EFDA–JET–PR(11)12

T. Craciunescu, A. Murari, I Tiseanu, P. Lang
and JET EFDA contributors

Motion Estimation within the MPEG Video Compressed Domain for JET Plasma Diagnostics
Motion Estimation within the MPEG Video Compressed Domain for JET Plasma Diagnostics

T. Craciunescu\textsuperscript{1}, A. Murari\textsuperscript{2}, I Tiseanu\textsuperscript{1}, P. Lang\textsuperscript{3} and JET EFDA contributors*
ABSTRACT.
Video cameras have recently become diagnostic tools widely used on JET for fusion plasma diagnostic and control. Since video streams are usually compressed for storage, MPEG-2 compressed domain information is processed to obtain a very fast and reasonably accurate 2-D motion estimation of the video scenes for the JET diagnostics, whose computational costs are prohibitively high. These methods can be used for the manipulation of the large JET video databases and, in specific cases, even for real-time data processing. Plasma instabilities which can trigger harmful disruptions are detected and tracked by means of motion segmentation. Motion segmentation is used as a key contrivance to allow very fast optical flow estimation for the determination of the deuterium ice extrusion velocity of JET pellet injector. Experimental validation is performed on significant JET video data.

1. INTRODUCTION
Video cameras have recently become common diagnostic tools in Magnetic Confinement Nuclear Fusion [1]. Camera based instruments provide essential information for both the control of the experiments and the physical interpretation of the results. In JET, the applications are ranging from protection of the first wall to the analysis of plasma instabilities and even the characterisation of turbulence. In the last campaigns many experiments relied on IR and/or visible cameras and about 15 new cameras are being installed for the next experiments with the new ITER-like wall.

These cameras can produce up to hundreds of kiloframes per second and their information content can be very different, depending on experimental conditions. However, the relevant information about the underlying processes is generally of much reduced dimensionality compared to the recorded data. The extraction of the relevant information which allows the full exploitation of this diagnostic, is a challenging task. In the last few years, new tools and methods were developed in order to manipulate this huge amount of information and to retrieve the desired information. A series of pattern recognition techniques were developed and implemented, which allows content based search of video database [2-3]. Algorithms for the automatic identification and tracking of objects in videos (e.g. MARFE’s and UFO) were developed in order to provide information useful for the improvement of disruption prediction techniques [4]. Various machine learning methods proved to be indispensable to properly classify the various objects appearing in the frames recorded by the visible cameras. [5]. New image processing algorithms, were implemented and validated successfully on JET for real-time identification of hot spots with a time resolution of tens of milliseconds. Specific real-time algorithms, based on the computational paradigm of Cellular Nonlinear Networks, have been implemented on Field-Programmable Gate Arrays (FPGA) to identify hot spots on the vacuum vessel and therefore to protect JET plasma facing components. [6]. The methodology of optical flow has allowed deriving information about movement of objects in 3-D space even if they have been detected by a single camera [7].

Even if the information provided by the cameras is in a form which may be particularly intuitive
for scientists to interpret, the extraction of quantitative information from the images can be a complex task. Another main challenge resides in the huge amount of data. With the rapid increase of imaging diagnostics, very large databases, with millions of signals and Tbytes of data, have to be automatically analyzed. For a number of physical processes, like e.g. plasma instabilities, at least a basic level of information has to be obtained from the videos in real-time. This is an extremely difficult task, as the events revealed by the video sequences have time constants ranging from ms to tens of ms.

To reduce costs most videos are stored in some kind of compressed format. Therefore it would be more efficient to retrieve the information, needed for physical studies, in the compressed domain. A typical example is the velocity of video objects, associated with different physical phenomena like e.g. instabilities, pellets, and filaments, captured by the cameras. In case of MPEG compression format, widely accepted as a video compression standard, a kind of motion estimation is performed for compression purposes. It exploits the redundancy between consecutive or temporally close video images. This MPEG motion field (MPEG-MF) estimation is a crude approximation to the optical flow although it is heavily corrupted by the sudden occurrence of lack of correlation between estimated motion vectors and real motion in the video scene. However, this information can be exploited, as presented in this paper, for separating the video sequence into multiple spatio-temporal regions corresponding to different rigid-body motion. Part of the rigid-body moving objects can be associated with different relevant physical phenomena. In this case the motion segmentation is further used for fast accurate optical flow estimation.

The main advantage derives from the fact that only partial decoding of the compressed video is needed to extract the useful information. This allows fast manipulation of large databases. For specific cases, the calculation speed is compatible with on-line evaluation of the motion parameters. In this case the algorithms for motion segmentation and optical flow estimation can be implemented inside the MPEG compressing code, allowing the evaluation and storage of the physical parameters together with the storage of the video sequence.

Summarizing, the aim of the paper is to prove that a very fast and reasonably accurate motion estimation of video scenes, relevant for the analysis of JET plasma, can be obtained using information within the MPEG compressed domain.

2. OVERVIEW OF THE METHODS

2.1. THE OPTICAL FLOW

Optical flow, or “projected motion”, is a dense representation of visual motion important for dynamic scene understanding. It can be defined as the (perspective or orthographic) projection of 3-D motion in the real world on the 2-D image plane. The motion field is the field in the image plane that is associated with the spatio-temporal variations of intensity pattern. It gives the distribution of apparent velocities corresponding to brightness patterns moving in the image.

The basic assumption used by most algorithms is the brightness constancy: when a pixel flows
from one image to another, its intensity or colour does not change. This assumption was introduced by
Horn and Schunk, in a pioneering paper [8]. It combines a number of assumptions: i) the reflectance
of the scene is Lambertian (a surface that obeys Lambert’s Law appears equally bright from all
viewing directions), ii) the illumination in the scene is uniform and iii) the image formation process
in the camera is not affected by distorting effects like e.g. vignetting. The brightness constancy
assumption implies that all changes in the image are caused by the translation of brightness patterns
leading to the gradient constraint equation:

$$\vec{f}_E \cdot \vec{v} + \vec{f}_t = 0$$

(1)

where \(f_t = (f_x, f_y)\) and \(f_t\) are the spatial and temporal gradients and \(v\) is the optical flow velocity. The
brightness constancy assumption provides just one constraint on the two unknowns at each pixel.
Thus the gradient constraint equation is ill-posed and requires additional constraints.

The ill-posed optical flow problem can be regularized by introducing a spatial term that models
how the flow is expected to vary across the image. The simplest approach is to use a spatial
smoothness assumption. In a quadratic formulation, the optical flow \((u, v)\) is determined as the
minimiser of the global energy functional:

$$E_{global}(u,v) = \int_{\Omega} \left( (f_x u + f_y v + f_t)^2 + \alpha (|\nabla u|^2 + |\nabla v|^2) \right) dx dy$$

(2)

where \(\alpha > 0\) determines the amount of smoothness and serves as regularization parameter: larger
values for \(\alpha\) result in a stronger penalization of large flow gradients and lead to smoother flow
fields. \(u\) is the \(x\) component of the flow and \(v\) is the \(y\) component of the flow. Minimising this convex
functional comes down to solving its corresponding Euler–Lagrange equations:

$$\Delta u - f_x (f_x u + f_y v + f_t) = 0$$

$$\Delta v - f_y (f_x u + f_y v + f_t) = 0$$

(3)

A Gauss-Seidel iterative method can be used to solve this system of equations [9].

Unfortunately the quadratic formulation (2) is not robust for outliers. A wide variety of other penalty
functions have been reported – for a review see Ref. 10. Different objective functions that define
the problem were introduced together with different optimization algorithms and implementations.
From a practical point of view, it remains unclear which of these approaches is best.

A different way to tackle the ill-posed optical flow problem comprises the so called local methods,
which were introduced in [11]. This approach assumes that the unknown optical flow vector is
constant within a neighbourhood of size \(\rho\). Therefore \(\vec{v}\) can be determined at the location \((s, t)\) from
a weighted least square fit by minimizing the function:

$$E_{local}(u,v) = K_\rho \times (f_x u + f_y v + f_t)^2$$

(4)
where $K_\rho$ is a Gaussian smoothing operator which is used in order to remove noise and to stabilize the differentiation process. A sufficiently large value for $\rho$ is very successful in rendering the method robust against noise. The main drawback of this class of local methods is related to the existence of flat regions of the emission, where the image gradient vanishes. In this case the method is unable to produce dense (one motion vector per pixel) flow fields. Brun et al. [12] noticed the similarity between equation (4) of the local OF approach and the first term under the integral (2), in the global approach. This suggested a way to extend the Horn–Schunck functional to a Combined Local-Global (CLG) one, by replacing the first term under the integral with the local term with some integration scale $s > 0$. The hybrid CLG class of methods incorporates the advantages of both paradigms: they are highly robust under Gaussian noise and at the same time they provide dense flow fields. The price paid for obtaining this synergy consists of the hardly more complicated form of the CLG equations:

$$\Delta u - \frac{I}{l} \left( K_\rho * (f_x^2) u + K_\rho * (f_x f_y) v + K_\rho * (f_x f_t) \right) = 0$$

$$\Delta v - \frac{I}{l} \left( K_\rho * (f_y^2) v + K_\rho * (f_x f_y) u + K_\rho * (f_y f_t) \right) = 0$$

(5)

The sparse linear systems of Eq. (5) may be solved iteratively, usually by using the Successive Over-Relaxation (SOR) method [13], which ensures a good compromise between simplicity and efficiency. SOR is a method of solving a linear system of equations derived by extrapolating the Gauss-Seidel method. This extrapolation takes the form of a weighted average between the previous iterate and the computed Gauss-Seidel iterate successively for each component. The idea is to choose a value for weighting factor $\omega$ that will accelerate the rate of convergence of the iterates to the solution. SOR fails to converge if $\omega$ is outside the interval $(0,2)$ [14].

The CLG model was proved to work with good results for the very specific case of JET images, as described in detail in Refs. 7 and 15.

The optimization method and the practical implementation strongly influence the accuracy of the optical flow evaluation. When working with real JET video images there are several issues that need to be solved. Particularly important is to deal with large displacements of objects between consecutive frames. A current practice when dealing with this problem is the multi-resolution coarse-to-fine procedure [16]. A pyramid of multi resolution images is derived from the original frame by successive down-sampling and Gaussian smoothing steps. Optical flow calculations start at the coarse level, where the displacements are small and consequently the linearization of the gray value constancy assumption is satisfied. This estimate is then refined step by step along the pyramidal structure. Occlusions, which appear when two objects that are spatially separated in the 3D space might interfere with each other in the projected 2D image plane, can be treated using the method suggested by Alvarez et al. [17]. This method simultaneously computes forward and reverse flow fields, labeling pixels as occluded where the two disagree. High-order filter constancy can be used to reduce the influence of lighting changes [18]. This strategy is somewhat similar to filtering the input images using gradient filters [19].
The optical flow encodes the information provided by time-varying images for dynamic scene understanding. In particular it identifies distinct image regions based on their perceived motion in the image plane. Therefore the resulting information can be used to decompose the image sequence into sets of pixels moving coherently across the sequence with associated motions. Extracted spatio-temporal information obtained after motion segmentation may be used in order to track the moving objects. Motion segmentation and tracking is of particular interest for JET plasma diagnosis for detection of instabilities. This may contribute to increasing the success rate of disruption prediction and therefore to the safe operation of large devices such as JET.

In general, the optical flow algorithms are accurate but cannot fulfill the performance requirement of real-time applications. They produce good results at the expense of being too slow for real-time applications. However, in the last years, significant efforts were made to improve the speed of optical flow computation. First attempts relied on wavelet-based methods. For example Bernard [20] proposed an approach based on the projection of the optical flow equation on vectors of a wavelet basis. Liu et al. [21] proposed a multiscale algorithm which uses wavelet filters to estimate dense optical flow from two frames. These approaches succeeded to reduce the complexity of the optical flow algorithm to $O(N)$, where $N$ is the number of pixels in one image. The algorithms efficiency is better than 300 flops per pixel. Bruhn et al. [22] introduced a multigrid scheme for the linear system of equations resulting from the discretised Euler–Lagrange equations, that have to be solved in order to minimize the convex functional of the CLG method. The method has to be tailored for a specific application and under certain conditions it is able to provide optical flow calculation at the speed of ~27 frames per second (fps) for images of size 200 x 200 pixels. Improved performances – ~64 fps for images of size 640 x 480 pixels - were obtained using FPGA technology [23].

Recently, the field of video object extraction and motion estimation has been transferred from the pixel domain to the compressed domain, since video data is usually compressed for storage and transmission. This allows real-time performance which is obtained at the price of a certain sacrifice of accuracy. It has to be mentioned that these approaches have been developed especially for applications like object-based video coding, video summarization, intelligent video surveillance, where the typical video frame rate is around 20 fps.

### 2.2. THE MPEG COMPRESSED DOMAIN

Video sequences contain a significant amount of data similarity within and between frames. These statistical redundancies in both temporal and spatial directions are exploited for moving picture compression in MPEG-2 standard [24]. Although alternatives exist (H.261, H.263, MPEG-1, MPEG-4), this standard keeps a high level of popularity and is widely used in DVD and digital TV. The bit-rate reduction for storage and transmission is achieved by using the Discrete Cosine Transform (DCT) to reduce the spatial redundancy, and a Motion-Compensation (MC) scheme based on a block matching method to exploit the temporal redundancies. A Huffman entropy encoder is used afterwards to obtain a compact bit stream.
The basic assumptions of the MPEG encoding process are the existence of an inter-pixel correlation and of a simple translatory motion between consecutive frames. Then the magnitude of a particular image pixel can be predicted from nearby pixels within the same frame (intra-frame coding) or from pixels of a nearby frame (inter-frame coding).

MPEG encoding is implemented using macroblocks (MB16) which consist of 16 x 16 pixels. For intra-frame coding each MB16 is further divided into 8 x 8 pixel blocks (MB8). In case of a 4:2:0 chroma sub-sampling this results in 6 MB8 for a MB16.

MPEG stream consists of three types of pictures: I, P and B. Intra (I) frames are coded using only information present in the picture itself by the DCT. Each MB8 block of the I frame is processed independently with an 8 x 8 DCT. The distribution of the DCT coefficients is non-uniform. The DCT concentrates the energy into the low-frequency coefficients and many of the other coefficients are near zero. This is a result of the spatial redundancy present in the original image block. Consequently a bit rate reduction is obtained by neglecting the low value coefficients. Afterwards quantization is used to reduce the number of possible values to be transmitted. High-frequency coefficients are more coarsely quantized than the low-frequency coefficients. Quantization makes the coding/decoding processes ‘lossy’ due to the irreversible noise introduced. The result is similar to JPEG compression.

The high degree of compression specific to MPEG is achieved by means of the inter-frame coding. Inter-frame coding uses the correlation between the current frame and a past or future frame to achieve compression. P (predicted) frames are coded with forward motion compensation, using the nearest previous reference (of type I or P) images. Bi-directional (B) pictures are also motion compensated, this time with respect to both past and future reference frames. The parts of the image that do not change significantly are simply copied from other areas or other frames. In case of the other parts, for each MB16, the best matching block is searched in the reference frame(s). Residuals are then encoded, in a similar manner to I frame coding. A trade-off must be found between the accuracy in predicting complex motion in the image and the expense of transmitting the Motion Vectors (MV). Therefore only one MV is estimated for each MB16. The motion vectors are directly entropy coded without transformation and quantization.

Summarizing, the MPEG stream carries both intensity and motion information of the underlying scene. Intensity is represented by a set of DCT coefficients while motion is represented by a field of MV vectors. The motion between frames is described by a limited number of parameters (i.e. MV translation vectors). MPEG video stream may be of different types, i.e. I-, P or B-, and can occur in a variety of GOP (Group Of Pictures) patterns. 2-D motion estimation of a video scene within the MPEG compressed domain is presented in certain details in the following paragraph.

In order to transform the MV field in a smooth motion field, Coimbra [25] introduced a regularization set of rules that do not rely on future images (in case of B-frames). Therefore they are appropriate for real-time application. Assuming that consecutive images are strongly correlated, in case of I-frames, which have no motion, the MV field is supposed to be the same as for the previous
image. In case of B-frames, for which two sets of MV exists, the one pointing backward is reversed and averaged with the one pointing forward. Applying these rules creates a motion field with one MV per macroblock.

A confidence measure can be introduced to ensure that the MV field is meaningful. Starting from the assumption that areas with strong edges exhibit better correlation with real motion than textureless ones, the DCT coefficients AC(1) and AC(8) (see Fig. 1) can be used to measure the edge strength in the horizontal and vertical directions within the MB8 block. In Ref. 25 it is shown that these coefficients can be interpreted as weighted averages of the image gradients $f_x$ and $f_y$. Shen et al. [26] proved that the AC(1) and AC(8) coefficients outperform the Sobel edge detector.

Simoncelli [27] introduced a confidence measure for the optical flow by mean of the following algorithm:

$$v = -M^{-1} \left( \sum_i b_i \right)$$

where:

$$M = \begin{pmatrix} f_x^2 & f_x f_y \\ f_x f_y & f_y^2 \end{pmatrix} = \frac{\pi^2}{2} \begin{pmatrix} -AC(1) & -AC(8) \\ -AC(8) & -AC(1) \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} f_x \\ f_y \\ f_t \end{pmatrix}$$

The confidence of the optical flow estimate is determined by the matrix $M$. An eigenvalue decomposition of the matrix $M$ can be performed. The size of the eigenvalues ($\lambda_1, \lambda_2$) is a measure of uncertainty in the direction of the corresponding eigenvector. Barron et al. [28] argue that the first eigenvalue $\lambda_1$ alone provides a good confidence measure for the optical flow field. The eigenvalue $\lambda_1$ can be calculated using the following relation:

$$\lambda_1 = \frac{\pi^2}{2} \frac{\left( -AC(1) \right)^2 - AC(8)^2}{2}$$

A different approach consists in using the MV field only as a crude initial estimation for optical flow recovery which can be performed using the classical framework, outside the compressed domain [see e.g. Ref. 29]. The intensity information for the optical flow algorithm can be provided by the DC coefficients. The use of the (DC + 2AC) image, instead of just the DC image, is adequate to provide more intensity information for the estimation of the compressed domain optical flow field and ensures improved results [30]. The (DC + 2AC) image is a reconstructed image from the DC and two lowest-order AC coefficients (AC(1) and AC(8)) of each block.

The DCT coefficients are readily accessible for I frames, but they must be estimated for P and B frames. Therefore the DCT coefficients for the MB16 of the reference frame, that the current block was predicted from, need to be calculated. As the MV vectors have half-pel accuracy, they do not necessary point to the top-left corner of a MB16 DCT block (Fig. 2). Therefore interpolation is necessary. Liu et al. [31] proposed an efficient technique for calculating reasonable approximations to the DC coefficients of a MB of a P or B frame.
3. APPLICATION TO JET VIDEO DIAGNOSTICS

3.1. OFFLINE APPLICATION: INSTABILITY IDENTIFICATION

The above described techniques were used as a basic skeleton for developing an efficient procedure adapted for the characteristics of the JET video data.

Several plasma instabilities are revealed as distinct moving objects in videos recorded by visible cameras. The motion estimation of these instabilities may provide useful information for the improvement of the safe operation by increasing the success rate of disruption prediction techniques. An illustrative example, which will be used here to describe part of our implementation, is the Multifaceted Asymmetric Radiation from the Edge (MARFE) [32]. MARFEs are characterized by a small region where the radiative losses are higher than the input conducted power and usually occur on the high field side of the torus. MARFEs can reduce confinement and, more importantly, they can cause disruptions which lead to the sudden losses of plasma confinement which determine the abrupt end of the discharge. They may represent a risk for the integrity of the devices.

A typical MARFE sequence of images recorded with the fast visible camera (Photron APX) installed at JET is presented in Fig. 3. The camera views the full poloidal cross-section of the vacuum vessel and covers a toroidal extent of ~90°. The wide angle view is appropriate for the study of pellet ablation, large scale instabilities and plasma wall interactions [33]. The MPEG MV field obtained after applying the above described Coimbra’s regularization rules is represented in Fig. 4. The regularization procedure removes the dependence on specific MPEG-2 characteristics such as picture and macroblock type and creates a motion field with one vector per macroblock with standardized magnitude. The size of the each image representing the MV field is 16 x 14 pixels.

A brief and simple analysis of Fig. 4 reveals that the MV field, affected by irregular random patterns, cannot be used as an initial guess for the optical flow problem. However, this information can be exploited for MARFE detection and tracking. A reasonable assumption is to consider that just a few moving objects exist within a frame, with a size larger than the noise. Therefore a 3x3 spatial median filter is then applied to the motion field, removing isolated vectors that have a low probability of reflecting the real movement in the image. The median filtering and also the further morphological image processing steps are very fast since they are performed on images with a very low size, representing the MV field. The computation time for a single image processing step is accounted in Table 1. A pair of dilation and erosion operators is used to reveal “large” features of the image, using a structuring element which takes into account the first order vertical and horizontal neighbors. Afterwards the image object associated with the MARFE can be accompanied by other objects related to other physical phenomena, reflections on the vessel structure, false objects determined by the saturation of the image, etc. Analyzing the JET database for videos containing MARFE’s it was found that the image object describing the MARFE is always of the largest size. Therefore a labeling operation is further applied in order to detect the connected components in the image and select the moving object with the biggest area. The result is presented in Fig. 5.

The shapes evolving in Fig. 5 clearly include the flying MARFE. However the masks are not
tightly fastened on the shape of the MARFE. A block-based manipulation was performed to obtain a stretched enclosure. A moving object can be theoretically identified as a group of pixels in which the motion vectors are consistent with each other. A block grouping criterion for grouping neighboring pixels may be performed by using a consistency model [34]. In our approach the consistency criterion is based on a 4-connected model. A pixel \( P_{ij} \) is consistent with its neighbors if the difference between its gray-level and the gray-level of its neighbors \( \{ P_{i-1, j}, P_{i, j-1}, P_{i+1, j}, P_{i, j+1} \} \) is under a certain predefined value \( T_{\text{diff}} \). The grouping is performed by a block growing procedure which uses a seed pixel as starting point. The seed pixel corresponds to the local maximum gray-level. Afterwards the seed block and its four direct neighbors are first grouped together to form a new region \( R \). To grow the region, each ungrouped surrounding pixel that borders \( H \) is grouped into the region only if the consistency criterion between this pixel and all the pixels in \( H \) is satisfied. \( T_{\text{diff}} \) is an adjustable parameter of the algorithm. However, a unique threshold of 20% of the seed gray-level was found to be satisfactory for all the analyzed videos. The final result is presented in Fig. 6.

A MV field bounding box can be also easily calculated. The object centroids can be used for the MARFE tracking. In principle they can be used as inputs to a Support Vector Machines classifier for automatic identification of this kind of instabilities (see e.g. Refs. 5 and 35 for details about this procedure).

Summarizing, a set of simple and fast image processing steps have proven successful in deviding the image sequence into semantically meaningful regions along the time axis. The MARFE detection and tracking is achieved with very good quality. The procedure is very fast (see Table 1) because it performed on low size images representing the MV field retrieved from the MPEG compressed domain. The total processing time per frame is 6.9 ms.

The MV field can be further used for optical flow calculation. The non-zero MV region is used as a mask which is superimposed, after appropriate resizing, on the original image. In this way a Region Of Interest (ROI) is defined, enclosing the image area containing all the information needed for the determination of the speed of the objects moving in the images. The advantage is the dramatic reduction of the size of the image where the optical flow must be calculated. For example, for the sequence of images displayed in Fig. 6, the ratio \( R_{\text{OF}} \) between the area of the region of interest which is useful for the calculation of the optical flow and the total area of the image is, on average, less than 3%.

### 3.1 REAL TIME APPLICATION: PELLET VELOCITY

However, even if the size of the image area where calculation have to be performed has been dramatically reduced with the approach described in the previous subsection, an optimization of the optical flow procedure is still needed in order to achieve a computation times compatible with on-line processing of the input data. The procedure will be illustrated for an experimental example related to the determination of the deuterium ice extrusion velocity which is an important parameter for pellet injection.
Injection of solid, cryogenic hydrogen isotope pellets in tokamaks is used for particle fuelling as well as for Edge Localised Mode (ELM) instabilities control (triggering and mitigation). The method has been demonstrated to open access to operational regimes not reachable by gas puffing [36]. ELM triggering by pellets has been recognized as a potentially useful tool to mitigate type-I ELMs in large fusion experiments [37]. However, a controlled high fuelling efficiency is needed as otherwise the beneficial effects are spoiled by the increase of neutral pressure from fuel losses. ELM triggering and the variation of the ELMs dynamics depend on technical control variables such as pellet size, velocity, frequency and poloidal launch position.

Recently, optical flow was used in order to evaluate the ice extrusion velocity (see Ref. 15 for details). The method is based on the image sequences provided by a CCD camera viewing the ice at the exit of the nozzles of the extrusion cryostat [38]. A set of typical images recorded by this camera are presented in Fig. 7. A significant reduction of the computation time can be obtained if the optical flow calculation is performed only for the image region which overlaps the ribbons of ice. But as the ribbons of ice are floating in the image, the ROI must be determined dynamically, for each pair of images in the sequence. Segmentation techniques which minimize the measures of fuzziness of the image [39] were used in order to shrink the optical flow calculation to the image region where motion activity occurs. The image processing time was reduced but it is still far from on-line processing range. It is proved here that the information retrieved from MPEG compressed domain is useful for further optimization of the optical flow method. This optimization succeeds to satisfy the on-line requirements.

The segmentation can be performed in similar manner as for the flying MARFE. However, in this case the strong noise and saturation of the images provided by the video camera must be taken into account. The identification of each moving object in the image (i.e. each ice pellet) may increase significantly the image processing time. Therefore it is more efficient to reformulate the objective of the segmentation procedure. A tight enough region which encompasses the whole motion activity can be determined in two fast steps: applying the regularization rules for the MV field and median filtering. The computation time is accounted in Table 2. The segmented region, illustrated in Fig. 8 can be used further for optical flow calculations. The average value of the ratio $R_{OF}$ is below 8%.

An optimization for the CLG formulation of the optical flow problem has been necessary in order to obtain fast computation time. The sparse linear systems of equations (5) may be solved iteratively, by using the successive over-relaxation (SOR) method which ensures a good compromise between simplicity and efficiency. SOR introduces an additional relaxation parameter $\omega \in (0.2)$; for $\omega = 1$ the well-known Gauss–Seidel method is retrieved. The relaxation parameter has a strong influence on the convergence speed. A value in the range $(1.9, 2.0)$ leads to a numerically inexpensive over-relaxation step which results in a speed-up by one order of magnitude compared with the Gauss-Seidel approach. We vary also the number of pyramid levels in the multi-resolution coarse-to-fine procedure. We found that this number can be reduced to three levels, with no significant loss in what concerns the quality of the results. The small number of levels is allowed by the reduced size
of the image on which the optical flow algorithm is applied. From the point of view of the quality of the estimated optical flow, a key role is played also by median filtering of intermediate flow fields during optimization between successive iterations. As shown in Ref. 40 this operation leads to higher energy solution and it is equivalent to the optimization of a different objective function, which regularizes the optical flow over a large spatial neighborhood.

As can be seen in Table 2, the segmentation time is ~1.4 ms and the optical flow computation time, for three iterations, is ~13.2 ms (the image derivatives are calculated only at the beginning of the iterative process while the median filtering is performed after each SOR iteration). The total optical flow computation time is ~16.4 ms. As the image acquisition framing rate is 50 Hz, it results that, in principle, the optical flow calculations can be performed on-line. The optical flow can be estimated together with the storage of the video sequence.

High speed and good accuracy optical flow estimation are obtained. A representative result is reported in Fig. 9. Due to strong saturation of the images provided by the video camera, the ice ribbon-structure is only partially visible. Therefore the speed velocity field is reconstructed only when the visible ribbon structure is clearly identified. The difference between the speeds of the different pellets in the ribbon structure, calculated using the optical flow, is below 12.5%. This factor can provide an estimate of the error associated with the optical flow calculation. It was evaluated for more than 300 image pairs in the video sequence corresponding to the JET Pulse No: 76379. Its value is always below 16%.

Another criterion to discern between correct and wrong calculated flow fields can be based on the difference between one image of the image pair and its reconstruction obtained using the other image and the estimated optical flow:

\[
I_{\text{diff}} = (I_2 - I_2^{\text{rec}}), \text{ where: } I_2^{\text{rec}} = I_1 + OF
\]  

\[I_1, I_2\) is the image pair and OF is the optical flow calculated using \(I_1\) and \(I_2\). \(I_{\text{diff}}\) is presented also in Fig. 9. The Peak-Signal-to-Noise Ratio (PSNR), calculated for \(I_{\text{diff}}\) can also be used as a confidence criterion to determine the validity of the results. PSNR is defined as the ration between the maximum possible power of a signal and the power of the corrupting noise that affects the fidelity of the representation:

\[
\text{PSNR} = 10 \times \log \frac{128^2 \times N}{\sum_{i=1}^{M} \sum_{j=1}^{N} (I_{2,ij} - I_{2,ij}^{\text{rec}})^2}
\]

where \(N\) is the total number of pixels in the ROI and the images which are represented on 128 gray-levels. A higher PSNR indicates a higher quality of the reconstruction. For the image difference in Fig. 9 PSNR has a value of 19 dB. For the video corresponding to the JET Pulse No: 76379, PSNR has always a value above 14 dB. The main limitation on the accuracy of the optical flow estimate is determined by the poor quality of the input images which are affected mainly by strong noise and saturation.
The methods presented in this paper have been implemented in ANSI C. They have been inserted in the MPEG converter [41] which is a MATLAB mex file. This converter represents a convenient solution allowing a fast and friendly interaction via MATLAB during the development phase. However, the motion segmentation and optical flow algorithm can be engrafted, in principle, on any other MPEG-2 coding/decoding software. The software was run on the JETNET PC network at JET. The computing times presented in Table 1-2 are reported for this environment.

CONCLUSION
This paper proves the applicability of the compressed domain motion information extraction in the case of video cameras used on JET for fusion plasma diagnostic and control. The video sequences of images are stored in a compressed format for further analysis. In the case of the MPEG-2 format, crude motion estimation is performed for compression purposes. A set of methods, which exploit this information, have been adapted to the specific JET specific requirements. The methods achieve efficient motion segmentation and optical flow estimation. They can be used for the manipulation of the large JET video databases and even for real-time processing. The methods can be engrafted in the decompressing routines in case of analyzing stored video streams and in the compressing ones in case of real time estimation of the speed of moving objects.

The motion segmentation is appropriate for automatic detection of plasma instabilities which can trigger disruptions. This is exemplified by the automatic detection and tracking of MARFEs. The motion segmentation is also used as the main tool for optimizing the optical flow estimation. The time performances achieved allow real-time determination of the deuterium ice extrusion velocity, an important parameter for pellet injection.

ACKNOWLEDGEMENTS
This work was supported by the European Communities under the contract of Association between EURATOM and MEdC, ENEA and IPP and was carried out within the framework of the European Fusion Development Agreement. The views and opinions expressed herein do not necessarily reflect those of the European Commission.

REFERENCES


### Table 1: Time needed for MARFE detection and tracking.

<table>
<thead>
<tr>
<th>Image processing step</th>
<th>Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applying the regularization rules for the MV field</td>
<td>0.5</td>
</tr>
<tr>
<td>Median filtering</td>
<td>0.7</td>
</tr>
<tr>
<td>Dilation/Erosion</td>
<td>3.0</td>
</tr>
<tr>
<td>Labeling</td>
<td>2.4</td>
</tr>
<tr>
<td>Objection centroid determination</td>
<td>0.3</td>
</tr>
</tbody>
</table>

### Table 2: Time needed for ice extrusion velocity estimation.

<table>
<thead>
<tr>
<th>Image processing step</th>
<th>Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segmentation using information from MPEG video compressed domain</td>
<td></td>
</tr>
<tr>
<td>Applying the regularization rules for the MV field</td>
<td>0.6</td>
</tr>
<tr>
<td>Median filtering</td>
<td>0.8</td>
</tr>
<tr>
<td>Optical flow calculation performed on the segmented image region</td>
<td></td>
</tr>
<tr>
<td>Image derivatives</td>
<td>1.8</td>
</tr>
<tr>
<td>SOR iteration</td>
<td>4.1</td>
</tr>
<tr>
<td>Median filtering</td>
<td>0.3</td>
</tr>
</tbody>
</table>
Figure 1: AC[1] and AC[8] in the MB8 DCT block. Figure 2 illustrates the DCT coefficient of an 8 x 8 block. The top left coefficient is referred to as DC, which is the first DCT coefficient of each block and it is eight times the average intensity of the respective block [24].

Figure 2: Reference block (MB16ref), motion vector (v) and current block (MB16cur).

Figure 3: MARFE image sequence recorded during JET Pulse No: 70050. Each image has a size of 256 x 224 pixels.
Figure 4: The MV field obtained for the image sequence presented in Fig. 3 after applying Coimbra's regularization rules. The size of each MV field is 16 x 14 pixels.

Figure 5: The MV field obtained after median filtering, dilation/erosion, labeling with the selection of the image object with the largest size and binarization.
Figure 6: MARFE detection and tracking using the MPEG MV field and simple image processing.

Figure 7: Different frames from the image sequence showing the extruded deuterium ice in case of JET Pulse No:76379. Each image has a size of 223 x 177 pixels.
Figure 8: Illustration of the motion activity segmentation.

Figure 9: Illustration of the optical flow estimation. Two successive images from the video sequence corresponding to the JET Pulse No: 76379 are presented together with the region of interest (ROI) determined by the segmentation procedure (left). The velocity field (middle) and the image difference $I_{diff}$ (right) are calculated only inside ROI.