Scene Understanding (2)

perception, multi-sensor fusion, spatio-temporal reasoning
and activity recognition.

Francois BREMOND

PULSAR project-team,
INRIA Sophia Antipolis, FRANCE

Francois.Bremond@sophia.inria.fr
http://www-sop.inria.fr/pulsar/

Key words: Artificial intelligence, knowledge-based systems, cognitive vision, human behavior representation, scenario recognition

Event Recognition

Outline:

- Event Representation
- Temporal Scenario Recognition
  - Scenarios representation
  - recognition process
- Results: recognition of several scenarios

Event Representation

Several entities are involved in the scene understanding process:

- **Moving region**: any intensity change between images.

- **Context object**: predefined static object of the scene environment (entrance zone, wall, equipment, door...).

- **Physical object** : any moving region which has been tracked and classified (person, group of persons, vehicle, … etc).

- **Physical object of interest**: meaningful object, but depending on applications (person/ door, parked vehicle, … etc).

Event Representation

Events and scenarios : **large variety**

- more or less composed of sub-events (running/fighting).
- involving few/many actors (football game).
- general (standing)/sensor and application/view (sit down, stop) dependent.
- spatial granularity: the view observed by one camera/the whole site.
- temporal granularity: instantaneous/long term with complex relationships (synchronize).

➢ 3 levels of complexity depending on the complexity of temporal relations and on the number of physical objects :

- non-temporal constraint relative to one physical object (sitting). Intuitive association of probabilities to get precision.
- temporal sequence of sub-scenarios relative to one physical object (open the door, go toward the chair then sit down).
- complex temporal constraints relative to several physical objects (A meets B at the coffee machine then C gets up and leaves). Need of logic expressiveness but sensitive to vision errors.

Event Recognition

Several formalisms can be used:

- **Event representation**:
  - n-ary tree, frame, aggregate (structure).
  - finite state automaton, sequence (evolution).
  - graph, set of constraints.
- **Event recognition**:
  - feature based routine.
  - Classification, Bayesian, neural network, SVM, clustering.
  - DBN, HMM.
  - Petri net, stochastic grammar,
  - constraint propagation, verification of temporal constraints.

Event Representation

Video events: real world notions corresponding to short actions up to activities.

- *Primitive State:* a spatio-temporal property linked to vision routines involving one or several actors, valid at a given time point or stable on a *time interval*

  Ex : « close», « walking», « sitting»

- *Composite State: a combination of primitive states*
- *Primitive Event:* significant change of states

  Ex : « enters», « stands up», « leaves »

- *Composite Event:* a combination of states and events. Corresponds to a long term (symbolic, application dependent) activity.

  Ex : « fighting», « vandalism»

Event Representation

A video event is mainly constituted of five parts:

- Physical objects: all real world objects present in the scene observed by the cameras
  - Mobile objects, contextual objects, zones of interest
- Components: list of states and sub-events involved in the event
- Forbidden Components: list of states and sub-events that must not be detected in the event
- Constraints: symbolic, logical, spatio-temporal relations between components or physical objects
- Action: a set of tasks to be performed when the event is recognized

Event Representation

Representation Language to describe Temporal Events of interest.

Example: a “Bank_Attack” scenario model

composite-event (***Bank_attack***,
  physical-objects ((employee : **Person**), (robber : **Person**))
  components(
    (e1 : primitive-state ***inside_zone*** (employee, "Back"))
    (e2 : primitive-event ***changes_zone*** (robber, "Entrance", "Infront"))
    (e3 : primitive-state ***inside_zone*** (employee, "Safe"))
    (e4 : primitive-state ***inside_zone*** (robber, "Safe")))
  constraints ((e2 ***during*** e1)
    (e2 ***before*** e3)
    (e1 ***before*** e3)
    (e2 ***before*** e4)
    (e4 ***during*** e3))
  action (“Bank attack!!!”))

Scenario Representation

A "Bank attack" scenario instance

Event Representation: Video Event Ontology

- Video Event Ontology: a set of concepts and relations is used as a reference between all the actors of the domain to describe knowledge

- Enable experts to describe video events of interest (e.g. composite event) and to structure the knowledge: ontology of the application domain.
- Share knowledge between developers: ontology of visual concepts (e.g. a stopped mobile object)
- Ease communication between developers and end users and enable performance evaluation: ontology of the video understanding process (what should be detected: mobile object (a parked car), object of interest (a door), visible object (occluded person))

- Architecture interoperability: separation between specification and knowledge description

Scenario Recognition: Specific Routines
Advisor project: F. Cupillard, A. Avanzi,…

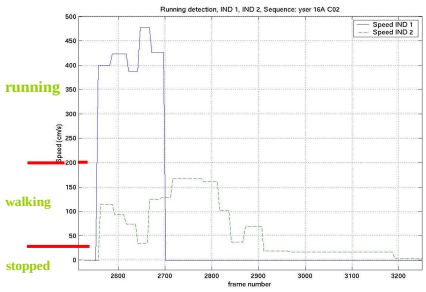
Scenario Recognition: Specific Routines
Results in metro station

Scenario Recognition : Running

Scenario:Running -> ALARM

State: walking

State: stopped

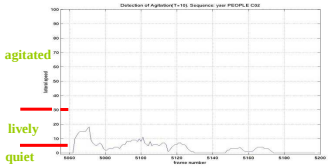


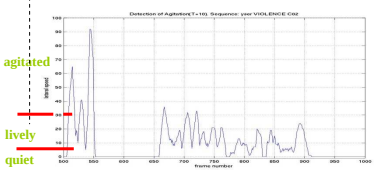
Scenario Recognition: Specific Routines

Results in metro station

Scenario:Agitated Behaviour
-> ALARM

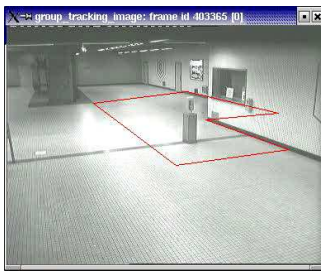
State: Lively





Scenario Recognition: automaton

The scenario "A Group of people blocks an Exit" is based on a Finite state automaton

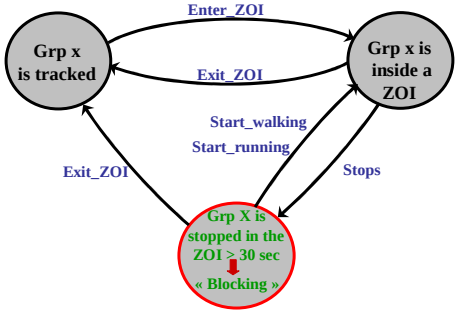


Mobile objects Detection

Group Tracking

Blocking

Recognition of the behaviour « a Group of people blocks an Exit»



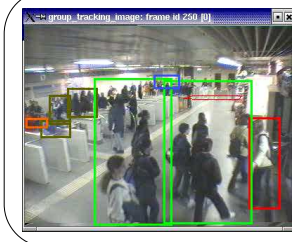
Scenario Recognition: automaton

Group behavior

Blocking

Group behavior

Fighting



Crowd behavior

Overcrowding

Individual behavior

Jumping over barrier

Scenario Recognition : automaton

- Recognition of five behaviors "Blocking", "Fighting", "Jumping over barrier", "Vandalism" and "Overcrowding"
- Tested on 50 metro sequences (10 hours) and one week live recognition
- True positive per sequence: 70% ("Fighting") to 95% ("Blocking")
- False positive per sequence: 5% ("Fighting", "Jumping over barrier") to 0% (others)

Scenario Recognition: Temporal Constraints

Work done in collaboration with T. Vu

Scenario Recognition: Temporal Constraints

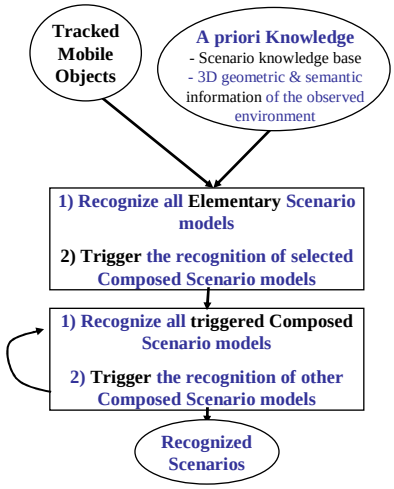
- Overview of the recognition process
- Recognition of elementary scenarios
- Scenario compilation
- Recognition of composed scenarios
- Prediction and uncertainty
- Example of the recognition of a “Bank attack” scenario and more…

Scenario Recognition: Temporal Constraints
(T. Vu)

- **Scenario** (algorithmic notion): any type of video events

- Two types of scenarios:
  - **elementary** (primitive states)
  - **composed** (composite states and events).

- Algorithm in two steps.



Scenario Recognition: Elementary Scenario

- The recognition of an elementary scenario model m_e consists of a loop:

  1. Choosing a physical object for each physical-object variable

  2. Verifying all constraints linked to this variable

  m_e is recognized if all the physical-object variables are assigned a value and all the linked constraints are satisfied.

Scenario Recognition: Composed Scenario

- **Problem**:

  given a scenario model $m_c = (m_1$ *before* m_2 *before* $m_3)$;

  if a scenario instance i_3 of m_3 has been recognized

  then the main scenario model m_c may be recognized.

  However, the classical algorithms will try all combinations of scenario instances of m_1 and of m_2 with i_3

  ➔ a combinatorial explosion.

- **Solution**:

  decompose the composed scenario models into simpler scenario models in an initial (compilation) stage such as each composed scenario model is composed of two components: $m_c = (m_4$ *before* $m_3)$

  ➔ a linear search.

Scenario Recognition: Composed Scenario

Example: original "Bank_attack" scenario model

```
composite-event(Bank_attack,
   physical-objects((employee : Person), (robber : Person))
   components(
      (1)   (e1 : primitive-state inside_zone(employee, "Back"))
      (2)   (e2 : primitive-event changes_zone(robber, "Entrance", "Infront"))
      (3)   (e3 : primitive-state inside_zone(employee, "Safe"))
      (4)   (e4 : primitive-state inside_zone(robber, "Safe"))   )
    constraints((e2  during  e1)
                (e2  before  e3)
                (e1  before  e3)
                (e2  before  e4)
                (e4 during  e3)  )
      alert("Bank attack!!!") )
```

Scenario Recognition: Composed Scenario

Compilation: Original scenario model is decomposed into 3 new scenarios

```
composite-event(Bank_attack_1,
   physical-objects((employee : Person), (robber : Person))
   components(
   (1)      (e1 : primitive-state inside_zone (employee, "Back"))
   (2)      (e2 : primitive-event changes_zone (robber, "Entrance", "Infront"))
   constraints((e1 during e2)   ))

composite-event( Bank_attack_2,
   physical-objects((employee : Person), (robber : Person))
   components(
    (3)      (e3 : primitive-state inside_zone (employee, "Safe"))
    (4)      (e4 : primitive-state inside_zone (robber, "Safe"))     )
   constraints((e3 during e4)    ))

composite-event( Bank_attack_3,
   physical-objects((employee : Person), (robber : Person))
   components(
            (att_1 : composite-event Bank_attack_1 (employee, robber))
            (att_2 : composite-event Bank_attack_2 (employee, robber))    )
   constraints(((termination of att_1) before (start of att_2))   )

   alert("Bank attack!!!")      )
```


Scenario Recognition: Composed Scenario

- A compiled scenario model m_c is composed of two components: start and termination.
- To start the recognition of m_c, its termination needs to be already instantiated.
- The recognition of a compiled scenario model m_c consists of a loop:
  1. Choosing a scenario instance for the start of m_c,
  2. Verifying the temporal constraints of m_c,
  3. Instantiating the physical-objects of m_c with physical-objects of the start and of the *termination* of m_c,
  4. Verifying the non-temporal constraints of m_c.
  5. Verifying forbidden constraints

Uncertainty Representation

PrimitiveState (**Person_Close_To_Vehicle**,
Physical Objects ((p : Person, 0.7), (v : Vehicle, 0.3))
Constraints ((p distance v ≤ close_distance)
(recognized if likelihood > 0.8)))

CompositeEvent (**Crowd_Splits**,
Physical Objects ((c1: Crowd, 0.5), (c2 : Crowd, 0.5), (z1: Zone))
Components ((s1 : CompositeState Move_toward (c1, z1), 0.3)
(e2 : CompositeEvent Move_away (c2, c1), 0.7))
Constraints ((e2 during s1)
(c2's Size > Threshold)
(recognized if likelihood > 0.8)))

Scenario recognition: Results

Evaluation: the experts of 20 projects in video interpretation have realized three types of tests.

- on recorded videos: to verify whether the recognition algorithm can recognize effectively scenario occurrences (correct detections).
- on live videos: to verify whether the recognition algorithm can work on a longtime interval (no false alarms).
- on recorded/simulated videos: to estimate the processing time and efficiently of the recognition algorithm.

Scenario recognition: Results

Bank agency monitoring in Paris (M. Maziere)

Scenario recognition: Results

Vandalism scenario example (temporal constraints) :

Scenario(vandalism_against_ticket_machine,

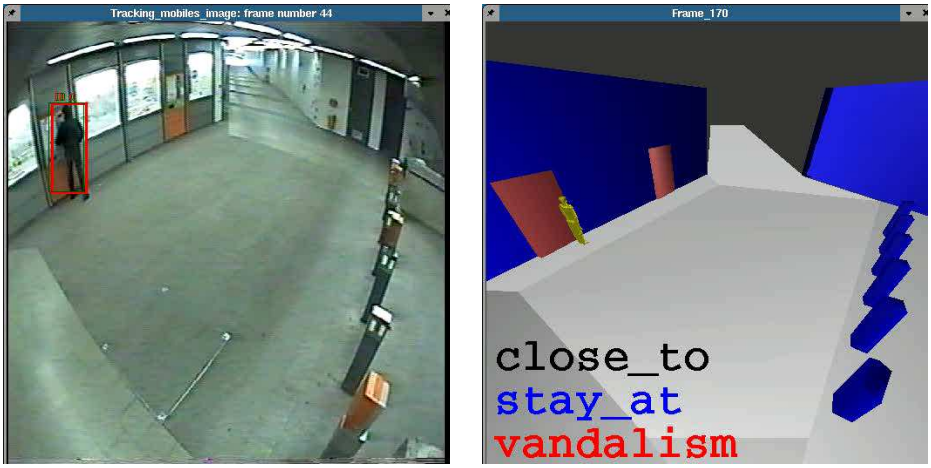
Physical_objects((p : **Person**), (eq : **Equipment**, ***Name***="Ticket_Machine"))

Components ((event s1: *p* **moves_close_to** *eq*)
(state s2: *p* **stays_at** *eq*)
(event s3: *p* **moves_away_from** *eq*)
(event s4: *p* **moves_close_to** *eq*)
(state s5: *p* **stays_at** *eq*))

Constraints ((s1 **!=** s4) (s2 **!=** s5)
(s1 **before** s2) (s2 **before** s3)
(s3 **before** s4) (s4 **before** s5))))

Scenario Recognition: Results

Vandalism in metro in Nuremberg



Scenario recognition: Results

Example: a “Vandalism against a ticket machine” scenario

Scenario recognition: Results

Example: a “abandoned goods” scenario

Scenario recognition:

Results Example: “Unloading Front Operation ” event

- Example of the **Unloading Front Operation** (global)

CompositeEvent (UnLoading_Front_Global_Operation,
PhysicalObjects ((v1 : Vehicle), (v2 : Vehicle),
(z1 : Zone), (z2 : Zone), (z3 :Zone))
Components ((c1 : **CompositeEvent** Loader_Arrival(v1, z1, z2))
(c2 : **CompositeEvent** Transporter_Arrival(v2, z1, z3))
Constraints ((v1->SubType = LOADER)
(v2->SubType = TRANSPORTER)
(z1->Name = ERA)
(z2->Name = RF_DoorC_Access)
(z3->Name = LOADER_BackZone)
(c1 before c2)))

Scenario recognition: Results

Example: "Unloading Global Operation" event

- **"Unloading Global Operation"**

Scenario recognition: Results

Example: "Unloading Front Operation " event

- Example of the **Unloading Front Operation** (detailed)

```
CompositeEvent (UnLoading_Front_Detailed_Operation,
  PhysicalObjects ( (p1 : Person), (v1 : Vehicle), (v2 : Vehicle), (v3 : Vehicle),
                    (z1 : Zone), (z2 : Zone), (z3 :Zone), (z4 : Zone))
   Components ( (c1 : CompositeEvent Loader_Arrival(v1, z1, z2))
                (c2 : CompositeEvent Transporter_Arrival(v2, z1, z3))
              (c3 : CompositeState Worker_Manipulating_Container(p1, v3, v2, z3, z4)))
  Constraints ( (v1->SubType = LOADER)
                (v2->SubType = TRANSPORTER)
                (z1->Name = ERA) (z2->Name = RF_DoorC_Access)
                (z3->Name = LOADER_BackZone)
                (z4->Name = Behind_RF_DoorC_Access)
                (c1 before c2)
                (c2 before c3)))
```

Scenario recognition: Results

Parked aircraft monitoring in Toulouse (F Fusier)

- **"Unloading Front Operation"**

SCENARIO UNLOADING_DETAILED_OPERATION

PHYSICAL OBJECTS :
VEHICLES : {Loader, Transporter}
PERSONS : {Worker}
STATIC ZONES : {ERA}
AIRCRAFT ZONES : {Front_Unloading_Area, Baggages_Unloading_Area}
DYNAMIC ZONES : {Transporter_Parking_Area}

VIDEO EVENTS :
Loader_Arrival
Transporter_Arrival
Worker_Arrived
Worker_Manipulating_Container

Scenario recognition: Results

Example: "Aircraft Arrival Preparation " event

- Aircraft Arrival Preparation (involving the GPU)

SCENARIO AIRCRAFT_ARRIVAL_PREPARATION_SCENARIOS
Vehicle: GPU
Person: Handler
Zones: ERA, GPU_Access, Arrival_Preparation
Dynamic Zone: GPU_Door

Vehicle_Arrived_In_ERA
Gpu_Enters_Gpu_Access_Area
Gpu_Stopped_In_Gpu_Access_Area
Handler_Gets_Out_Gpu
Handler_From_Gpu_Deposites_Chocks_Or_Stud

Scenario recognition: Results

Example: “Tow Tractor Arrival” event

- Tow Tractor Arrival

Scenario recognition: Results

Example: “vandalism_against_window” event

CompositeEvent(***vandalism_against_window***,

PhysicalObjects((vandal : Person)), (w : Equipment))

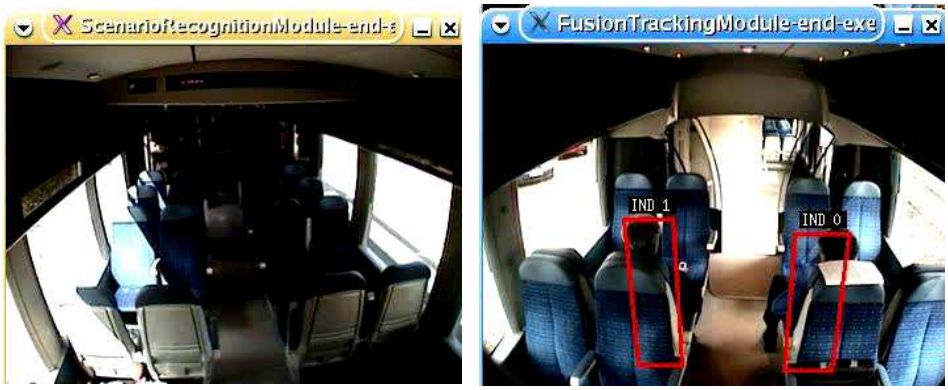
Components((vandalism_against_window_VIDEO :
CompositeEvent vandal_close_to_window(vandal, w))
(vandalism_against_window_AUDIO :
CompositeEvent tag_detected_close_to_person(vandal)))

Constraints((vandalism_against_window_VIDEO ***during***
vandalism_against_window_AUDIO))

Alarm(AText("Vandalism against window")
AType("URGENT")))

Scenario recognition: Results

Example: “Scratch & theft in a train” scenarios

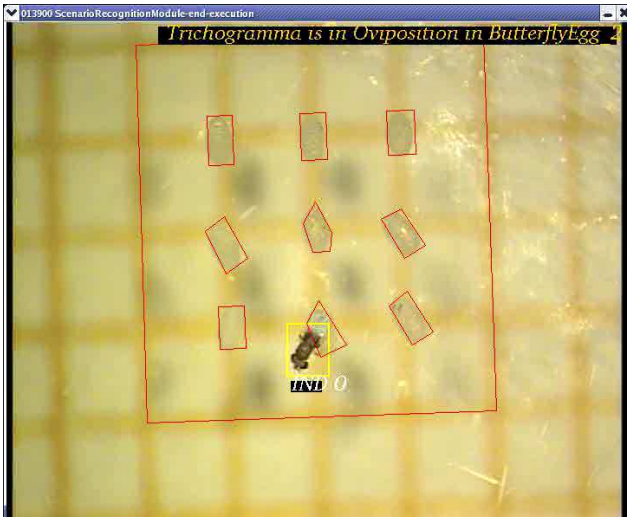


Scenario recognition: Results

Example: a “Disturbing people in a train” scenario

Scenario recognition: Results

Video Understanding for Trichogramma Monitoring

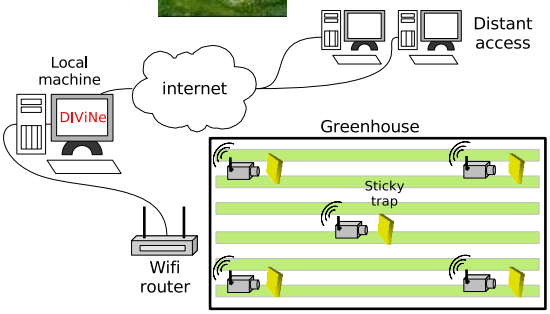


Scenario recognition: BIOSERRE

Early detection of infestation in greenhouses in the framework of Integrated Pest Management Methods

Whitefly
curiosity : Inra (Brun)

Greenfly
curiosity: Inra (Brun)



Wifi Video Camera sensing a sticky trap (Curiosity : PULSAR/INRIA)

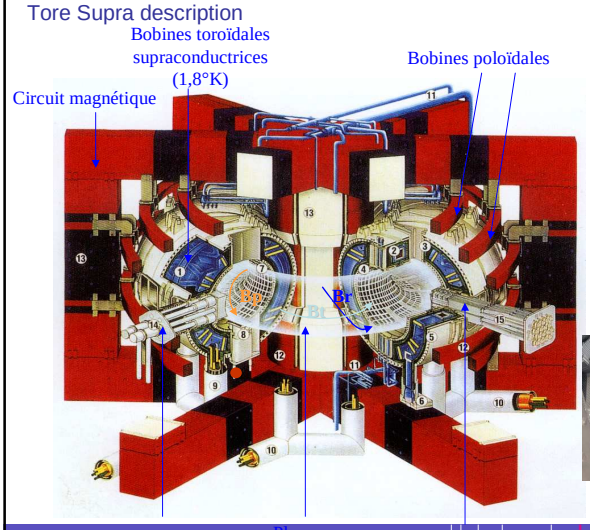
Scenario recognition: BIOSERRE

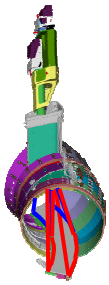
Early detection of infestation in greenhouses

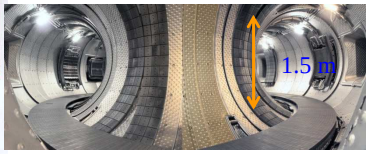
INRIA SOPHIA ANTIPOLIS

Real-Time Monitoring of Imaging Diagnostics Applied to Tore Plasma Operation (ITER)

Tore Supra description







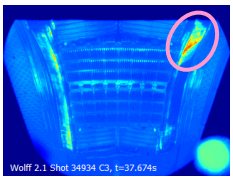
INRIA SOPHIA ANTIPOLIS

Real-Time Monitoring of Imaging Diagnostics Applied to Tore Plasma Operation (ITER)

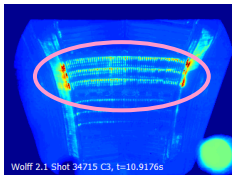
Goal 1: real time detection of thermal events for the protection of plasma facing components

Goal 2: multimodal knowledge discovery for modelling precursors of critical thermal events (e.g. disruptions)

Thermal Events:

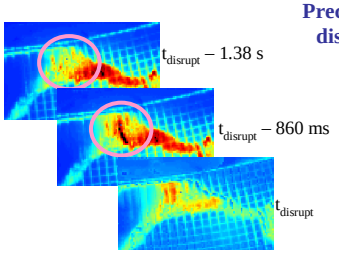


Hot Spot



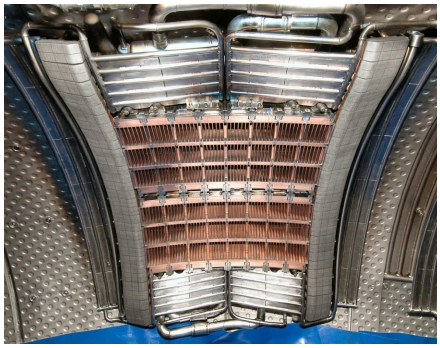
Electrical Arcs

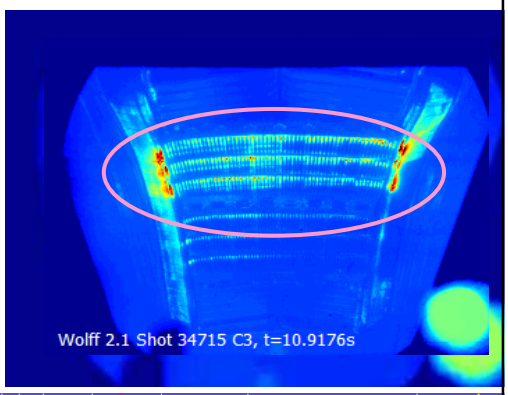
Precursors of disruption



Real-Time Monitoring of Imaging Diagnostics Applied to Tore Plasma Operation (ITER)

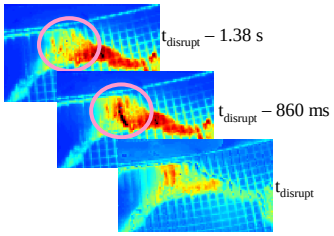
Thermal Events: Electrical Arcs on heating antenna

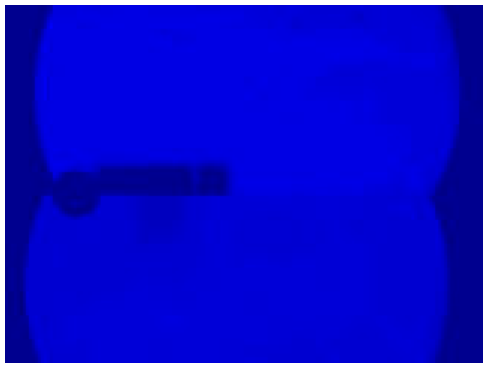




Real-Time Monitoring of Imaging Diagnostics Applied to Tore Plasma Operation (ITER)

Thermal Events: **Precursors of disruption**





Real-Time Monitoring of Imaging Diagnostics Applied to Tore Plasma Operation (ITER)

Thermal Events: **UFO with visible, ultra fast camera (15000 frame per s.)**