yuan, and K. Egiazarian, "Face recognition using scale-adaptive directional and textural features," *Pattern Recognition*, vol. 47, no. 5, pp. 1846 – 1858, 2014.

[3] S. W. Taylor, D. G. Woolford, C. B. Dean, and D. L. Martell, "Wildfire prediction to inform fire management: Statistical science challenges," *Statistical Science*, vol. 28, no. 4, pp. 586 – 615, 2013.

[4] M. Rabiner and S. Young, "The application of hidden markov models in speech recognition," *Foundations and Trends in Signal Processing*, vol. 1, no. 3, pp. 195 – 304, 2007.

[5] W. Dyrka, J. Nebel, and M. Kotulska, "Probabilistic grammatical model for helix-helix contact site classification," *Algorithms for Molecular Biology*, vol. 8, no. 1, 2013.

[6] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occuring in the statistical analysis of probabilistic functions of markov chains," in *Annals of Mathematical Statistics 1*, 1970, pp. 164 – 171.

[7] C. C. Yen and S. S. Kuo, "Degraded documents recognition using pseudo 2d hidden markov models in gray-scale images," in *Proceedings of SPIE 2277*, 1994, pp. 180 – 191.

[8] J. Li and R. Gray, *Image Segmentation and Compression using Hidden Markov Models*. Kluwer Academic Publishers, 2000.

[9] J. Li, A. Najmi, and R. M. Gray, "Image classification by a two dimensional hidden markov model," *IEEE Transactions on Signal Processing*, vol. 48, no. 2, pp. 517 – 533, 2000.

[10] D. Joshi, J. Li, and J. Z. Wang, "A computationally efficient approach to the estimation of two- and three-dimensional hidden markov models," in *IEEE Transactions on Image Processing*, vol. 15, no. 7, 2006, pp. 1871 – 1886.

[11] X. Ma, D. Schonfeld, and A. Khokhar, "Image segmentation and classification based on a 2d distributed hidden markov model," in *Proc. SPIE 6822, Visual Communications and Image Processing*, 2008.

[12] A. Meraoumia, S. Chitroub, and A. Bouridane, "2d and 3d palmprint information, pca and hmm for an improved person recognition performance," *Integrated Computer-Aided Engineering*, vol. 20, no. 3, pp. 303 – 319, 2013.

[13] A. G. Flesia, J. Baumgartner, J. Gimenez, and J. Martinez, "Accuracy of map segmentation with hidden potts and markov mesh prior models via path constrained viterbi training, iterated conditional modes and graph cut based algorithms," Send to Pattern Recognition, Elsevier, Julio 2013, iSSN: 0031-3203.

[14] J. Baumgartner, A. G. Flesia, J. Gimenez, and J. Pucheta, "A new approach to image segmentation with two-dimensional hidden markov models," in *1st BRICS Countries Congress on Computational Intelligence BRICS-CCI*, Recife, Brasil, September 2013.

[15] K. P. Pyun, J. Lim, and R. M. Gray, "A robust hidden markov gauss mixture vector quantizer for a noisy source," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1385 – 1394, 2009.

[16] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society*, vol. 39, no. 1, pp. 1 – 38, 1977.

[17] J. Rice, *Mathematical Statistics and Data Analysis*. Brooks Cole, 2007.

[18] J. E. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society*, vol. 48, no. 3, pp. 259 – 302, 1986.

[19] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37 – 46, 1960.