32nd EPS Conference on Plasma Phys. Tarragona, 27 June - 1 July 2005 ECA Vol.29C, P-2.030 (2005)

Pattern recognition techniques in plasma turbulence imaging

H. Thomsen, T. Klinger , R. König, A. Alonso†, C. Hidalgo†

Max-Planck-Institut für Plasmaphysik, EURATOM Association, Greifswald, Germany

†Laboratorio Nacional de Fusión. Asociación EURATOM-CIEMAT, Madrid, Spain

Introduction

The availability of fast framing CCD cameras with up to 1 million frames per second enables the study of edge plasma turbulence by means of the Gas Puff Imaging (GPI) technique [1]. The data contain a lot of information about the turbulent dynamics in the plasma boundary. Potentially, one could gain deep insight into edge plasma turbulence, however, the data analysis is a challenge for several reasons: 1. high complexity, 2. huge amounts of data (2d spatial information with good temporal resolution), 3. no established data analysis schemes for such data. Typical data rates of todays fast framing cameras are of the order of 100 Mbyte/s. E.g., the Photron Ultima 40k camera (specs: 40500 frames/s with 64x64 pixel resolution, 8 bit) has a data rate of 160 Mbyte/s with a recording time of up to 2 s (limited by the amount of camera memory, only). Taking into account the technical improvements in camera resolution which can be expected over the next few years, the data storage problem will be getting more severe.

Due to the stochastic nature of turbulence, a statistical characterization scheme is appropriate. Useful characterizing quantities (or meta-information) are sizes and life times of structures as well as propagation velocities. For one spatial and one temporal dimension only, the average values of these quantities can be derived from fits to the 2d-cross-correlation function [2]. This analysis method is not appropriate for the fast camera data with two spatial dimensions, since the dimensionality in the cross-correlation function is conserved. Here, we describe a different approach for the extraction of meta-information by means of pattern recognition techniques and demonstrate the capability of this method for preliminary data from turbulence imaging

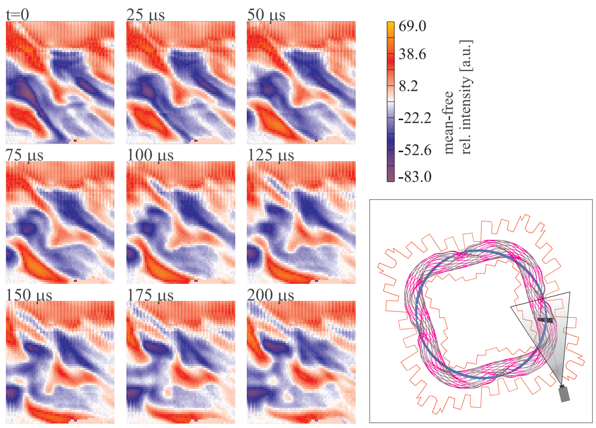

Figure 1: Sequence of video frames recorded by an Ultima 40k fast framing camera with a rate of 40.5 kFrames/s. The camera is viewing in toroidal direction and records light from the radial-poloidal plane of the edge plasma of the TJ-II stellarator (discharge #11431). The moving average intensity of each pixel over 100 frames is subtracted. Inset: The view of the camera and the magnetic configuration in TJ-II (top view).

measurements on the TJ-II stellarator. The Photron Ultima 40k camera recorded the turbulent fluctuations in the visible light in a view approximately parallel to the magnetic field. A sequence of 9 frames is shown in Fig. 1.

Region hierarchy and extraction of meta-information

Pattern recognition generally consists of three basic steps, namely: segmentation, feature extraction and classification [3]. The segmentation divides the data in regions, which should represent existent structures in the data. If we consider a single video frame, the interesting subgroups are the local maxima and minima. For the multi-scale data in our case, a watershed algorithm [3] yields good extraction results. The algorithms used here are programmed in IDL [4]. In Fig. 2, a sample input frame and the detected local maxima are shown. The recursive

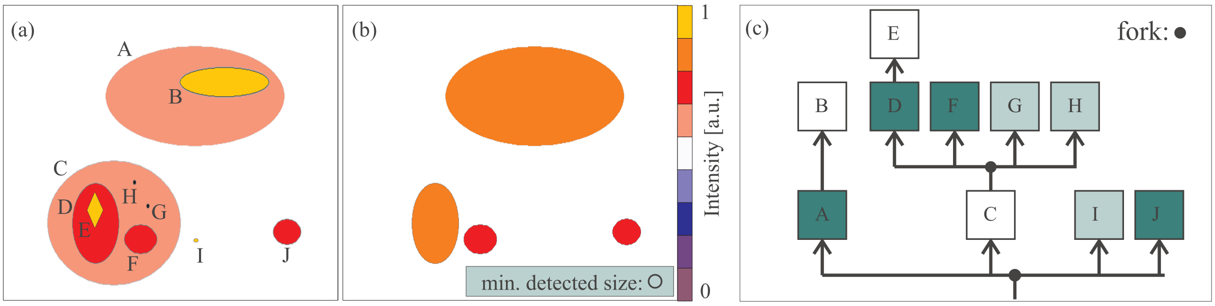

Figure 2: (a) Sample input for the segmentation: Objects A–J with different amplitudes and sizes. (b) Extraction result are the 4 objects, which do not contain more than one sub-object and are larger than the minimum detected size. The amplitude is the average over the intensity of the detected structure size. (c) Region object hierarchy in tree representation. The green objects follow the selection rule (see text), the darker green objects are also larger than the minimum size (rejection of pixel artifacts).

algorithm looks for regions with a higher amplitude than a given base-level. In each of the obtained region, all regions with the next-higher level are extracted and so on, until the regions with the highest level are found. At the end of the segmentation, a hierarchy of region objects (which all know their parent and their children objects) describes each frame (Fig. 2c). Pixel artifacts are suppressed by the definition of a minimal allowable structure size. A useful subset of these region objects are those regions, which are one level above a fork (that is a parent object with two or more children objects) and which have no fork in their descendants (cf. Fig. 2). This subset maps the basic structure of the region hierarchy. Moreover, the region objects in the subset have no intersections between each other, which is advantageous for the region tracking described below.

In a next step, the features (as mean intensity, size, position of the centroid, elongation, aspect ratio and tilt angle) are extracted from those region objects by principle components analysis. This way, the shapes of the region objects are approximated by ellipsoids. Each region object is represented by a single point in a low-dimensional feature space. Provided that this representation is sufficient, i.e. the features contain the relevant turbulent structures, the problem of the large data amounts to be handled is solved: the typical packing ratio is of the order of 1 / 1,000,000. It is of course necessary to check the assumption of a sufficient approximation by comparison of raw and meta data. However, as will be shown in the next section, the utilized segmentation algorithm is already quite powerful for the statistical description of the turbulent dynamics.

For the classification, sequences of frames are considered, since we are interested in the turbulent dynamics in the spatio-temporal domain (not only spatial). This is possible by a simple object tracking based on a minimal distance criterion between region objects in subsequent frames. The paths of region objects can be followed over several frames and the velocities are calculated accordingly. One main advantage of the pattern recognition technique, which cannot be easily accessed by other methods, is the possibility to get the feature *distributions* in contrast to just averages. Histograms are a good representation, e.g. of the size of region objects, life time and velocity . We note here, that effects like region splitting (which might occur in a sheared flow) will not always be detected in a correct way by this simple approach. In a further step it is planed to incorporate the full hierarchy of region objects for a better matching of features to the turbulent structures and in order to improve the tracking results.

First results of pattern recognition characterization

In Fig. 3(a) the paths of the basic structures are plotted for frame #7 of Fig. 1. The positive structure ’1’ is moving from the middle of the left edge towards the lower right corner. Structure ’3’ starts higher at the top left edge and then shows a similar trajectory. The negative structure ’2’ seems to follow structure ’1’, suggesting a dipole structure. However, this effect might be an artifact caused by the data preprocessing, since the subtraction of the mean from the original data causes a negative ’hole’ for a moving positive structure. In Fig. 3(b) the paths of region objects (maxima only) within a sequence of 100 frames are plotted, and the corresponding velocity field is shown in Fig. 3(d). A similar flow as previously described for the single region objects ’1’ and ’3’ is also evident in this averaged data representation. From the velocity field it is thus possible to get valuable information on the turbulence dynamics without the need of (the very time-consuming) frame-by-frame evaluation of the video data.

Further useful statistics for the turbulence characterization are histograms of the area, mean value, life time and velocity of the basic structures (cf. Fig. 3(c,e)). The detection of large coherent structures (’blobs’) is of particular interest in the GPI investigations [1], since these blobs seem to be responsible for a large fraction of the anomalous transport in the edge plasma.

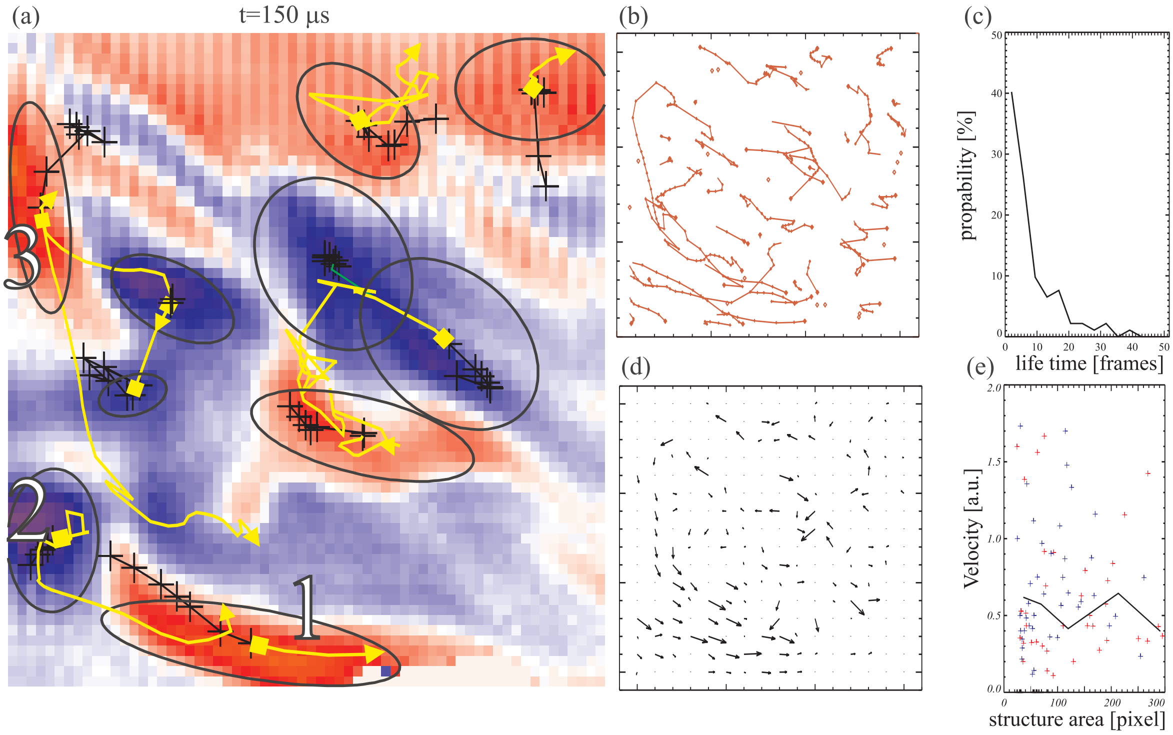

Figure 3: (a) Result of the segmentation and feature-extraction (black ellipsoids: the region objects) and of the object tracking (black crosses: previous positions, yellow lines: path in subsequent frames). Numbers mark region objects discussed in the text. (b) Object paths from tracking, connected to the object life time. (d) Velocity field.

These blobs are characterized by a relatively large life time and a strong radiation due to their high density. They can easily be detected within the pattern recognition approach by searching for region objects with a large product of life time and intensity.

Conclusions

Pattern recognition techniques for data analysis of turbulence imaging is a very promising approach. This method has the potential to significantly simplify the analysis of the huge data amount collected by fast framing cameras. Combined with object tracking a unique insight into the turbulent dynamics is possible. It is possible to estimate distributions of feature parameters like life times of structures, etc. Moreover, the extracted features are a good basis for novelty detection techniques, which statistically classify turbulence data, e.g. for different discharge scenarios.

References

[1] S. Zweben et al., Nucl. Fusion **44**, 134 (2004)

[2] J. Bleuel et al., New J. Physics **4**,38 (2002)

[3] M. Sonka, V. Hlavac, R. Boyle. Image processing, Analysis and Maschine Vision -2nd ed. (PWS Publishing, Pacific Grove, 1998)

[4] http://www.rsinc.com/idl