Comparison of Partitions of Two Images for Satellite Image Time Series Segmentation

Alain Giros
Image Processing and Products Office
Centre National d'Etudes Spatiales
Toulouse, France
Alain.Giros@cnes.fr

***Abstract*—The availability of high resolution image time series raises new problems in the field of image processing. This paper has the perspective of achieving a consistent segmentation of a time series and goes in this direction by proposing a non trivial composite segmentation of the time series. For doing so we need a segmentations comparison metric which is robust and meaningful. We thus use an information theory based distance which measures the amount of information which is not shared by two random variables in order to compute the distance between two partitions. Applied to the watershed segmentation algorithm, we are then able to handle its regions fusion tree in several ways which lead to the desired composite segmentation.**

Keywords- image segmentation; image time series; partition; mutual information; information theory; segmentation comparison.

I. INTRODUCTION

High resolution remote sensing satellites have long been used for studying steady phenomena, but few work has been done with image time series as they were not easily available. Nowadays, remote sensing systems capabilities allow gathering image time series made of several tens of images covering a period of some months on a given geographic location. These new datasets are very promising as they allow to conduct scene analysis based not only on the instantaneous pixels values but also on the temporal behaviour of every object contained within the scene. Indeed a lot of new questions are raised by such datasets. Among all these questions we focus our work on obtaining a meaningful segmentation of such a time series. This choice is motivated by the conviction that such a segmentation could be useful in several standard image time series processing problems as well as in other application oriented tasks as for instance object behaviour modelling, change detection, removal of clouds and many others.

In section II of this paper we briefly present the possible strategies for designing such a segmentation and our preferred choice. Then in section III we explain what a segmentation is and why we restrict our analysis to their partitioning property. This restriction allows us to devise a comparison criterion presented in section V, based on entropic measures on partitions defined in section IV. In section VI and VII we show how we use this criterion for segmentations comparison and image time series segmentation. Finally section VIII concludes.

II. STRATEGIES FOR IMAGE TIME SERIES SEGMENTATION

Given a set of N registered images, each of them possibly with several spectral bands, the question is: how can we make a meaningful segmentation of such a dataset ?

A first approach is to build a 2D space with, at each pixel location, a vector containing N times the number of spectral components of each image. Then we make a 2D segmentation of this vector arrangement. This approach gives a clean result which is presented in fig. 1, but with several drawbacks, the most important being that the chronology of the samples in the time series is lost, which is not acceptable.

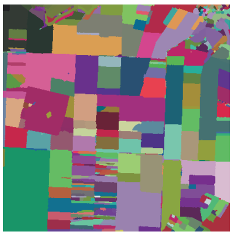


Figure 1. Segmentation of a 2D space, with a temporal vector at each pixel.

We can then imagine a second approach which consists in building a 3D space made by gathering the 2 spatial axis with the temporal one, this space being segmented in order to identify 3D regions. The result of this approach is shown in fig. 2. But, even if we have kept the right chronology of the data, the result suffers from several drawbacks: first the regions are not sharply delineated along the temporal axis and second the segmentation algorithm must cope with incommensurable quantities (space and time).

a) time t-1 b) time t c) time t+1

Figure 2. Three consecutive slices along the time axis of a 3D segmentation. Note that some regions span over several consecutive time samples.

Finally the third strategy is called the 2D+T segmentation and is our preferred one. It consists in segmenting each time sample independently and finding some significant links between the time-localized consecutive segmentations. For doing so must be able to compare segmentations and this will be the topic of the next 4 sections. We also need to build a reference segmentation which is closest, in some sense, to each of the time-localized ones. This will be detailed in section VII.

III. SEGMENTATIONS CONSIDERED AS PARTITIONS

A. Segmentation of an image

Traditionally image segmentation is viewed as a process which split the entire image in a set of regions, called itself a segmentation. This set of regions is such that:

1. every pixel of the image belongs to a region,
2. no pixel belongs to several regions,
3. the regions are spatially connected.

Properties 1) and 2) tell us that the segmentation process deals with every pixel and takes hard decisions for each of them. The result is that a segmentation is essentially a partition of the image space in the sense of the set theory. Property 3) is associated to the concept of neighborhood: given some connecting schema (e.g. 4- or 8-neighborhood in case of a spatial 2D space) defined over S, a subset R of S is said to be connected if for every pair of elements in R, it exists a connecting path between these elements which remains completely inside R. In the context of segmentation, this topological property is desired in order to get regions with spatial consistency, thus leading to a result which can be viewed as a simplification of the original image. However in the rest of the paper, we will not be interested in this topological property which is often viewed as one of the most important for segmentation. We will restrict ourselves to the partition view and we will devise a method for comparing segmentations which does not rely on any topological property. Other properties can also describe what a segmentation is, for instance the fact that some predicate must remain true within a region in order to insure consistency with the underlying pixels. As we focus our work on the result of the segmentation, we are not interested in the regions building process and will therefore restrict our discussion to comparison of partitions.

B. Partition of a set

1) Notation. Let S be a set of discrete elements. $|S|$ will denote the number of elements of S (i.e. its cardinality).

2) Definition. Let P be a set of subsets of S: $P = \{R_1, \ldots, R_{|P|}\}$. P is a partition of S if and only if :

- The union of all elements of P is equal to the set S:
$$\bigcup_{i=1}^{|P|} R_i = S$$
- The elements of P are pair wise disjoint: $\forall i, 1 \le i \le |P|,\ \forall j, 1 \le j \le |P|,\ i \ne j,\ R_i \cap R_j = \varnothing$

3) Specific partitions. Among all the possible partitions of S, two of them are of particular interest:

- $P_{\max}$, the partition composed of $|S|$ elements, each being a singleton in S: $P_{\max} = \{R_1, \ldots, R_{|S|}\}$
- $P_{\min}$, the partition composed of a single element, S itself: $P_{\min} = \{S\}$.

IV. ENTROPIC MEASURES ON PARTITIONS

This section introduces several entropic quantities measured on partitions, which will serve as a basis for the comparison criterion in the next section. These quantities are the entropy of a partition, the joint entropy of two partitions, their conditional entropy, and their mutual information.

A. Probability of a subset

First we define the probability $p(R_i)$ of a subset R_i belonging to P as the probability that a given element of S belongs to R_i. Its formulation is trivial: $p(R_i) = \frac{|R_i|}{|S|}$.

B. Entropy of a partition

The entropy of a partition P is defined as the uncertainty on the subset of P to which an element picked at random in S belongs:

$$H(P) = -\sum_{i=1}^{|P|} p(R_i)\log(p(R_i)) = -\sum_{i=1}^{|P|} \frac{|R_i|}{|S|}\log\left(\frac{|R_i|}{|S|}\right)$$

Its value lies between a lower and an upper bounds which are respectively $H(P_{\min}) = 0$ and $H(P_{\max}) = \log(|S|)$.

C. Joint entropy of a pair of partitions

Given a pair of partitions P_1 and P_2 of S we can form the joint entropy of this pair by considering all the intersections of the elements R_i of P_1 with the elements Q_j of P_2:

$$H(P_1, P_2) = -\sum_{i=1}^{|P_1|}\sum_{j=1}^{|P_2|} p(R_i \cap Q_j)\log(p(R_i \cap Q_j))$$

The following properties hold for the joint entropy:

- Symmetry : $H(P_1, P_2) = H(P_2, P_1)$

- If P_1 and P_2 are statistically independent: $H(P_1,P_2)=H(P_1)+H(P_2)$
- $H(P_{\min},P)=H(P)$
- $H(P_{\max},P)=H(P_{\max})=\log(|S|)$

D. *Conditional entropy of a pair of partitions*

We can now define the conditional entropy of a partition P_1 of S given the partition P_2 of S by the classical relation:

$$H(P_1|P_2)=H(P_1,P_2)-H(P_2).$$

The following properties hold for this conditional entropy:

- If P_1 and P_2 are statistically independent: $H(P_1|P_2)=H(P_1)$
- $H(P_{\min}|P)=H(P|P_{\max})=0$
- $H(P|P_{\min})=H(P)$
- $H(P_{\max}|P)=\log(|S|)-H(P)$

E. *Mutual information of 2 partitions*

Finally we define the mutual information of a pair of partitions P_1 and P_2 of S by the relation:

$$I(P_1,P_2)=H(P_1)+H(P_2)-H(P_1,P_2).$$

The following properties hold for the mutual information:

- Symmetry : $I(P_1,P_2)=I(P_2,P_1)$
- If P_1 and P_2 are statistically independent: $I(P_1,P_2)=0$
- $I(P,P_{\min})=0$
- $I(P,P_{\max})=H(P)$

V. PARTITIONS COMPARISON DISTANCE

A lot of criteria for comparing partitions can be found in the clustering and the segmentation literature (cf. [2] and [4]). Here we present a criterion based on the information theory paradigm. As a premise we adopt a communication framework in which we code one partition knowing the other one. Such a transmission would cost $H(P_1|P_2)$ bits but this quantity, which reflects yet the discrepancies between the partitions, is not symmetric and thus unsatisfactory. Therefore we build a symmetric quantity by adding the inverse conditional entropy:

$$\Delta(P_1,P_2)=H(P_1|P_2)+H(P_2|P_1)$$

This quantity can be viewed as the cost of coding both partitions in a full duplex transmission schema. It is shown in [3] that it is a metric on the set of partitions of S. We can also write the following equivalent formulations:

$$\Delta(P_1,P_2)=2H(P_1,P_2)-H(P_1)-H(P_2)$$

$$\Delta(P_1,P_2)=H(P_1)+H(P_2)-2I(P_1,P_2)$$

$$\Delta(P_1,P_2)=H(P_1,P_2)-I(P_1,P_2)$$

The last equation shows this distance as a dissimilarity measure which measures the amount of unshared information; interestingly mutual information has the opposite meaning and even if used as a similarity measure, it is not a metric.

The following properties hold for the distance Δ:

- If P_1 and P_2 are statistically independent: $\Delta(P_1,P_2)=H(P_1)+H(P_2)$
- $\Delta(P,P_{\min})=H(P)$
- $\Delta(P,P_{\max})=\log(|S|)-H(P)$

VI. SEGMENTATIONS COMPARISON

Using the distance defined in the previous section, we are now able to compare 2 segmentations by considering them as partitions only. In order to present the results we choose a specific segmentation algorithm, the watershed algorithm which has two interesting characteristics: its result is a fusion tree in which nodes correspond to regions, and the selected segmentation within the tree depends on a single scalar parameter called the flood level. By exploring the fusion tree with a flood level going from 100% to 0%, we are thus able to generate a family of partitions which are progressively refined. We make our tests with the ADAM dataset. This dataset is made of 39 SPOT multispectral images taken over Fundulea (Romania) during a period of 10 months. These images have been calibrated, registered and constitute a high quality image time series. The segmentation chosen as reference is extracted at flood level 7% from the watershed segmentation of a chip taken in one of these images (cf. fig 3). The test consists in computing the distance between this partition and all the partitions which can be computed from the fusion tree of the watershed segmentation of another image. We make the test with 4 different images and present the curves in fig. 4.

Figure 3. SPOT image reference excerpt (left); and its segmentation obtained with a 7% flood level in the watershed algorithm (right)

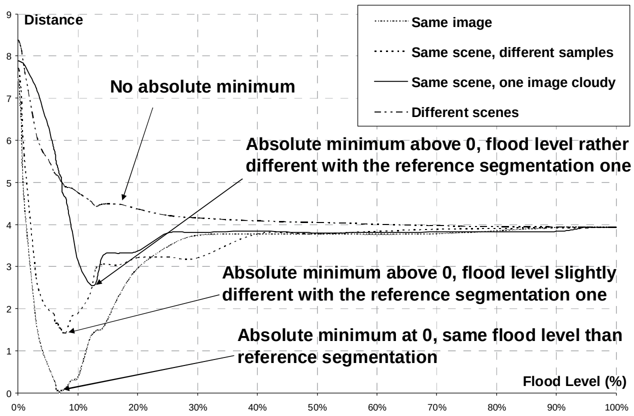
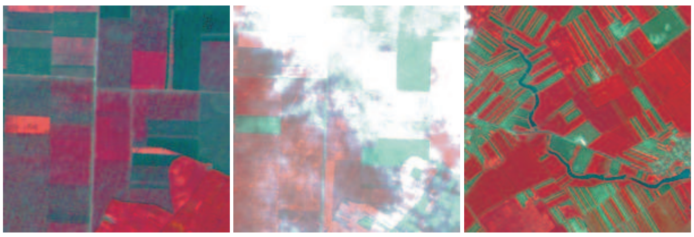
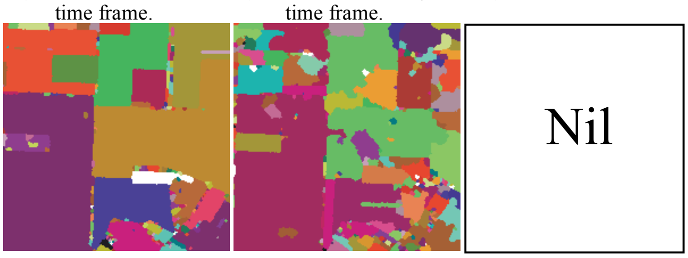


Figure 4. Δ distance between a reference segmentation (right of Fig. 3) and every segmentation within the fusion tree provided by the watershed algorithm for 4 images (left of Fig. 3 and images a), b) and c) of Fig.5)

In the first case we take the same image as the reference one. As foreseen, the distance curve exhibits a global minimum which reaches 0 and which is located exactly at a 7% flood level; the partition in the fusion tree closest to the reference one is thus this partition itself. In the second case we take a different image over the same area (fig. 5a). The distance curve exhibits a global minimum which is higher than 0 and located slightly aside the 7% flood level; the corresponding segmentation is shown in fig 5d. In the third case we take a very cloudy image over the same area (fig 5b). the distance curve continues to exhibit a global minimum, even if its value is much higher than 0 and if it is far away from 7% flood level; the corresponding segmentation is shown in fig 5e). Finally we take an image representing a completely different scene (fig. 5c). The distance curve does not show any global minimum and we conclude that there is no closest partition to the reference one in the fusion tree of this image.



a) Same scene, other time frame. b) Same scene, cloudy time frame. c) Different scene.



d) e) f)

Figure 5. First row shows test images used in segmentations comparison. Second row shows the selected partition in the fusion tree of the watershed segmentation of the upper image, closest to the reference partition.

Figure 6. Composite segmentation of the whole time series.

VII. 2D+T Time series segmentation

Starting with the vector segmentation of the image time series (cf. fig 1), we can select for each image the time-localized segmentation which is closest to this vector segmentation. Among these segmentations we choose the one which is closest to the vector segmentation and adjust the other time-localized segmentations by propagating this selected segmentation along the time series and selecting the segmentations which are consecutively closest to each other. Then we can select in the vector segmentation fusion tree the level which minimizes the sum of the distance between this vector segmentation and all the adjusted time-localized segmentations. By iterating the algorithm, we are thus able to build a composite segmentation (cf. fig 6) which is not trivial while simultaneously closest to the individual segmentations.

VIII. Conclusions

With the goal of comparing segmentations, we have shown that we can consider them as partitions only. This view allows us to adopt a partitions comparison framework in which we introduce an information theoretic comparison metric which has a clear meaning and a very interesting behavior. Equipped with this tool we are then able to compare segmentations and to adjust different segmentations coming from different images in an image time series. The result is a reference segmentation for the whole time series which constitutes a first step towards a 2D+T segmentation of such image time series.

Acknowledgment

Author would like to thank Yann Le Meur from ENSIEG who conducted the experiments on time series segmentation; also, he would like to stress the fruitful discussions with Jordi Inglada from CNES during which the concept of the segmentations comparison distance emerged.

References

[1] Y. Le Meur, "Segmentation multi-temporelle d'une séquence d'images SPOT", MsC thesis, unpublished, september 2004.

[2] M. Meila, "Comparing clusterings – An axiomatic view" in Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany, 2005.

[3] M. Meila, "Comparing clusterings by the variation of information" in Proceedings of the 16th Annual Conference of Computational Learning Theory (COLT), Springer, 2003.

[4] A. Almudevar, C. Field, "Estimation of single generation sibling relationships based on dna markers", J. Agricult., Biol., Environ. Stat., vol.. 4, pp. 136-165, 1999.