Jornada de Seguimiento de Proyectos, 2005
Programa Nacional de Tecnologías Informáticas

PASSIVIM: Panoramic based Summarization and Semantic Interaction with Digital Video for Multimedia Messages TIC2003-06075

Xavier Binefa Valls [*]
Computer Science Department
Universitat Autònoma de Barcelona

Abstract

This project aims to summarize the events present in clips coming from sports or news sequence as an augmented image. The summarization will be adapted to the capabilities of the multimedia mobile devices and networks. The user shall be provided with capabilities to view the information as a current video clip but he also will be able to retrieve semantically related visual information. To reach the proposed objectives, this project shall start from a clip acquired with a fixed camera. The project is divided into three main tasks, which are : **a)** Clustering and identification of trajectories generated from the different moving regions within the scene, **b)** representation of the clip with a panoramic image including information from the dynamic regions and from the camera movements (pan, tilt, zoom) enabling the reconstruction of the video clip at the user side and **c)**, take advantage of the representation obtained in *b* in order to classify and structure a sports video sequence by events of interest (goals, etc.). Finally, the project will add user interacting capabilities using the representations mentioned above.

Keywords:

1 Project Aims

The main goal of the PASSIVIM project is to apply computer vision technologies to develop new ways of transmission of video sequences, resulting in the bandwith saving by transforming the sequences into panoramic images while keeping track of all the information about the moving regions in the video. This allows to transmit the video as an image and a set of information about the motion of the camera and the objects, as well as permits also the inclusion of enhanced features, such as relevance-based highlight detection in videos and interactive capabilities. The main objectives of the project are next detailed :

[*] Email: xavier.binefa@uab.es

1.1 Clustering and identification of motions in the sequence

Motion analysis in video sequences is the cornerstone of the PASSIVIM project. The accuracy and correctness of the motion modelling of the sequence dominant and local motions is directly related to the quality of the augmented image. Therefore our very first aim is to achieve a reliable analysis of all the existing motions along the sequence. This first goal of the project is divided in several tasks.

1.1.1 Identification of different motions in a video sequence

Firstly, the different moving regions of the sequence that obey the same motion model must be identified and separed. In order to do so, algorithms regarding motion vector fields and non parametric clustering have to be developed or adapted to our problem. The frames of the video sequence must be processed in order to find motion descriptors of the regions of the sequence; these descriptors will provide the source data to perform the identification and clustering of the moving regions.

1.1.2 Motion modelling in video sequences

The identification of these regions allows us to infer the followed motion model from the motion descriptors (vectors) in terms of a set of parameters for each of the moving regions. These regions have to be tracked for a futher use in the final transmission and representation of the video sequence so tracking and parameter estimation must be used and/or developed to extract an accurate representation of each region motion and trajectory.

1.1.3 Object tracking and trajectories

Each of the objects is analyzed throughout the sequence in order to extract the trajectory of the players during a clip. This extraction provides further features such as an improved video browsing or an enhanced sports analysis, but yields new and complex problems like collisions handling when occlusions occur which are not trivial and must be addressed.

1.2 Representation of the sequence using a panoramic image

Once the motion regions have been identified the dominant motion is assumed to be the camera motion, that is to say that the largest region of the frames that follow the same motion model is considered the background, static in most of the cases. The parametric representation obtained in the previous step helps inferring the transformation between each pair of frames in order to warp all the frames to build the panoramic image.

1.2.1 Action localization

One of the enhanced features sought to achieve by our project is to know the context of a sequence in each time instant. That means to know where in the match field the action is taking place. In order to do so, an algorithm to perform global mosaic image registration must be developed so the location of each one of the frames of the sequence in the match field can be obtained.

1.3 Addition of semantic information to the processed sequence

Taking advantage of both de motion descriptors and the parametric models extracted in the previous task, context-related information such as relevance ratio can be included to the transmission to enable further enhanced features. The main aim of this task is to analyze the retrieved information in order to provide this semantic information to the sequence, using mainly the motion of the camera (analyzing when a zoom occurs, or when the camera is pointing to a goal position in the field, for example).

1.3.1 Abstract model

A very useful application of all the previous analysis (panoramic image construction, object trajectory extraction and contextual localization) is to generate an abstract model (a match field schema) and represent all the trajectories of the objects into it. It proves to be very useful for further applications, such as match analysis, but an abstract model also becomes a crucial tool when working in multicamera environments, because all the views can be synchronized using the generated artificial field.

1.4 Transmission and user interaction

Finally, when the sequence has been processed, our objective is to show the advantadges of our work by transmitting it into a mobile device (mobile phone or palm) and to compare the bandwith required when applying our approach and when transmitting normal video. Moreover, our aim is to provide a prototype where all the motion analysis performed can be exploited in form of enhanced capabilities, such as see the players trajectories or watch the highlights of a complete game automatically.

1.4.1 Optimized transmission

The usefulness of panoramic image from a video sequence in terms of bandwidth saving is far proven, as the MPEG-4 standard also takes advantage of this fact. The whole background of a video sequence can be transmitted as a single image, with only the moving regions and players remaining to be transmitted as video.

1.4.2 Multicamera environments

As mentioned above, having an abstract model of the field and player trajectories allows us to synchronize multiple views between them, using the positions in the artificial model as a nexus. The main aim of having multiple views synchronized is to offer, for example, interactive replays of the most interesting game moments, as well as a user-friendly reverse angle, providing the user the possibility to switch between cameras when there are occlusions and multicamera availability.

1.4.3 Automatic Highlight detection and user-friendly video browsing

The processes mentioned above can help when detecting automatically the most interesting plays of the game. A slow motion replay, a camera zoom, or the proximity of the action to one of the goals are indicators of relevance. These indicators are all detected by the previous tasks and stored as a handful of parameters, easily findable in a standard database. One of the objectives of the project is to offer a highlight-based sequence summary to the end user, taking advantage of these relevance indexs.

1.5 Project Progress

Table 1 summarizes the status of the project. In general terms, most of the project goals regarding computer vision algorithmic have been achieved, while the end-user part remains to be done in 2006. It is obviously the last part of the project because it depends on the reliability of the tasks developed in the steps 1.1, 1.2 and 1.3.

Progress of the Project Objectives			
Ref.	**Description**	**Status**	**Remarks and Schedule**
1.1	**Clustering and motion Identification**	**100%**	
1.1.1	Identification of different motions in a video sequence	100%	Goal achieved
1.1.2	Motion modelling in video sequences	100%	Goal achieved
1.1.3	Object tracking and trajectories	100%	Goal achieved
1.2	**Representation of the sequence using a panoramic image**	100%	Goal achieved
1.2.1	Action localization	100%	Goal achieved
1.3	**Addition of semantic information to the sequence**	50%	Currently working
1.3.1	Abstract model	100%	Goal achieved
1.3.1	Relevance Index Extraction	30%	Researching
1.4	**Transmission and user interection**	0%	2006
1.4.1	Optimized transmission	0%	2006
1.4.2	Multicamera environments	0%	2006
1.4.3	Automatic Highlight detection and user-friendly video browsing	0%	2006

Table 1: Status of the project goals at August, 2005

2 Tasks and development

In this section we detail the progress of the tasks related to each of the project goals, as well as problems encountered, changes and a brief summary of the techniques used or developed.

2.1 Clustering and identification of motions in the sequence

The objectives of this part have been acomplished. We have had to change the initial formulation in order to achieve a more reliable one, but our group has developed efficient techniques regarding identification and clustering of moving regions.

The starting point in order to solve this task was the polynomial fiber models presented in [1], but the lack of reliability of this approach in singular cases that involved zooms forced us to find an alternate method to identify different moving regions or objects. We explored an alternative by combining the optical flow techniques and the robust parameter estimators. With these methods we are able to extract an accurate parametric description of the camera motion that is fundamental to obtain a correct panoramic image.

The motion vector field proves to be a valid descriptor in order to cluster different moving regions. We first employ a traditional robust estimator (*Least Median of Squares*, LMedS henceforth) to describe the dominant model by a handful of parameters. We use a reduced affine model, that takes 4 parameters to describe the dominant motion of the sequence. From that point, different regions of interest that follow different motions can be isolated by studying the residual space of the vectors that do not follow the dominant motion model in terms of the module and angle of its difference vector. This method was presented in [10],[7] and [9].

Once the different regions are identified, an accurate tracking of the objects is performed using an adaptation of a mean-shift based algorithm presented in [13]. Problems arise when we deal with sequences that present occlusions between players, so we improved the tracking algorithms with a sports sequence management method presented in [6], capable of handling collisions.

2.2 Representation of the sequence using a panoramic image

This part is the most important of the work because it allows all the further processing. The goals have been achieved, our group has published several methods to perform robust real-time panoramic images building.

It is very costly to implement a full mosaicing algorithm that relies on a dense motion vector field. Calculating the dense field and estimating the camera parameters for each frame is an overwhelming task that cannot be carried out at real time. For this reason, our group has become expert in dealing with compressed domain data. Compressed domain, or MPEG data provides a pre-calculated sparse motion vector field that allows avoiding the costly process of calculating a dense motion vector field and estimating a parameter model from thousands of samples, permitting to extract motion descriptios and build panoramic images at real time. Mentioned papers [10],[7] and [9] also contain parts referring the panoramic image construction technique.

Experimental results have shown that our latter approach outperforms by far the starting polynomial fiber method. In figure 2 results of the panoramic image building are shown.

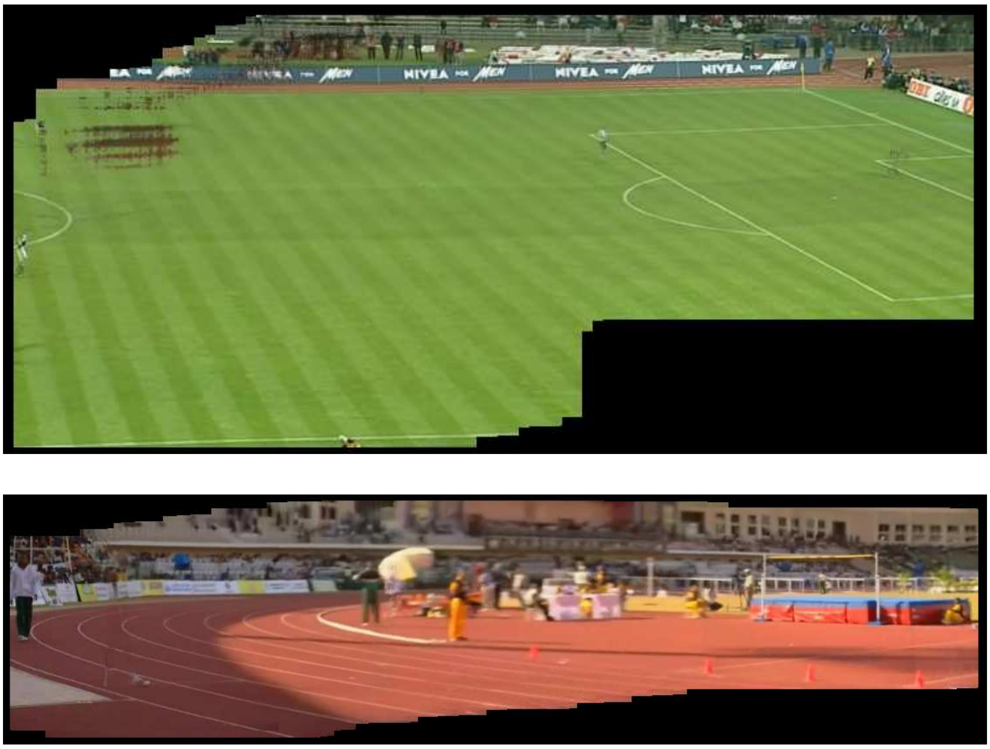

Figure 1: Resulting panoramic images of the application of our robust statistics algortihm to sports sequences. The lines of the field are kept perfectly straight, proving that our mosaicing technique performs flawlessly in sports sequences.

We have developed improved mosaicing algorithms by taking advantage of highly robust parameter estimators such as the *vbMDPE*[14] and a novel idea based on imaginary straight lines tracking. Our methods have been published in several conferences ([8], [4]). Our work with robust estimators has also resulted in publications in major conferences [2].

2.3 Addition of semantic information to the processed sequence

We are currently working in this task.

Algorithms of modeling the match field using a synthetic model have been developed. We are now able to generate an artificial model of the field with all the objects' trajectories embedded. This work was presented in [6].

We are currently working on extracting relevance indexs of the plays to enable a further automatic selection of highlights.

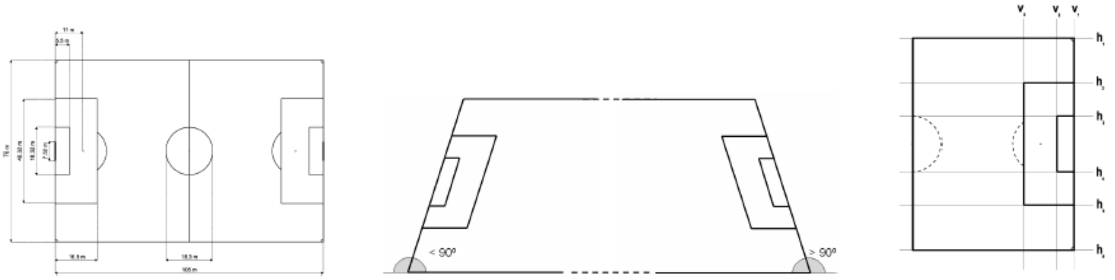

Figure 2: Schema of relationship between a match field and our abstract model.

2.4 Transmission and enhanced features of the processed sequence

This is the final part of the project, and it is scheduled for the next year. We depend on the results achieved in the previous part of the project, but the first results are promising and we think we may be able to accomplished the goals we propose :

- **An intelligent system for browsing the video sequences**, offering the possibility to overview a complete sequence by watching only the highlights, automatically extracted.
- **User interaction**. Our system will provide mechanisms for the user to choose the camera from where the action is being shown.
- **Optimized Transmission**. We aim to take advantadge of the process applied to the sequence by using the panoramic image to transmit video, with the consequent bandwidth saving.

3 Relevant scientific-technical results

3.1 Personnel involved in the project

Master Students

Luis Ferraz
Sira Ferradans (Master Thesis at July, 2006)
Brais Martínez (Master Thesis at July, 2006)

PhD. Students

Juan M. Sánchez (Ph.D November, 2003)
Xavier Orriols (Ph.D February, 2004)

Lluis Barcelo (Ph.D Thesis in December, 2005)
Ramon Felip (Ph.D Thesis in December, 2006)

3.2 Publications

3.2.1 PhD. Thesis

- *Multiple Feature temporal models for the semantic characterization of video contents*, Juan Maria Sanchez Pujades, December, 2003
- *Generative Models for Video Analysis and 3D Range Data Applications*, Xavier Orriols i Majoral, February 2004

3.2.2 Master Thesis

- *Application of Robust Statistics and Mean Shift Analysis to Digital Video and LADAR Imaging*, Ramon Lluís Felip Rodríguez, September, 2004

3.2.3 Journal Papers

- R. Toledo, X. Orriols, J. Sanchez, X. Binefa Automatic Cataloguing of Advertisement in Magazines. *Multimedia Tools and Applications (In Press, from July 04)*
- L. Barceló, R. Felip, X. Binefa. Robust Dominant Motion Estimation using MPEG Information and a Helmoltz Principle based Parameter Estimator. *Pattern Recogntion Journal (Submitted July 05)*

3.2.4 Conference Papers

- R.L. Felip, X. Binefa and J. Diaz. A New Parameter Estimator Based on The Helmholtz Principle. *IEEE Conference on Signal Processing (ICIP'05), Genova, September 2005 (In Press)*
  In this paper we present a novel parameter estimator that follows a general perception law as

an objective function, the Helmoltz principle. We present an interpretation of this postulate in statistical terms and apply it along with low order statistics in order to obtain a flexible parameter estimator with a high breakdown point, capable of suiting the specific needs of each problem. The performance of the algorithm is tested applying the estimator to find lines in synthetic data with up to 90% of noise and real LADAR images

- L. Barceló, R.L. Felip and X. Binefa. A new Approach for Real Time Motion Estimation Using Robust Statistics and MPEG Domain Applied to Mosaic Images Contruction. *IEEE conference on Multimedia and Expo., Amsterdam July, 2005 (In Press)*
  Dominant motion estimation in video sequence is a task that must be often be solved in Computer Vision problems but involves a high computational cost due to the overwhelming amount of data to be treated when working in image domain. In this paper we introduce a novel technique to perform motion analysis in video sequences taking advantage of the motion information of MPEG streams and its structure, using imaginary line tracking and robust statistics to overcome the noise present in compressed domain information. In order to demonstrate the reliability of our new approach, we also show the results of its application to mosaic image construction problem.

- J.M. Sanchez, R. L. Felip and X. Binefa Preatentional Filtering in Compressed Video. *IEEE conference on Multimedia and Expo, Amsterdam July 2005 (In Press)*
  We propose the use of attentional cascades based on the DCT and motion information contained in an MPEG coded stream. An attentional cascade is a sequence of very efficient classifiers that reject a large number of negative candidate regions, while keeping all the positive candidates. Working directly on the compressed domain has two main advantages: computationally expensive features are already computed, and the stream is only partially decoded without the additional cost of full decompression, which will be reached by a very small number of the initial candidate regions. We have applied these concepts to skin color detection, as a pre-attentive filtering prior to face detection, and to text region detection with particular focus on license plates for vehicle identification. In both cases, a reduction of the number of candidate regions close to 95% is achieved, which turns into an enormous performance increase in video indexing processes.

- LL. Barceló, X. Binefa and J.R. Kender. Robust Methods and Representations for Soccer Player Tracking and Collision Resolution. *4th International Conference on Image and Video Retrieval (CIVR'05), Singapore, July 2005. (In Press in LNCS)*
  We present a method of tracking multiple players in a soccer match using video taken from a single fixed camera with pan, tilt and zoom. We extract a single mosaic of the playing field and robustly derive its homography to a playing field model, based on color information, line extraction, and a Hausdorff distance measure. Players are identified by color and shape, and tracked in the image mosaic space using a Kalman filter. The frequent occlusions of multiple players are resolved using a novel representation acted on by a rule-based method, which recognizes differences between removable and intrinsic ambiguities. We test the methods with synthetic and real data.

- R. L. Felip, J. M. Sanchez and X. Binefa. Video summarization with moving objects in the compressed domain. *Proc. Of the IST/SPIE Conference Internet Imaging VI , Vol 5670, San Jose, pp. 143-154. January 2005*

The vast amount of video sequences avaliable in digital format presents considerable challenges for descriptor extraction and information retrieval. The dominant motion in a video scene proves to be very important to characterize video sequences, but the cost to compute it is high when working in image domain. In this paper we present a method to extract an affine description of the global motion of a video sequence using a robust estimator and data from the compressed domain, where the motion vector field is already calculated. We also introduce a criterion to measure the reliability of the affine model estimated. Finally, we apply our approach to motion-based segmentation of video sequences and video summarization.

- LL. Barcelo, X. Binefa. Contextual Soccer Detection Using Mosaicing Techniques. *4th Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA2005), pp. 77-84. Springer-Verlag LNCS 3522. June 2005, Estoril, Portugal*
  Sports Video understanding aims to select and sumaries important video events that occur in only special fragments of the whole sports video. A key aspect to this objective is to determine the position in the match field where the action takes place, that is, the location context of the play. In this paper we present a method to localize where in the match field the play is taking place.We apply our method to soccer videos, although the method is extensive to other sports. The method is based on constructing the mosaic of the first sequence that we process: this new mosaic is used as a context mosaic. Using this mosaic we register the frames of the other sequences in order to put in correspondence all the frames with the context mosaic, that is, put in context any play. In order to construct the mosaics, we have developed a novel method to register the soccer sequences based on tracking imaginary straight lines using the Lucas-Kanade feature tracker and the vb-QMDPE robust estimator.

- R. L. Felip, X. Binefa. Affine Description of Motions in Compressed Domain. *Pp: .93-100 In Recent Advances in Artificial Intelligence Research and Development. Frontiers in Artificial Intelligence and Applications, Ed. IOS Press. 2004.*
  This paper considers the problem of finding a robust solution to the parametric description of multiple motions in video sequences problem, using compressed domain data. Here, we present an effective way to recover an affine global motion description from the MPEG motion information using a robust parameter estimator. The error structure shown by the outliers of this first estimation allows us to cluster them into several local motions, using a non parametric estimator of the density (the mean shift) over a measure of the residuals of our estimation. Finally, results of our algorithm are shown in synthetic and real sequences.

- LL. Barceló, X. Binefa and J. R. Kender. Visual Events Detection in Well-Known Environtments. *Pp .217-224. In Recent Advances in Artificial Intelligence Research and Development. Frontiers in Artificial Intelligence and Applications, Ed. IOS Press. 2004*

3.3 Collaboration with other research groups

Thanks to this project, collaboration with important international research groups has been possible. Our group has established a good relationship with Professor John Kender, from Columbia University. Our collaboration has resulted in several published papers as well as research stages of Juan M. Sanchez and Lluís Barceló in New York. Professor Kender has come several times to Barcelona to give lectures and to take part in technical committees and thesis tribunals. Professor John Kender is a renowned researcher in the video analysis field.

We have also established collaboration links with Dr. Nicu Sebe, from the Amsterdam University. We have participated together in several Technical Committees of conferences as well as revisors for special issues of video analysis journals.

3.4 Technological Transfer

The PASSIVIM project has also brought us the oportunity to participate in other projects thanks to the prior work regarding robust statistics and object tracking performed. Specifically, we are also contributing in two more projects.

The first one is in collaboration with CIDA *(Centro de Investigación y Desarrollo de la Armada)* and it consist in automatic target detection and recognition in aerial 3D LADAR images. The algorithms of robust statistics developed in order to extract the motion parameters is our project fit perfectly in CIDA's project, due to the nature of source data and the model chosen to describe the scenes (a parametric model).

The second one aims to perform moving object tracking in infrared (IR) imagery. In this project we are working with a spanish company, *Tecnobit.* The tracking algorithms developed to follow the players in the PASSIVIM project were adapted to the IR imagery, yielding excelent results.

Both projects are in early stages but have provided some publications [11], [12], which are to be also credited to the PASSIVIM project.

References

[1] X.Orriols, Ll. Barceló, X. Binefa. Polynomial Fiber Description of Motion for Video Mosaicing. *Proceedings of the International Conference on Image Processing, ICIP2001*

[2] R.L. Felip, X. Binefa and J. Diaz. A New Parameter Estimator Based on The Helmholtz Principle. *IEEE Conference on Signal Processing (ICIP'05), Genova, September 2005 (In Press)*

[3] LL. Barceló, X. Binefa and J. R. Kender. Visual Events Detection in Well-Known Environtments. *Pp .217-224. In Recent Advances in Artificial Intelligence Research and Development. Frontiers in Artificial Intelligence and Applications, Ed. IOS Press. 2004*

[4] L. Barceló, R.L. Felip and X. Binefa. A new Approach for Real Time Motion Estimation Using Robust Statistics and MPEG Domain Applied to Mosaic Images Contruction. *IEEE conference on Multimedia and Expo., Amsterdam July, 2005 (In Press)*

[5] J.M. Sanchez, R. L. Felip and X. Binefa Preatentional Filtering in Compressed Video. *IEEE conference on Multimedia and Expo, Amsterdam July 2005 (In Press)*

[6] LL. Barceló, X. Binefa and J.R. Kender. Robust Methods and Representations for Soccer Player Tracking and Collision Resolution. *4th International Conference on Image and Video Retrieval (CIVR'05), Singapore, July 2005. (In Press in LNCS)*

[7] R. L. Felip, J. M. Sanchez and X. Binefa. Video summarization with moving objects in the compressed domain. *Proc. Of the IST/SPIE Conference Internet Imaging VI , Vol 5670, San Jose, pp. 143-154. January 2005*

[8] LL. Barcelo, X. Binefa. Contextual Soccer Detection Using Mosaicing Techniques. *4th Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA2005), pp. 77-84. Springer-Verlag LNCS 3522. June 2005, Estoril, Portugal*

[9] R. Felip. *Application of Robust Statistics and Mean Shift Analysis to Digital Video and LADAR Imaging*, Master Thesis, September, 2004.

[10] R. L. Felip, X. Binefa. Affine Description of Motions in Compressed Domain. *Pp: .93-100 In Recent Advances in Artificial Intelligence Research and Development. Frontiers in Artificial Intelligence and Applications, Ed. IOS Press. 2004.*

[11] R. Felip, S. Ferradans, J. Diaz-Caro and X. Binefa. Target detection in LADAR data using robust statistics. *In Proceedings of the SPIE Europe Symposium on Optics/Photonics in Security and Defence, Bruges, 2005*

[12] B. Martínez, L. Ferraz, J. Diaz-Caro and X. Binefa. Optimization of the SSD multiple kernel tracking applied to IR video sequences. *In Proceedings of the SPIE Europe Symposium on Optics/Photonics in Security and Defence, Bruges, 2005*

[13] D. Comaniciu, V.Ramesh and P. Meer, Kernel-based object tracking, *Transactions on Pattern Analysis and Machine Intelligence 25, May 2003.*

[14] H. Wang and D. Suter. Variable Bandwidth QMDPE and Its Application in Robust Optical Flow Estimation, *In Proceedings ICCV03, International Conference on Computer Vision, Nice, France*