STAG: Smart Tools and Apps in computer Graphics (2016)
Giovanni Pintore and Filippo Stanco (Editors)

Tracking the evolution of rainfall precipitation fields using persistent maxima

S. Biasotti[1], A. Cerri[1], S. Pittaluga[1], D. Sobrero[1], M. Spagnuolo[1],

[1]Istituto di Matematica Applicata e Tecnologie Informatiche "E. Magenes", CNR, Italy

Abstract

In this paper we propose a novel methodology for tracking the maxima of rainfall precipitation fields, whose changes in time may give interesting insights on the evolution of storm. Our approach is based on a topological analysis of rainfall data allowing for the extraction of the most prominent, and hence meaningful, rainfall field maxima. Then, an ad-hoc bottleneck matching is used to track the evolution of maxima along multiple time instances. The potential of our method is exhibited through a set of experiments carried out on a collection of observed punctual rainfall data and radar measurements provided by Genova municipality and Regione Liguria.

Categories and Subject Descriptors (according to ACM CCS): I.3.6 [Computer Graphics]: Methodology and Techniques—

1. Introduction

Recent catastrophic events caused by flooding rain in Genova and in various areas of the Liguria region (Italy, October 2014) have highlighted once again the importance of computer systems in the analysis of environmental data. More and more digital data are available, which provide an extremely rich, yet difficult to process, amount of information about our environment and its dynamic phenomena. This is the case of observed rainfall data, measured either by meteorological radar or rain gauges distributed over the Liguria territory. The rain gauges measure at regular intervals the amount of rain and provide therefore a close to real-time measures about the ongoing precipitation. Together with wether forecasts, this information is used to monitor critical precipitation events, as one of the many input for alarm forecasting and civil protection plans.

The observed rain data are also stored in time series, which contain valuable knowledge that could concur to a deeper comprehension of storms and their evolution in time. We believe that an effective and automatic method for an efficient analysis of precipitation fields could suggest effective statistical analysis of events, and correlation studies among the evolution of storms and several other relevant data such as terrain morphology, satellite imagery and meteorological situation at the large. While this is the long-term plan of our target application, in this paper we present the results of the first step of the analysis pipeline, which relates to the detection and tracking of precipitation maxima.

In order to understand the evolution in time of precipitation events, it is important to focus on the main features of the associated rainfall fields and their configuration, keeping only what is important and discarding irrelevant details that do not contribute to understand the overall event structure. For this reason, we think it is crucial to adopt a description that captures the important elements of the field, such as its maxima, which have a relevant semantic content and, at the same time, are formally well-defined. Indeed, the maxima of a scalar field are a subset of its critical points and their configuration. Moreover such a concept is related to differential topology thus giving a suitable framework to formalize the problem. From the practical point of view, computational topology techniques provide several tools and measures for data analysis and coding, which can be used in several applications including visualization [TG09, WBP07], understanding [DSNW13, HHC*13, WG09], simplification [GJR*14] and comparison [SWC*08] of data. Extended surveys on these topic can be found in [BDF*08, BDFP07].

In this paper we take advantage of tools offered by computational topology to propose a novel methodology for tracking the maxima of rainfall precipitation fields, whose changing in time may offer insights about the evolution of storms. The main contribution of the proposed approach is twofold. First, we apply topological methods to the analysis of rainfall data, which allow for the extraction of the most prominent, and hence meaningful, rainfall field maxima. Then, we introduce a new bottleneck matching between sets of rainfall field maxima, which is used to track their evolution along multiple time instances. The proposed method is validated by experiments carried out on a collection of heterogeneous observed rainfall data (rain gauges and radar measurements) as provided by Genova municipality and Regione Liguria, and its implementation is integrated in the European Integrating Project IQmulus: A High-volume Fusion and Analysis Platform for Geospatial Point Clouds, Coverages and Volumetric

DOI: 10.2312/stag.20161361

Data Sets (http://www.iqmulus.eu/).

2. Related work

Storm-tracking algorithms are a key ingredient of forecasting systems, as they can provide important information about assessing storm birth, evolution and decay. A storm identifies a prominent precipitation event, and tracking a storm consists in collecting, along time, all the locations spanned by that specific precipitation event: the ideal tracking starts from the first moment in which the storm has been detected till the last time it has been observed. A number of approaches have been proposed to associate locations at a time frame t_{i+1} with storms identified at the time t_i. Many of them first identify regions of interest on radar images, usually characterized by high reflectivity and sufficiently large area, and computes their characteristics such as centroids, area, major/minor radii and orientation. Then, regions are matched across two consecutive time frames, according to the idea that the best candidate for matching minimizes some distance between the considered characteristic [LS09]. For example, the TITAN algorithm discussed in [DW93] combines both centre of mass and area of regions for final decision of tracking. The SCIT algorithm [JMW*98] forecasts the centroid locations of cells at time t_i: regions at time t_{i+1} are then assigned to the closest centroid location within a certain radius. The approach proposed in [HFZ*09] takes inspiration from the TITAN algorithm, but also includes the overlapping of regions across consecutive time samples in the tracking process.

The methods based on a region-oriented tracking strategy work sufficiently well when the morphological characteristics of the land are relatively simple: in our case, the Liguria region instead is characterized by an articulated orography close to the sea, with many small catchment basins that are highly influenced by local maxima of precipitation and that can cause quite critical flash floods. Therefore, a tracking which aims at detecting and following punctual maxima instead of regions appears here to be more appropriate.

In this context, the approach we propose contributes to the current state-of-the-art scenario of storm tracking algorithms by introducing the topological perspective for the analysis of rainfall fields and their local maxima, in order to better understand the structure and the evolution of precipitation events. To the best of our knowledge, this is the first time that topological data analysis is used for this target application, although the topological approach has been used in related scenarios involving geo-spatial data, such as GPS trajectories simplification [KSG] and change detection [Val13].

Also, topological methods have been widely investigated for the visualization and analysis of time-varying scalar fields [BBD*07, CSEM06, EHM*08, FOTT08, WCBP12]. However, most of these techniques mainly focus on localizing and tracking features of interest, sometimes lacking a measure to quantitatively assess the amount of changing in the considered features along time. On the other hand, our method pairs a tracking procedure with a notion of distance that can be used to quantify how much a set of rainfall maxima and its configuration has changed from time to time.

3. Persistent rainfall field maxima

Our goal is to study the evolution of precipitation events along time. To achieve this, the idea is to capture the precipitation field originated by the storm under study at several, sufficiently close time samples, and then to detect meaningful changes in the considered field instances as time moves on.

Our working assumption is that precipitation events at a given time sample are sufficiently well represented by the maxima of the corresponding precipitation field. This is actually part of the information analysed by experts in order to better understand and foresee the evolution in time and space of precipitation events. Hence, our starting goal can be recast into the one of tracking the temporal evolution, in terms of both geographical displacement and rainfall field value, of such maxima.

Note, however, that in general not all maximum points of a precipitation field are useful to characterize meaningful information. For example, local maxima characterized by small rainfall field values usually correspond to non-relevant events; therefore, it makes sense to track their evolution only in case the associated precipitation values become bigger than a threshold depending, e.g., on wind speed and direction or territorial geography.

Also, relevant maxima should be characterized by some notion of *prominence*. For instance, two maxima that are close in both geographical displacement and field values, such as in the case of a small bump occurring in the neighbourhood of a field peak, will be probably talking about the same storm front, the bump being the result of some approximation error or non-relevant fluctuation of the precipitation field.

Motivated by the above remarks, we propose a methodology based on the use of *topological persistence* [ELZ02] for the detection of meaningful rainfall field maxima. Indeed, topological persistence provides a theoretically sound framework to formally introduce the prominence (also called persistence) of rainfall field maxima, and hierarchically organize them according to this notion; in particular, low-valued maxima cannot have large prominence according to persistence. In this way, a persistent-based pruning can be easily induced on the sets of rainfall maxima to simplify data and remove noise, also improving computational efficiency.

3.1. Topological persistence

Topological persistence is at the heart of topological data analysis that deals with the study of global features of data to extract information about the phenomena that data represent. The topological persistence approach is based on computing topological features of data at different scales to see which ones are long-lived and which are short-lived. The basic assumption is that relevant features and structures are the ones that persist longer. These ideas are currently receiving increasing attention from the research community, finding applications in various fields ranging from shape description and comparison [CZCG05, DLL*10, DL12] to data simplification [BLW12] and clustering [CGOS13].

In the classical topological persistence setting, data are usually represented by a topological space X, while its topological exploration is driven by a continuous scalar field $f : X \to \mathbb{R}$. The role of f is to describe some property which is considered relevant for the analysis, in our case a rainfall field at a given time sample.

For the present contribution, we use topological persistence to study the evolution of the connectivity for the superlevel sets $X^u = \{x \in X | f(x) \geq u\}$ for $u \in (-\infty, +\infty)$. To simplify the exposition, assume that the local maxima of f are such that all their values are different. As we sweep u from $+\infty$ to $-\infty$, new connected components are either born, or previously existing ones are merged together. A connected component C is associated with a local maximum $x \in X$ of f, that is, the point of X at which the component is first born. The value $f(x)$ is referred to as the birth time of C. When two components corresponding to local maxima x_1 and x_2, such that $f(x_1) < f(x_2)$, merge together, we say that the component corresponding to x_1 dies. In other words, the component associated with the smaller local maximum is merged into that associated with the larger one.

In this way, it is possible to define a hierarchy of components, and hence of the corresponding local maxima. In particular, each local maximum $x \in X$ of f can be associated with a quite natural notion of prominence: the *f-persistence* $\mathrm{pers}_f(x)$ of x is simply the difference between the birth and the death time of the corresponding connected component. The global maximum of f, which is associated with the eldest component, is considered to have f-persistence equal to $\max f - \min f$.

The added value in using persistence is that it is known to be more stable than other measures of magnitude such as absolute height: to have an intuition of this, it it is sufficient to think of a small bump occurring in the neighbourhood of a high-valued peak of f, which will be characterized by large absolute height but small persistence. In general, persistence is robust to small perturbations of the considered function: assuming for instance that g is a noisy approximation of f, there is a one-to-one mapping of small variation from the prominent local maxima of g to those of f, the remaining ones being associated with topological noise, see Figure 3 for a visual intuition.

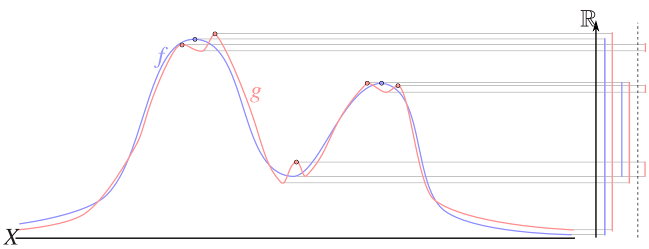

Figure 1: *Two functions $f, g : X \to \mathbb{R}$ and the associated local maxima. On the right, pictorial representation for the persistence of each local maxima. Segments on the right of the dotted line stand for the persistence of topological noise.*

3.2. Implementation

In our implementation, the discrete counterpart of the space X is a triangle mesh M representing the whole Liguria Region. We consider several functions $f_i : V \to \mathbb{R}_{\geq 0}$ defined on the set of vertices V of M and taking values in the set $\mathbb{R}_{\geq 0}$ of non-negative real numbers: each f_i comes from rainfall data at a time t_i, as specified in Section 5.1. In practice, the value $f_i(v)$ represents the cumulated rain at v in the temporal interval $(t_{i-1}, t_i]$. For a function f_i, we process the vertices of M in decreasing values, from $\max f_i$ to $\min f_i$. To compute the local maxima of f_i and their prominence, we use the classical persistence algorithm for $0th$ homology [ELZ02, EH10]. An example of the outcome is given in Figure 2, showing one of the functions f_i and its local maxima at three different persistence levels.

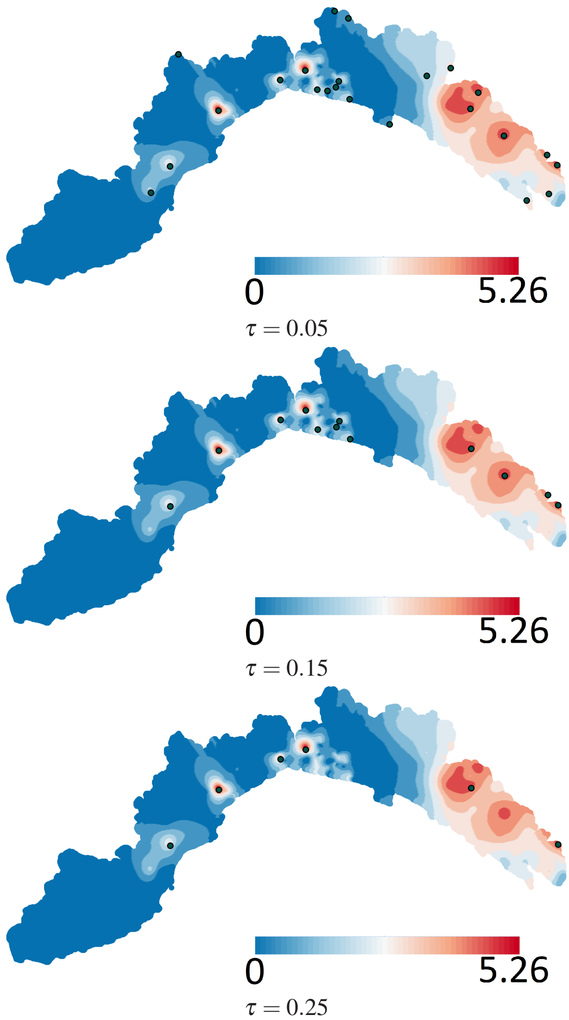

Figure 2: *A function $f_i : V \to \mathbb{R}_{\geq 0}$, colour coded from blue (low) to red (high) values, equipped with the local maxima having persistence greater than $\tau(\max f_i - \min f_i)$.*

Sorting the n vertices of M takes $O(n \log n)$. After that, by using a union-find data structure, the persistence algorithm requires linear storage and running time at most proportional to $m\alpha^{-1}(m)$, with m the number of edges in the mesh, and $\alpha^{-1}(\cdot)$ is the Ackermann function.

4. Tracking persistent rainfall field maxima

In this section we discuss how to match the sets of local maxima of two functions $f, g : V \to \mathbb{R}_{\geq 0}$ representing a rainfall field at two different time instances, and how to derive from the considered matching a measure of (dis)similarity for the two sets. The idea is to compare the two sets by measuring the cost of moving the points associated with one function to those of the other one, with the requirement that the longest of the transportations should be as short as possible. Since, in general, the number of points in the two sets may differ, we also enable points to be "annihilated", paying some cost in terms of the final dissimilarity distance. By assuming that the number of local maxima is finite for both f and g, which is actually the case in our application scenario, our goal can be related to the bottleneck transportation problem [EIK01, GR71], and in particular to the notion of bottleneck distance [CSEH07, VH01].

4.1. Matching rainfall field maxima

Let F, G be the sets of local maxima associated with two rainfall fields f and g, respectively. In order to compare the two sets, we interpret each of their elements as a point of $\mathbb{R}^2 \times \mathbb{R}_{>0}$, with $\mathbb{R}_{>0}$ the set of strictly positive real numbers. In practice, each local maximum is associated with a triplet of coordinates representing its geographical position and the associated persistence, either for f and g. Note that geographical and rainfall measurements have nothing to do with each other, hence a normalization step is needed beforehand, see Section 5.2 for details. After normalization, we thus have $p = (x(p), y(p), \mathrm{pers}_f(p))$ for each maximum $p \in F$; similarly, $q = (x(q), y(q), \mathrm{pers}_g(q))$ for all $q \in G$. We further assume to augment both F and G by adding all points of the plane $xy : z = 0$, still denoting F and G the resulting subsets of $\mathbb{R}^2 \times \mathbb{R}_{\geq 0}$. This last technical requirement allows us to compare the collections of local maxima by making use of the bottleneck distance between F and G, which is defined as:

$$d_B(F,G) = \inf_{\gamma} \sup_{p} d(p, \gamma(p)), \qquad (1)$$

where $p \in F$, γ ranges over all the bijections between F and G, and d is a measure of the distance between p and $\gamma(p)$. A bijection between F and G has three types of point pairs: both off the plane xy, one off xy and the other on that plane, and both on xy. Roughly speaking, the most important type is the first, matching points in terms of their geographical displacement and persistence, and the least important is the last, completing the matching in a way that does not affect the final distance. The remaining type of pairing is used to annihilate local maxima by moving them to xy. In order to make the above reasoning more precise, we have to specify d. For two points $p, q \in \mathbb{R}^2 \times \mathbb{R}_{\geq 0}$, let $\|p - q\|$ be a distance (e.g., the standard Euclidean distance) between p and q. We now consider the following pseudo-distance d on $\mathbb{R}^2 \times \mathbb{R}_{\geq 0}$ to measure the cost of moving p to q:

$$d(p,q) := \min\{\|p - q\|, \max\{\mathrm{pers}_f(p), \mathrm{pers}_g(q)\}\}. \qquad (2)$$

In other words, the pseudo-distance d between two points p and q compares the cost of moving p to q with that of annihilate them by moving both p and q onto the plane xy, and takes the most convenient. Therefore, $d(p,q)$ can be considered a measure of the minimum of the costs of moving p to q along two different paths (i.e. the path that takes p directly to q and the path that passes through the plane xy). This observation easily yields that d is actually a pseudo-distance. We also remark that the inf and the sup in the definition of the bottleneck distance are actually attained; this is quite easy to see under the assumption that the local maxima of F and G are finite in number. In other words, there always exists a matching between the elements of F and G. In what follows, such a matching will be referred to as a *bottleneck matching*.

4.2. Tracking rainfall field maxima

The notion of bottleneck matching that naturally arises from the formulation of the bottleneck distance provides us with a tool to follow the evolution of rainfall field maxima along time. Indeed, the bottleneck matching can be used to pair the local maxima associated with two functions representing a rainfall field at two consecutive time samples, see Figure 3 for an example.

Consider now the functions f_i introduced in Section 3.2. By composing the bottleneck matchings obtained for each pair of consecutive functions, we get a procedure to track the temporal evolution of the rainfall field under examination. In our application scenario this translates, for example, into the following typical situations. Suppose that two functions f_i, f_{i+1} are such that f_{i+1} is obtained through minor variations of f_i. In this case, there is a one-to-one mapping that pairs local maxima with large persistence; the remaining points can be either annihilated or matched each other without affecting the final value of the bottleneck distance. On the other hand it could happen that, passing from f_i to f_{i+1}, new local maxima characterized by a large persistence value appear, representing the birth of new events; similarly, existing maxima might disappear, revealing the death of meaningful events. In both cases, some maxima for either f_i or f_{i+1} have to be annihilated because they have no counterpart among those of the other function, possibly producing the final value of the bottleneck distance.

4.3. Interpreting tracking

The analysis of the above situations can be complemented by considering the resulting value of the bottleneck distance. Indeed, it provides a quantitative insight about the changing in the configuration of local maxima. In particular, a large value for the bottleneck distance can be used as a warning highlighting a brusque variation in the storm evolution. In the first situation, for example, the bottleneck distance might reveal the significance of geographical displacements for some maxima with large persistence; in the second situation, a high value for the bottleneck distance might be associated with a split or merge event. The role of the bottleneck distance would be particularly useful in the latter case, as the bottleneck matching is not conceived, in the current formulation, for dealing with one-to-many or many-to-one pairings, which is actually another possible way to represent split and merge events.

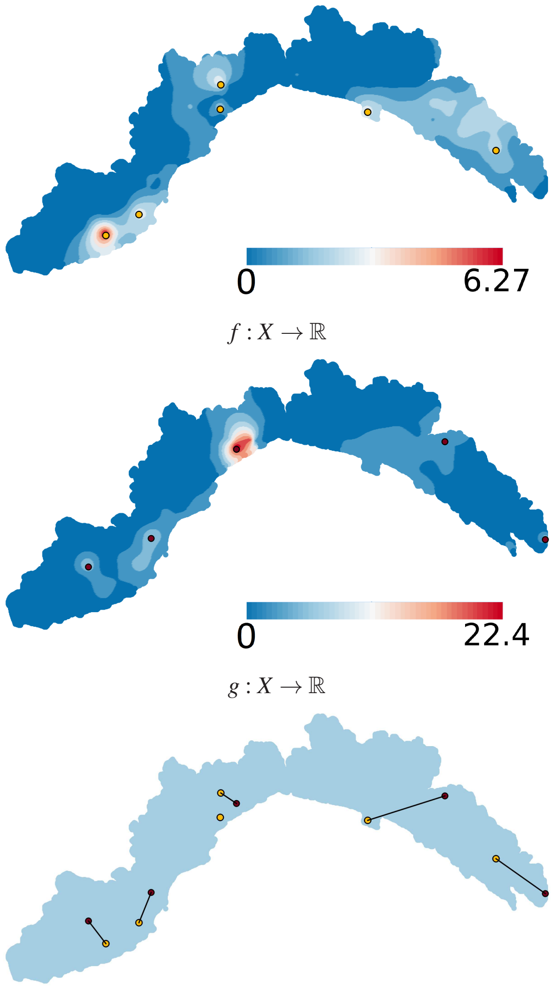

Figure 3: *Two functions and the associated local maxima, colour coded from blue (low) to red (high) values. On the bottom, the bottleneck matching between local maxima.*

4.4. Implementation

Computing the bottleneck distance can be formulated as a classical assignment problem, which can be usually handled by either following a pure graph-theoretic approach, or taking advantage of some geometric additional information possibly characterizing the assignment problem. The latter solution is generally more performing, achieving peaks of computational efficiency in case the points to be compared are in $\mathbb{R}^2$ and the metric underlying the bottleneck distance (i.e., d in our notation) is the Euclidean or the L_∞ one [EIK01]. However, this is not our case, as we consider points in $\mathbb{R}^3$, and d is actually a pseudo-distance. Therefore, we opt for a graph-theoretic approach, which is independent of any geometric constraint.

Our implementation is based on the push-relabel maximum flow algorithm [CG97]. For two sets of local maxima F and G, our algorithm simulates the aforementioned augmenting procedure, giving the value of the bottleneck distance and the pair of points realizing it as output. Note that not both points are necessarily local maxima: indeed, the bottleneck distance might come as the result of moving one local maximum onto the plane xy. To obtain the bottleneck matching, the selected local maxima are removed from the initial sets, and a new iteration of the algorithm is run: the process ends when one of the two sets is empty.

For each iteration, the algorithm runs in $O(n^{2.5})$, being n the number of local maxima involved in the comparison. Note, however, that the computational complexity is definitely not an issue in our application scenario, because the number of storms fronts to be tracked is very limited, usually no more than a dozen.

5. Experimental results

The case study on which we have tested our framework is defined by a collection of observed punctual rainfall and radar data covering the area of interest of the study. Measurements are organized in a number of time steps: for each time step, data are pre-processed and interpolated over the whole domain. The resulting precipitation field is sampled at the vertices of a triangle mesh, representing the Liguria region. Our procedure to track the temporal evolution of precipitation events is summarized as follows:

- for each time step, extract the most persistent local maxima of the corresponding precipitation field, according to a persistence threshold specified by the user;
- for each pair of consecutive time steps, compute the bottleneck matching to pair the persistent local maxima of the corresponding fields;
- chain all the computed bottleneck matchings to define the final tracking of persistent local maxima.

Before discussing the results obtained, we provide details on the dataset and the above procedural steps.

5.1. The dataset

We have selected two different precipitation events. The first one occurred on September 29, 2013, and was characterized by light rain over Liguria with 2 different thunderstorms that caused local flooding and landslides. The peculiarity of this event is that the two thunderstorms were qualitatively different, adding variability to the benchmark: the first thunderstorm translated linearly from south-west to north-east, while the second one was characterized by a first stationary phase before translating as well from west to east. Due to the morphology of the Ligurian territory, thunderstorms belonging to the latter category are particularly dangerous: one of them caused the catastrophic floods in various areas of Liguria in October 2014. The second event occurred between the 16th and the 20th of January, 2014, and is related to an Atlantic low pressure area. This kind of events often produces a secondary low pressure area, known as Genoa Low, over the Ligurian Sea. The depression was responsible of heavy rain for about five days over all the region.

The dataset are gathered from different devices, namely rain gauges and weather radar. The rain gauges networks are maintained by Regione Ligura and Genova municipality, and are deployed with

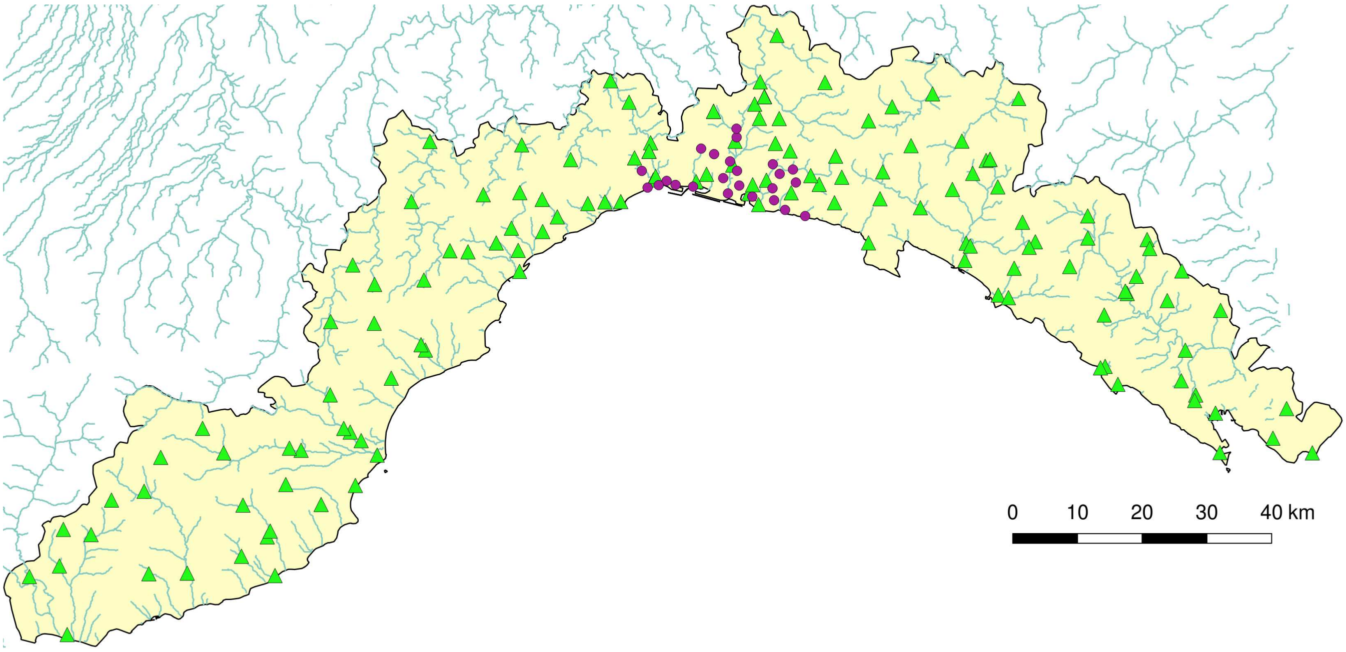

Figure 4: *Spatial distribution of the rain gauge networks of Regione Liguria (green triangles) and Genova municipality (purple circles).*

a different spatial distribution, see Figure 4. The network developed by Regione Liguria is spread over the whole region, with 143 measure stations. The measure system deployed by Genova municipality is entirely located within the city boundary, with 25 measure stations. The raw radar acquisitions come at first as reflectivity measurements with a range of 400 km. The frequency of mountains over the whole Ligurian territory affects the quality of radar acquisitions and a pre-processing step is needed to remove ground clutter effects; processed data are then combined with observations gathered from rain gauges, which are more reliable measurements but do not cover the whole region.

Since the temporal interval is different for each acquisition device, rainfall measurements have been cumulated. In this study, a 10 minutes cumulative step has been used for the more dynamic event of September 29, 2013, for a total amount of 144 time samples; for the event of January 2014, which is more stationary, measurements have been cumulated every 30 minutes (240 time samples).

For each time step t_i, the rainfall field f_i is obtained by interpolating both rainfall and pre-processed radar measurements on a regular grid by means of *ordinary kriging*, a point estimator algorithm in the best linear unbiased estimator family. The estimate is a linear combination of the available measurements; it tries to be unbiased by having the residual mean equal to zero to minimize the residual error. The estimate at a point v is expressed as $f_i(v) = \sum_{j=1}^{n} w_j f_i(v_j)$, where $f_i(v)$ is the estimated value at position v, $\{f_i(v_j)\}_{j=1}^{n}$ are the known samples (i.e. the rainfall measurements cumulated to the time step t_i at the point v_j) and $\{w_j\}_{j=1}^{n}$ are the corresponding weights. These weights are computed as $\mathbf{C}^{-1}\mathbf{D}$, where $\mathbf{C}$, $\mathbf{D}$ are covariance matrices calculated (i) among all the input points and (ii) among the points to be calculated and all the known data, respectively.

Each interpolated rainfall field f_i is finally re-sampled on a digital terrain model (DTM), represented as a triangle mesh. For our analysis, we consider the SRTM (Shuttle Radar Topography Mission [FRC*07]) DTM available in public domain at the URL `http://www2.jpl.nasa.gov/srtm/`.

5.2. The practical setting

For each field f_i, we extract its most relevant local maxima. In doing this, points are selected according to a threshold ε fixed a priori and chosen by the user. In practice, a maximum is considered relevant only if its persistence is larger than ε. In this way, it is possible to filter out local maxima associated with non-relevant information, namely minor variations of the considered rainfall field, as well as approximation errors. In our experiments, we have considered $\varepsilon = \varepsilon(i) = \tau \cdot (\max f_i - \min f_i)$, varying τ to check the impact of the threshold on the quality of tracking results, see Section 5.3 for details. In particular, for $\tau = 0.35$, data has been manually annotated by a geologist with experience in the analysis of precipitation events, finding a collection of meaningful tracks (some of them are displayed in Figure 5). Following [LS09], a track is considered meaningful only if it is given by the composition of at least two bottleneck matchings. This collection of paths has then been used as a ground truth to evaluate the results of our tracking procedure, which has been implemented by taking into account the following remarks depending on the specific real-world data under examination:

- geographic coordinates and rainfall field measurements come

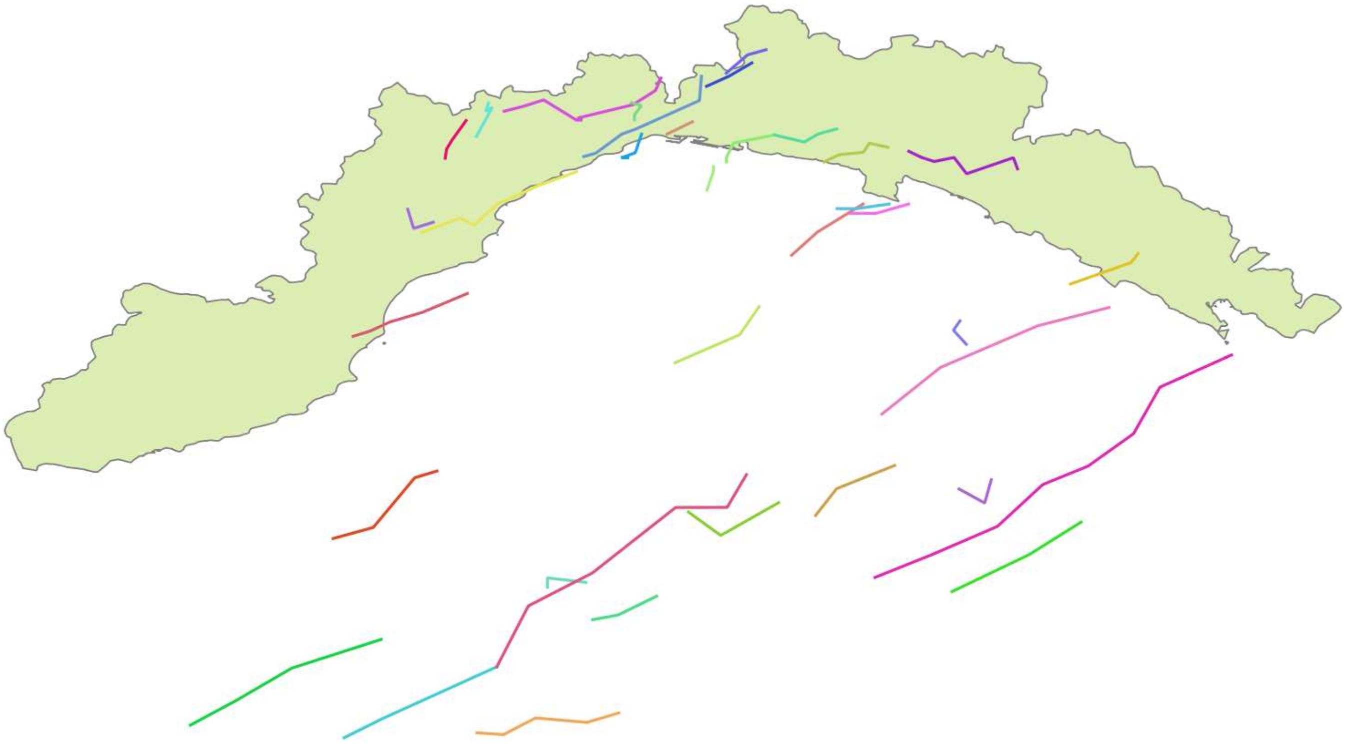

Figure 5: *Some of the validated tracks for the considered precipitation events.*

with different reference frames and at different scales: the former are expressed in (millions of) meters, while the latter in (tens of) millimetres for each time sample. Hence, for two sets of local maxima to be matched, points are first normalized to range in the interval [0,1], processed by computing the bottleneck matching and distance, and then projected back in the original reference frames for the final visualization;

- The high presence of mountains over the whole Ligurian territory penalises large geographical displacements of precipitation events in a short time. To put this information in our model, we compare local maxima by emphasizing their geographic proximity. For two local maxima p,q to be compared, we denote by $\|p-q\|$ the weighted combination of their Euclidean distance restricted to the geographic coordinates, say $\|p-q\|_{\text{geo}}$, and the absolute difference of their persistence, denoted by $\|p-q\|_{\text{pers}}$. Hence we have

$$\|p-q\| = \alpha\|p-q\|_{\text{geo}} + \beta\|p-q\|_{\text{pers}}. \tag{3}$$

The pseudo-distance d in (2) is then evaluated by considering $\|\cdot\|$ as in (3); the computation of the bottleneck distance d_B in (1) and the associated bottleneck matching is updated accordingly. To emphasize the contribution of geographic proximity, in our experiments we set $\alpha > \beta$, varying α and β to test the impact of weights on final results, see Section 5.3 for details;

- while pairing two local maxima p and q through a bottleneck matching, the pseudo-distance $d(p,q)$ provides an additional hint about the nature of those points. If $d(p,q) = \|p-q\|$, by equation (2) it follows that it is more convenient to directly match p and q: our interpretation is that the two local maxima are strongly related, that is, one point is the temporal evolution of the other. In this case, the pairing (p,q) is used to compose the final tracking of local maxima. On the other hand, having $d(p,q) = \max\{\text{pers}(p), \text{pers}(q)\}$ is equivalent to annihilating both p and q. This might occur because the two local maxima represent either unrelated events or non-relevant information. In both cases, the pairing (p,q) is not included in the final tracking. However, one of the causes for the annihilation procedure, namely non-relevant information to be handled, is in part achieved by filtering out maxima trough the persistence threshold ε. Motivated by this, we relax the annihilation process induced by d by mitigating the contribution of $\|p-q\|$ in (2), and assume $\alpha+\beta < 1$ rather than $= 1$.

5.3. Results

Results are shown in Tables 1 and 2. A first evaluation consists in analysing the performance of our tracking method according to different parameters settings, namely varying the persistence threshold τ and the weights α and β used to balance the contributions of geographic and rainfall information. In doing this, we have considered the following evaluation measures adapted from [YMV07]:

- **Detected Tracks (DT):** A track in the ground truth is considered to have been correctly detected if it is overlapped by a track retrieved by our system for at least one third of its segments, that is, matchings. The final TD score is given by the ratio between

Table 1: *Table of results for the DT, TCD and MT evaluation measures, according to different choices of the parameters τ, α and β. Best results are in bold text.*

DT (%) **TCD (%)** **MT (%)**	$\alpha = 0.3$ $\beta = 0.3$	$\alpha = 0.4$ $\beta = 0.2$	$\alpha = 0.5$ $\beta = 0.1$
	59,46	62,16	54,05
$\tau = 0.25$	27,03	27,03	24,32
	40,54	37,84	45,95
	51,35	**70,27**	62,16
$\tau = 0.35$	21,61	**27,03**	24,32
	48,65	**29,73**	37,84

Table 2: *Table of results for the three evaluation measures.*

	Duration	δ-rain	Distance
Track 1	0	0	0
Track 3	0.79	0.29	0.95
Track 4	0	0	0
Track 12	0	0	0
Track 14	0	0	0
Track 16	0.20	0.19	0.35
Track 22	0.25	0.02	0.35
Track 26	0.25	0.10	0.09
Track 29	0.86	0.13	0.89
Track 34	0.67	0.13	0.86

the number of correctly detected tracks and the total number of tracks in the ground truth;

- **Tracks completely detected (TCD):** A track in the ground truth is considered to have been completely detected if it is overlapped by a retrieved track for all its segments. Again, the final TCD score is the number of completely detected tracks normalized by the total number of tracks in the ground truth;
- **Missed tracks (MT):** These are the tracks in the ground truth whose overlapping with a retrieved track involves less than one third of their segments. The total number of missed tracks is finally normalized by the total number of tracks in the ground truth.

The above evaluation is summarized in Table 1 for different values of the persistence threshold τ and the weights α and β. All results are in percentage values.

Perhaps not surprisingly, the best results in terms of DT, TCD, and MT scores are achieved for the highest persistence threshold τ. Indeed, in this case a larger number of non-relevant events are thrown away, ensuring a more reliable tracking procedure. In particular, looking at the corresponding values for weights α and β, setting $\alpha = 0.4$ and $\beta = 0.2$ provides the best parameter configuration in terms of DT, TCD, and MT scores. Hence, for this specific setting, we have refined the evaluation of results by comparing the obtained tracks with those provided by manual annotation as detailed in what follows.

For those tracks that have been completely detected (10 in total), we first look for a counterpart among those retrieved by our system, and select the one sharing the highest number of segments, that is, matchings. Then, the comparison between the two tracks is performed according to the following evaluation measures, adapted from [LS09] to cope with our case study:

- **duration:** this is the difference between the number of segments in the retrieved and the target track, normalized by the number of segments of the longest one;
- **mean rain difference (δ-rain):** this is the absolute difference between the mean values of the cumulated rain along the retrieved and the target track, normalized by the largest of the two values;
- **distance:** this is a measure of how much the retrieved track is far from being exactly the target one. We simply sum the lengths (in the Euclidean norm) of all segments belonging to the retrieved track that are not in the target one. This value is then normalized by the length of the retrieved track.

All measures ranges in [0,1), with 0 the optimal value. A positive score means overestimation: in other words, in this case the target track is completely covered by a longer retrieved one. Related results for the collection of target tracks are reported in Table 2.

As a general comment, we can say that tracking the precipitation events associated with stationary thunderstorms has revealed to be definitely more difficult than for the more dynamic ones: the reason can be found in the fact that stationary events produce across time bunches of local maxima that are close to each other, and that can hardly be matched correctly by only relying on geometric information, which is actually our case.

6. Conclusions

In this paper we have presented a novel methodology for the effective tracking of rainfall field maxima along time. A persistence-based approach for the detection of the most meaningful local maxima has been complemented with the introduction of an ad-hoc bottleneck matching to track the evolution of maxima across different time instances. In spite of the encouraging results obtained on real-world data provided by Regione Liguria and the municipality of Genova, there is still a long road ahead. In this respect, the most promising research directions include to feed our system with additional multi-modal measurements complementing the purely geometric ones. We refer in particular to refine the matching process by including information about wind speed or territorial morphology, which may put constraints on the practical displacement of local maxima. Also, cross-correlation analysis could be considered, in order to somehow exploit the information about the already assigned matchings: indeed, in the current implementation the time history of a track is not considered in order to compute the subsequent matchings. Finally, it would be interesting to extend the notion of bottleneck matching to admit one-to many and many-to one pairings, to improve the detection and tracking of merge and split events.

Acknowledgements Work developed in the CNR research activity ICT.P10.009 (check!!!), and partially supported by the FP7 Integrated Project IQmulus, FP7-ICT-2011-318787, (2012-2016). The rainfall data sets are courtesy of ARPAL, Regione Liguria, and of the Genova Municipality.

References

[BBD*07] BREMER P.-T., BRINGA E. M., DUCHAINEAU M. A., GYULASSY A. G., LANEY D., MASCARENHAS A., PASCUCCI V.: Topological feature extraction and tracking. *J. Phys. Conf. Ser. 78*, 1 (2007), 012007. 2

[BDF*08] BIASOTTI S., DE FLORIANI L., FALCIDIENO B., FROSINI P., GIORGI D., LANDI C., PAPALEO L., SPAGNUOLO M.: Describing shapes by geometrical-topological properties of real functions. *ACM Comput. Surv. 40*, 4 (2008), 1–87. 1

[BDFP07] BIASOTTI S., DE FLORIANI L., FALCIDIENO B., PAPALEO L.: Morphological representations of scalar fields. In *Shape Analysis and Structuring*, Floriani L. D., Spagnuolo M., (Eds.). 2007. 1

[BLW12] BAUER U., LANGE C., WARDETZKY M.: Optimal topological simplification of discrete functions on surfaces. *Discrete. Comput. Geom. 47*, 2 (2012), 347–377. 2

[CG97] CHERKASSKY B. V., GOLDBERG A. V.: On Implementing the Push-Relabel Method for the Maximum Flow Problem. *Algorithmica 19*, 4 (1997), 390–410. 5

[CGOS13] CHAZAL F., GUIBAS L. J., OUDOT S. Y., SKRABA P.: Persistence-based clustering in riemannian manifolds. *J. ACM 60*, 6 (2013), 41:1–41:38. 2

[CSEH07] COHEN-STEINER D., EDELSBRUNNER H., HARER J.: Stability of persistence diagrams. *Discrete Comput. Geom. 37*, 1 (2007), 103–120. 4

[CSEM06] COHEN-STEINER D., EDELSBRUNNER H., MOROZOV D.: Vines and vineyards by updating persistence in linear time. In *Proceedings of the Twenty-second Annual Symposium on Computational Geometry* (New York, NY, USA, 2006), SCG '06, ACM, pp. 119–126. 2

[CZCG05] CARLSSON G., ZOMORODIAN A., COLLINS A., GUIBAS L.: Persistence barcodes for shapes. *Int. J. Shape Model. 11*, 2 (2005), 149–187. 2

[DL12] DI FABIO B., LANDI C.: Persistent homology and partial similarity of shapes. *Pattern Recogn. Lett. 33*, 11 (2012), 1445 – 1450. 2

[DLL*10] DEY T., LI K., LUO C., RANJAN P., SAFA I., WANG Y.: Persistent heat signature for pose-oblivious matching of incomplete models. *Comput. Graph. Forum 29*, 5 (2010), 1545–1554. 2

[DSNW13] DORAISWAMY H., SHIVASHANKAR N., NATARAJAN V., WANG Y.: Topological saliency. *Comput. Graph. 37*, 7 (2013), 787 – 799. 1

[DW93] DIXON M., WIENER G.: TITAN: Thunderstorm Identification, Tracking, Analysis, and NowcastingA Radar-based Methodology. *J. Atmos. Oceanic Technol. 10*, 6 (1993), 785–797. 2

[EH10] EDELSBRUNNER H., HARER J.: *Computational Topology: An Introduction*. American Mathematical Society, 2010. 3

[EHM*08] EDELSBRUNNER H., HARER J., MASCARENHAS A., PASCUCCI V., SNOEYINK J.: Time-varying reeb graphs for continuous spacetime data. *Comp. Geom. - Theor. Appl. 41*, 3 (2008), 149 – 166. 2

[EIK01] EFRAT A., ITAI A., KATZ M. J.: Geometry helps in bottleneck matching and related problems. *Algorithmica 31*, 1 (2001), 1–28. 4, 5

[ELZ02] EDELSBRUNNER H., LETSCHER D., ZOMORODIAN A.: Topological persistence and simplification. *Discrete Comput. Geom. 28* (2002), 511–533. 2, 3

[FOTT08] FUJISHIRO I., OTSUKA R., TAKAHASHI S., TAKESHIMA Y.: T-map: A topological approach to visual exploration of time-varying volume data. In *High-Performance Computing*, Labarta J., Joe K., Sato T., (Eds.), vol. 4759 of *Lect. Notes Comput. Sci.* Springer Berlin Heidelberg, 2008, pp. 176–190. 2

[FRC*07] FARR T. G., ROSEN P. A., CARO E., CRIPPEN R., DUREN R., HENSLEY S., KOBRICK M., PALLER M., RODRIGUEZ E., ROTH L.: The shuttle radar topography mission. *Rev. Geophys. 45*, 2 (2007). 6

[GJR*14] GUNTHER D., JACOBSON A., REININGHAUS J., SEIDEL H.-P., SORKINE-HORNUNG O., WEINKAUF T.: Fast and memory-efficiency topological denoising of 2d and 3d scalar fields. *IEEE Trans. Vis. Comput. Graph. 20*, 12 (2014), 2585–2594. 1

[GR71] GARFINKEL R. S., RAO M. R.: The bottleneck transportation problem. *Nav. Res. Logist. Q. 18*, 4 (1971), 465–472. 4

[HFZ*09] HAN L., FU S., ZHAO L., ZHENG Y., WANG H., LIN Y.: 3D Convective Storm Identification, Tracking, and Forecasting–An Enhanced TITAN Algorithm. *J. Atmos. Oceanic Technol. 26*, 4 (2009), 719–732. 2

[HHC*13] HUETTENBERGER L., HEINE C., CARR H., SCHEUERMANN G., GARTH C.: Towards multifield scalar topology based on pareto optimality. *Comput. Graph. Forum 32*, 3pt3 (2013), 341–350. 1

[JMW*98] JOHNSON J. T., MACKEEN P. L., WITT A., MITCHELL, STUMPF G. J., EILTS M. D., THOMAS K. W.: The Storm Cell Identification and Tracking Algorithm: An Enhanced WSR-88D Algorithm. *Wea. Forecasting 13*, 2 (1998), 263–276. 2

[KSG] KATSIKOULI P., SARKAR R., GAO J.: Persistence based online signal and trajectory simplification for mobile devices. to appear in Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. 2

[LS09] LAKSHMANAN V., SMITH T.: An Objective Method of Evaluating and Devising Storm-Tracking Algorithms. *Wea. Forecasting 25*, 2 (2009), 701–709. 2, 6, 8

[SWC*08] SCHNEIDER D., WIEBEL A., CARR H., HLAWITSCHKA M., SCHEUERMANN G.: Interactive comparison of scalar fields based on largest contours with applications to flow visualization. *IEEE Trans. Vis. Comput. Graph. 14*, 6 (2008), 1475–1482. 1

[TG09] TRICOCHE X., GARTH C.: Topological methods for visualizing vortical flows. In *Mathematical Foundations of Scientific Visualization, Computer Graphics, and Massive Data Exploration*, Möller T., Hamann B., Russell R., (Eds.), Mathematics and Visualization. Springer Berlin Heidelberg, 2009, pp. 89–107. 1

[Val13] VALLET B.: Homological persistence for shape based change detection between Digital Elevation Models. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences-II* (2013), 49–54. 2

[VH01] VELTKAMP R. C., HAGEDOORN M.: State of the art in shape matching. In *Principles of Visual Information Retrieval*, Lew M. S., (Ed.). Springer-Verlag, London, UK, 2001, pp. 87–119. 4

[WBP07] WEBER G., BREMER P.-T., PASCUCCI V.: Topological landscapes: A terrain metaphor for scientific data. *IEEE Trans. Vis. Comput. Graph. 13*, 6 (2007), 1416–1423. 1

[WCBP12] WIDANAGAMAACHCHI W., CHRISTENSEN C., BREMER P.-T., PASCUCCI V.: Interactive exploration of large-scale time-varying data using dynamic tracking graphs. In *Large Data Analysis and Visualization (LDAV), 2012 IEEE Symposium on* (2012), pp. 9–17. 2

[WG09] WEINKAUF T., GÜNTHER D.: Separatrix persistence: Extraction of salient edges on surfaces using topological methods. *Comput. Graph. Forum 28*, 5 (2009), 1519–1528. 1

[YMV07] YIN F., MAKRIS D., VELASTIN S. A.: Performance Evaluation of Object Tracking Algorithms. In *10th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS2007), Rio de Janeiro, Brazil* (Oct. 2007). 7