ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

Hierarchical and multiscale Mean Shift segmentation of population grid

Johanna Baro[1,2], Etienne Come[1], Patrice Aknin[1] and Olivier Bonin[2]

1- Université Paris-Est, IFSTTAR, GRETTIA, F-77447 Marne-la-Vallée, France

2- Université Paris-Est, IFSTTAR, LVMT, F-77447 Marne-la-Vallée, France

Abstract. The Mean Shift (MS) algorithm allows to identify clusters that are catchment areas of modes of a probability density function (pdf). We propose to use a multiscale and hierarchical implementation of the algorithm to process grid data of population and identify automatically urban centers and their dependant sub-centers through scales. The multi-scale structure is obtained by increasing iteratively the bandwidth of the kernel used to define the pdf on which the MS algorithm works. This will induce a hierarchical structure over clusters since modes will merge together when the bandwidth parameter increases.

1 Introduction

Since 2011, the French National Institute on Statistics and Economic Studies (INSEE) produces population density datasets at the national level on a regular 1-km^2 grid. These datasets are interesting to analyse urban areas in a new way because they are independent from all administrative divisions of the territory. Our goal here is to apply a hierarchical and multiscale segmentation approach to identify clusters of urban population, and to retrieve their spatial hierarchical structure composed of urban centers and sub-centers.

Here, we will use a multiscale and hierarchical implementation of the mean shift (MS) algorithm, considering our grid as an image. This algorithm follows an iterative procedure that defines trajectories starting from each point of a dataset and leading to its corresponding mode on the probability density function (pdf) of the data. Hence it can be used in a segmentation task where clusters consist of groups of points associated to the same mode i.e. cluster prototype.

2 Mean Shift Algorithm

The mean shift procedure can be applied to any kind of dataset as it is generally defined over any d-dimensional euclidean space $\mathbf{R}^d$, although sometime it is defined over $\mathbf{R}^d$ associated to a weight function $w : \mathbf{R}^d \to (0; \infty)$, as introduce by Cheng in [1]. This later formulation allows to work with binned data [2] where sets of elementary measurements are summarized by a single position (which corresponds to the center of the bin) associated to a weight. This setting corresponds to our image segmentation framework. And, in what follows, we therefore consider that a population grid is assimilated to an image with n pixels where $S = \{s_1, s_2..., s_n\}$ is the set of coordinates (s_{i_1}, s_{i_2}) on the image,

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

and $W = \{w_1, w_2, ..., w_n\}$ is the set of weights associated to S which correspond to the population per pixel.

2.1 Non-parametric density estimation

Application of MS algorithm in an image segmentation task rests upon the estimation of the 2-D pdf. The pdf of the data is estimated by the Parzen-Rosenblatt kernel estimation method [3]. In practice we used a weighted estimator of the pdf defined by :

$$\hat{f}_h(x) = \frac{1}{h^2 \sum_i w_i} \sum_{i=1}^{n} w_i K\left(\frac{x - s_i}{h}\right), \quad (1)$$

where $h > 0$ is the bandwidth parameter acting like a smoothing or scaling parameter, and K is a symmetric kernel that integrates to one. If h is small, the initial pdf estimation associates as many modes as pixels on the image. Here we use the gaussian kernel given by :

$$K(x) = (2\pi)^{-1} \exp\left(-\frac{1}{2}\|x\|^2\right) \quad (2)$$

2.2 Clustering algorithm

The principal function used in the algorithm is the *sample mean* defined in [1]:

$$m(x) = \frac{\sum_{s \in S} K(s - x)w(s)s}{\sum_{s \in S} K(s - x)w(s)}, \qquad \forall x \in \mathsf{R}^2 \quad (3)$$

The *sample mean* can be considered as the local mean of S around any x in R^2. The MS algorithm allows to find the mode on the pdf associated to each point $t \in T$, where T is a subset such as $T \subseteq S$. The mode, located at a point where the gradient $\Delta\hat{f}$ cancelled, is found by iteratively updating the position $m_t = m(m_t)$. It starts from $m_t = t$ and carries on until convergence to a stationary point, knowing that convergence can be monitored and iterations stopped if $m(m_t) - m_t < \alpha$ (a measurement of accuracy).

The difference $m(m_t) - m_t$, also called the *mean shift*, characterizes the movement of a point t near the *sample mean*. As showed in [1] and [4], regardless of the choice of kernel the algorithm consists of an gradient ascent method with an adaptive step linked to the gradient magnitude.

For the segmentation of an image we apply the following algorithm:

1. Input: $T = \{t \in S; w(t) > 0\}$ the cluster centers (initialization without the null regions of the image); h the bandwidth
2. For all $t \in T$, apply the MS procedure to find their modes stored in M
3. If m_i and m_j in M are separated by a distance smaller than a fixed value
   (a) Merge them into one cluster center in a new set T'
   (b) Assign them the cluster label c_k in P
   Else create : two clusters centers in T', two cluster labels c_k and c'_k in P
4. Output: T' the cluster centers, P the matrix of cluster labels

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

The result of this algorithm is that each pixel in the initial set T is assigned to a cluster in P, which represents the catchment area of a mode.

MS segmentation can be computationally expensive therefore in practice the pdf is not explicitly computed during the procedure, only local moves are. At every step of the track instead of using all the set S in the calculation of the sample mean, we consider a subset of S in which we removed negligible elements, that are the ones with values close to zero in the kernel, depending on the distance to the previously known m_t. Moreover tracking of each mode is independent from one another, therefore parallelization is possible to reduce computation time.

3 Multiscale and hierarchical implementation

Since the MS procedure relies on the kernel estimation of pdf, it is possible to consider a multiscale approach depending on the bandwidth of the Gaussian kernel h. The scale-space framework provides the basis for the development of multiscale and hierarchical implementation, by handling image feature points and their evolution through scales. Lakemond presents in [5] some properties of image features (such as modes in our application) in scale-space among which the two following are crucial to perform multiscale and hierarchical clustering:

- Modes may merge as the scale increases,
- Gaussian kernel ensures that a mode cannot diverge and split into multiple modes as the scale increases.

These properties allow to consider hierarchical adaptation of the MS algorithm such as in [6]. The idea is first to apply the MS algorithm on the set of all pixels to cluster, starting with a small value of h. Then the algorithm is called back iteratively only on the update set of cluster centers with an increasing value h. This way, we obtain a hierarchical tree of clusters.

Let T be the initial set of pixels. (P_i, T_i, h_i) is a triplet associated to the level i of the hierarchical tree where P_i is the segmentation of T at level i, T_i is the set of cluster centers and h_i is the bandwidth at level i.

1. Initialization : apply the MS algorithm (section 2.2) on T, with $h = h_0$ in order to compute the first triplet (P_0, T_0, h_0).
2. While $card(T_i) \neq 1$
   (a) $h_{i+1} = a * h_i$ (or $h_{i+1} = a + h_i$), where a is a constant
   (b) Apply the MS algorithm using (T_i, h_{i+1}) to obtain P_{i+1} and T_{i+1}.

Knowing that the algorithm starts with a small bandwidth and carries on until there is only one remaining mode (and so one cluster), it can provides numerous segmentation levels among which it is not trivial to retrieve significant structures in the dataset. One way to identify pertinent structures is to refer to the lifetime of modes (or of clusters) when the tree branches are scanned vertically as suggested in [7]. Lifetime is the range of scales over which the mode and its cluster remain stable (no merging occurs). "True" clusters in the dataset will be the ones associated to stable modes over a large range of scales.

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

4 Results and comparison

4.1 Application on a regional population grid

The dataset considered is a grid containing population count per 1-km^2 cell, covering mainland France. The grid is converted into an image and split in sub-images in order to treat urban structures at a regional level. The experiment exposed here is designed with ($h_0 = 1.1,\ h_{i+1} = 1.05 * h_i$).

Reminding that our goal is to identify urban areas, it appeared necessary to filter the grid so as to differentiate urban from rural areas. Indeed if we want to build a hierarchical structure of urban areas, we must first eliminate rural areas to ensure they are not merged with urban clusters during the construction of the hierarchy. To do so, we first applied the MS algorithm with a fixed and small $h = 1.1$ in order to remove clusters that did not contain enough population, by using a threshold on population per cluster. In the following example the threshold was fixed at 2000 pers./cluster.

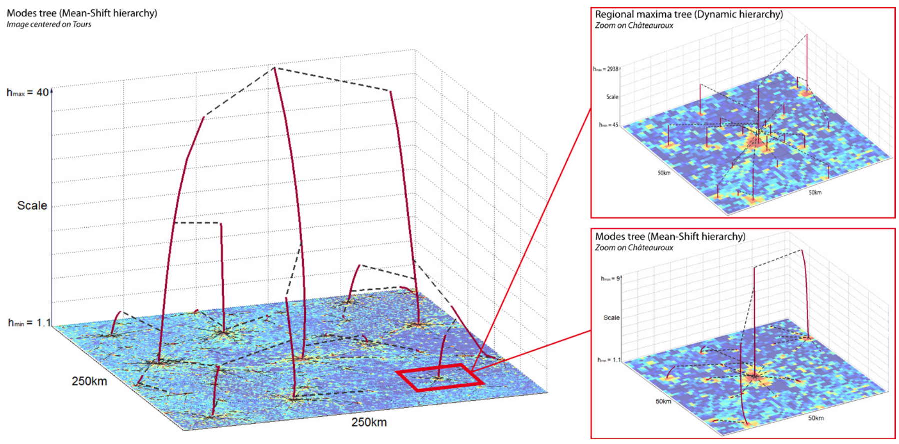

Fig. 1: Hierarchical trees of mode: (left) tree of the MS segmentation, (right) comparison of dynamic tree (top) and the MS tree (bottom) on a zoomed image

The left image on figure 1 shows the 3-D tree of modes placed on top of the original image of population density, centered on the city of Tours. It presents mode trajectories (and so mode lifetimes) through the space of the image and through scales. The stabler modes remaining in the bigger scales are the ones associated to the major agglomerations of the region, after they absorbed less stable modes that are the small population centers into their neighbourhood. Figure 2 shows segmentations for different scales. On the selected scales we observed long lifetimes on all the branches, expressing some kind of stability associated to what one would assumes true structures. It shows different levels of urban organisation, for instance core centers of urban areas appear in small scale (figure 2.b), knowing that unenclosed pixels on the image belong to the removed rural areas. Their clusters extend to include suburbs and periurban areas in

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

figure 2.d. Figure 2.f shows another type of spatial organisation consisting of big catchment areas where the hierarchy is more informative in itself than the segmentation in one particular scale.

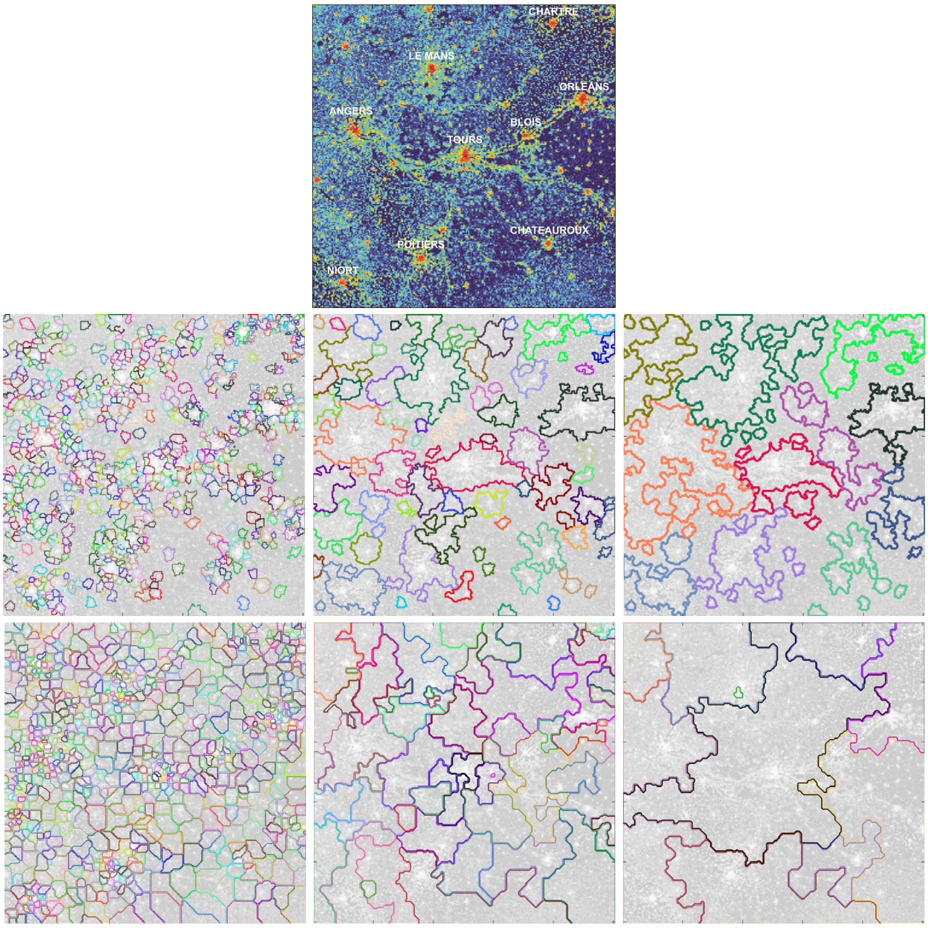

Fig. 2: Row 1 : Original population image, Row 2 : MS segmentations for $h = \{1.1; 5.5; 16\}$, Row 3 : Watershed segmentations for $dynamic = \{358; 1820; 5225\}$

4.2 Comparison to a hierarchical watershed segmentation

Vachier introduces in [8] a scale-space paradigm designed with morphological filters, allowing to compute hierarchical watershed segmentation. In this approach the set of markers used to control the segmentation are the regional maxima. The hierarchy is build according to the progressive merge of the clusters of the regional maxima, valued by their dynamic. Dynamic measures the contrast of the extrema on an image. In the scale-space, dynamic is the value where the extrema disappear on simplified images computed with opening by reconstruction.

Although the scales are not in the same unit and so directly comparable,

ESANN 2013 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2013, i6doc.com publ., ISBN 978-2-87419-081-0. Available from http://www.i6doc.com/en/livre/?GCOI=28001100131010.

we present comparisons based on the same number of modes/regional maxima at a given scale. The right side of figure 1 compares trees obtained by the MS approach and the morphological approach (rural areas have been filtered here too). The merges occurring in the trees differ according to the approach in the lower to the middle scales, but the top of the hierarchies appear similar as strong modes correspond to strong regional maxima. Row 2 of figure 2 shows that segmentation results are particularly different no matter the scale. Watershed segmentation is performed on the entire image because common boundaries are needed to merge the clusters in the hierarchical process. Hence boundaries include rural areas even if their regional maxima are not taken into account in the segmentation process. In the end the drawback of the watershed approach is that the data must include only urban areas, otherwise present rural areas could deform the identified urban centers boundaries.

5 Conclusion and future work

In this paper we presented a multiscale and hierarchical implementation of the mean shift algorithm, which we used to perform segmentation task. We applied it on a population data grid of 1-km^2 resolution in order to identify urban clusters and their hierarchy. This segmentation approach that can be apply to any kind of dataset suits us more than other image segmentation approach because modes and their catchment areas are readily interpretable and because the pixels to cluster can be the all image or a subset of it.

A crucial aspect of this multiscale approach is to determine which is or which are the interesting scales to cut the tree. In this paper we only focus on horizontal cuts of the tree corresponding to constant bandwidth on all the image. But it would also be interesting to study no standard cuts through scales. In that way, the spatial variation of scales of interested may be preserved.

References

[1] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):790–799, 1995.

[2] D. W. Scott and S. J. Sheatherb. Kernel density estimation with binned data. *Communications in Statistics - Theory and Methods*, 14(6):1353–1359, 1985.

[3] E. Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.

[4] D. Comaniciu. Mean-shift : A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

[5] R. Lakemond, D.N. McKinnon, C. Fookes, and S. Sridharan. A feature clustering algorithm for scale-space analysis of image structures. In *Proceedings of the International Conference on Signal Processing and Communication Systems ICSPCS2007*, pages 186–192, 2007.

[6] D. DeMenthon and R. Megret. Spatio-temporal segmentation of video by hierarchical mean shift analysis. Technical report, University of Maryland, INSA Lyon, 2002.

[7] S.J. Roberts. Parametric and non-parametric unsupervised cluster analysis. *Pattern Recognition*, 30(2):261–272, 1997.

[8] C. Vachier. Morphological scale-space analysis and feature extraction. In *Image Processing, 2001. Proceedings. 2001 International Conference on*, volume 3, pages 676–679, 2001.