FUSION FRAMEWORK FOR VIDEO EVENT RECOGNITION

Qiao Ma, Baptiste Fosty, Carlos Fernando Crispim-Junior, François Bremond

FUSION FRAMEWORK FOR VIDEO EVENT RECOGNITION

Qiao Ma[1], Baptiste Fosty[2], Carlos F. Crispim-Junior[3], François Brémond[4]
[1]Ecole Centrale de Pékin, Beihang University
37 Xueyuan Road, 100191 Beijing, China
[2,3,4]INRIA Sophia Antipolis – Mediterranee, STARS Team
2004 route des Lucioles, BP93, Sophia Antipolis, France
[1]Maqiao909@gmail.com, [2]Baptiste.Fosty@inria.fr, [3]Carlos-fernando.Crispim_junior@inria.fr, [4]Francois.Bremond@inria.fr

ABSTRACT
This paper presents a multisensor fusion framework for video activities recognition based on statistical reasoning and D-S evidence theory. Precisely, the framework consists in the combination of the events' uncertainty computation with the trained database and the fusion method based on the conflict management of evidences. Our framework aims to build Multisensor fusion architecture for event recognition by combining sensors, dealing with conflicting recognition, and improving their performance. According to a complex event's hierarchy, Primitive state is chosen as our target event in the framework. A RGB camera and a RGB-D camera are used to recognise a person's basic activities in the scene. The main convenience of the proposed framework is that it firstly allows adding easily more possible events into the system with a complete structure for handling uncertainty. And secondly, the inference of Dempster-Shafer theory resembles human perception and fits for uncertainty and conflict management with incomplete information. The cross-validation of real-world data (10 persons) is carried out using the proposed framework, and the evaluation shows promising results that the fusion approach has an average sensitivity of 93.31% and an average precision of 86.7%. These results are better than the ones when only one camera is used, encouraging further research focusing on the combination of more sensors with more events, as well as the optimization of the parameters in the framework for improvements.

KEY WORDS
object recognition and motion, architecture and implementation, Multisensor fusion, event recognition, Evidence theory

1. Introduction

The number of older people living alone around the world is increasing, highlighting the importance of solutions for the treatment of this population health care conditions and their life quality improvement.[1] Human activity recognition is an important part of computer vision, and therefore an active research topic in video surveillance areas, like human activity monitoring in outdoor places or indoor (like metro stations, bank agencies, daily living places) environments [2].

The mono-sensor approaches are used in detection, tracking, and recognition of activities in the scene without the stereo match nor synchronization processes of sensors[3][4]. Methods based on mono-sensor rely on image analysis have been carried out using different approaches like probabilistic approach [5-7] or constraint-based approach[8][9]. The constraint-based approach proposed by Romdhane *et al.* [10] handles the uncertainty of the activity recognition in complex event using probabilistic reasoning. It provides a convenient mechanism of reasoning to handle event uncertainty.

But, there are cases where the scene cannot be covered only by a single sensor or the distance from the sensor to the target can compromise the level of detail necessary for desired task accomplishment.

In these cases, the combination (fusion) of multiple video-cameras can increase the level of details of a scene. The combination of several sensors can help improve the low accuracy caused by the long distance as well as some other problems in visual recognition situations. Especially, multisensor fusion shows the values of different target features of an event (e.g. distance, velocity, weight), the combination of which can help use obtain a more complete view of the event and therefore a more promising performance of the recognition task compared to job done by single sensor.

The data fusion can take place at three levels [14]: data level (fusion of pixels), feature level (fusion of feature vectors of each sensor), and decision level (fusion of event from each sensor). The decision level fusion involves the multisensor fusion after each sensor has made a preliminary detection of an event occurring in the scene. Since the sensors for fusion can be either homogeneous or heterogeneous, the fusion at decision level makes it possible that changing or adding different sensors without completely changing event models (see Section 3.1).

This paper is organized as follows: The state of the art is presented in Section 2. In Section 3 we present the experimental material and the proposed data fusion framework, whose evaluation procedure and respective results are presented in Section 4. Section 5 describes our contribution and conclusion.

2. Related work

The recognition of different human postures (herein mapped as primitive states, see Section 3.1) is one of the basic traits to be detected in the event recognition systems. Fusion of video camera and environmental sensors [7][11] (like pressure, light, temperature, etc) are usually adopted in indoor places where specific sensors are already designed and installed. Data fusion between camera and inertial sensors has been also explored [12 - 14], as well as body sensor networks (BSNs)[15 - 17]. These approaches have been successfully used on the combination of homogeneous or heterogeneous sensors for event recognition. However, the use of wearable devices could be considered intrusive, and not as convenient and generic as non-contact video cameras [18].

Decision-level fusion methods include (but not limited to) weighted decision methods [19], Bayesian inference [20], and Dempster–Shafer's method [21] [22]. Dempster-Shafer evidence theory extends the Bayesian inference theory using incomplete information to make knowledge fusion, which resembles human reasoning process, and therefore is widely used for uncertainty modeling in many applications [23] [24]. However, the rule of combination of evidences as claimed by Dempster could give unexpected fusion results when conflicts among evidences exist (see Section 3.3.3.2). Consequently, a variety of improving methods have been proposed upon the Dempster-Shafer evidence theory. Yager [25] and Dubois *et al*. [26] proposed to assign the uncertainty of conflicting evidences to the frame of discernment, as a solution to the conflicting issue. Smets[27] proposes that the conflicting mass results from the non-exhaustively of the frame of discernment and Murphy[28] improved the basic probability assignment distribution instead of modifying the combination rule of the evidence theory. New combination rules are also discussed by several authors [29 - 31] in some specific applications.

Briefly, This paper proposes a decision-level multisensor fusion framework for event recognition by the combination of sensors on decision level, dealing with conflicting evidence by the adoption of Dempster-Shafer theory and statistical reasoning evidence method with the new combination rule proposed in [31]. Two spatially separated video cameras (a RGB camera and a RGB-D camera -Microsoft Kinect) are used, with each we have had managed to recognize the basic event (primitive state, see Section 3.1) occurring in the scene). Methods are used to deal with the basic event detection like "sitting" and "standing" human posture recognition when two sensors conflict. The framework proposed is tested with the real-world data by cross-validation and the evaluation shows promising results encouraging further research for improvements.

3. Material and Methods

3.1 Event modeling

This framework is based on a hierarchical approach for event modelling proposed by Vu *et al.* in [9] and extended by Zouba *et al*. in [11]. It categorizes events in respect to their complexity as follows:

The events are divided into four types: **Primitive state** (e.g. a person is standing) deals with the instantaneous values of a person. **Composite state** is a combination of primitive state. **Primitive event** corresponds to a change of primitive state (e.g. a person changes the posture), and **Composite event** is a combination of primitive states and/or primitive events. The model for event E includes all the physical objects in E, all the components involved in E, and a set of conditions (herein called "constraints") to be verified between physical objects and sub-events.

Briefly, a composite event stays at the top of the hierarchy in terms of complexity, and primitive state is the basic layer. A composite event recognition consists in the recognition of all the related primitive, composites states, and and/or primitive event described in the composite event model. Since the primitive state is the most basic layer of the hierarchy, we chose to fuse events at primitive state level to achieve an accurate recognition of higher level events.

3.2 Experimental set

An evaluation using videos of participants of is performed to verify the proposed framework. Experimental site is located in a test room of CHU Nice Hospital where ten videos of older person (5 females and 5 males) doing semi-directed activities like sitting and standing are taken.

In our case, two video cameras are used to record these experiments. The RGB camera is located from a wide view of the scenario but the long distance of detection leads to problems like noise, covering, etc. Besides, because of the Microsoft Kinect's 4-5 meters detection distance limitation, it has to be positioned nearer to the target person, and this leads to the problems of "missing" the objects and activities out of the view range. A data fusion is made between the RGB camera and the RGB-D camera which already have respectively the primitive state (sitting-standing posture in our case) recognition result. The original data reflect different sensors 'observations of the person's posture changes in a time interval (each is about 20 minutes). The scenarios performed in the experimental room are composed of a series of directed activities to access the person's physical profile (e.g. static and dynamic balance test, walking test)[14]. The goal of the proposed framework consists in managing conflicting evidences of spatially separated and time-synchronized sensors in the process of event recognition. It should be noticed, however, that the answer to the question that whether the fusion result of an event detection system can be improved or not compared

to what it would have been if only one sensor is used will depend on the reliability and the performance of each sensor as well as the data fusion method to fuse the data from each sensor.

3.3 Proposed approach

As in our case, we focus on the fusion framework of the system, two preliminary works are assumed to be done before entering the fusion module: firstly, each sensor detects, tracks and recognizes the person in the scene and its target event by its own with the pre-defined *a priori* knowledge like event model and scene information; Secondly, the 2D camera and the Microsoft Kinect are time-synchronized with the event timestamp in [14]. The step before data fusion of the framework gives us a set of XML files indicating the preliminary recognition results of each sensor. These data are provided for the next step: the fusion module.

The proposed fusion framework architecture is presented in Figure 1. In our framework, we consider the event level fusion between two visual sensors mentioned in the previous section, although the framework can take into account more sensors by the use of an iterative approach.

For each time instant *t*, the framework takes the test video data, the groundtruth of event annotated by experts as inputs, as well as the data obtained by training. After the detection, tracking and event detection of each sensor, the conflicting events are evaluated on the fusion module (FM). To be precise, on every instant of each sensor, the data input to the FM is composed of a pair of vectors that indicate the current detected primitive states and the character features of the target person (e.g. recognized event "Standing" and person's instantaneous 3D height).

The FM computes the likelihood of every possible event and manages the conflicts. It starts with the computation of the instantaneous likelihood, with which the local temporal is obtained for probability reasoning in the Evidence Theory. Based on the standard Gaussian distribution of the person feature value, the instantaneous likelihood and the local reliability are computed for each instant, the D-S data fusion center makes a decision on which sensor's primitive state detection should be taken into account. For event data fusion, two aspects are considered related to the event likelihood computation: person bounding box height's "distance" to Gaussian parameters of person's height obtained on training at a time instant, and also the local temporal relationship in

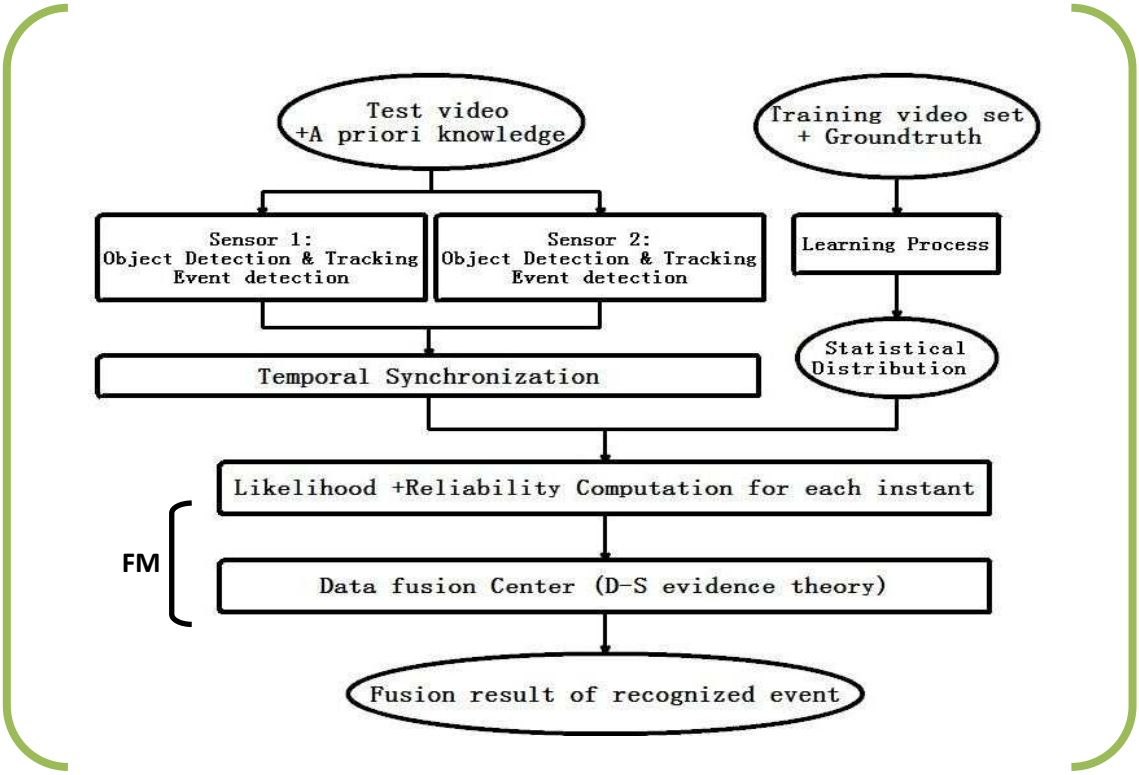

Figure 1. Proposed framework architecture

respect to previous instants. Both aspects have an influence on the belief level for the detected event.

3.3.1 Instantaneous likelihood function for primitive states

Based on the algorithms proposed in [10] dealing with the event recognition process, we propose an extension for human posture detection using a multi-sensor approach. The likelihood function consists in assessing the reliability value (belief level) of each sensor's event (i.e. primitive state detection) based on a standard Gaussian distribution obtained by a training step.

A simplified model of primitive state such as sitting-standing posture could be simplified by thresholding of a person's detected height. Herein, besides to the use of a fixed height threshold to distinguish "*sitting*" and "*standing*", the likelihood function of "*sitting*" and "*standing*" is computed based on the "distance" between the detected 3D height of a person and the trained Gaussian parameters results for "*sitting*" and "*standing*". This computed likelihood can also be interpreted as "belief level value".

A primitive state involves one or several features: for example, "Standing" can be defined by the person's detected height, and "in the zone of chair" is recognized when the constraint of person's position and pre-defined zone in the scene has been satisfied.

The distribution of a person's height for sitting and standing is considered as a Gaussian distribution. For each sensor, the Gaussian parameters (the mean μ, and the standard variance σ^2) for each posture state is learned by computing the average height of postures sitting or standing in the scene in respect to each sensor, with the training statistical data obtained a priori based on annotated data of person postures (Equation[1] 1 and 2).

$$\sigma_{e,i}^2 = E\left(Height_{i,j,e}^{av}{}^2\right) - E\left(Height_{i,j,e}^{av}\right)^2 = \sigma_{e,i}^2 + \frac{j-1}{j^2} Height_{i,j,e}^{av} - \mu_{e,i}^2 \quad (1)$$

$$\mu_{e,i} = E\left(Height_{i,j,e}^{av}\right) = \frac{j-1}{j}\mu_{e,i} + \frac{Height_{i,j,e}^{av}}{j} \quad (2)$$

For all feature values including the example feature in our case (sitting-standing posture's height), the instantaneous likelihood of the test video is computed for each frame using Equation 3[10] with the Gaussian parameters previously obtained.

$$PROB_{k,e,i}^{inst} = e^{-\frac{Height_{k,e,i}-\mu_{e,i}{}^2}{2\sigma_{e,i}^2}} \quad (3)$$

[1] Compute for each e: target event, i: sensor, j: video data number, k: instant, and "av" means "average". The notation is used throughout this paper.

Since the standard Gaussian distribution likelihood can be considered as a belief level value, this value is used as "*how strongly we believe that the event result of the sensor is true at the evaluated time instant*".

3.3.2 Local temporal reliability

Now another aspect of the likelihood computation is considered: whether sitting and standing is recognized or not should depend not only on likelihood based on person's detected feature value at this moment, but also the likelihood function values of previous instants in a pre-defined window size. The algorithm used can be the one computing the local temporal reliability values based on a fixed window size ω. (Herein a 5 seconds window is used.) As shown in Equation 4 and 5, the temporal reliability for current instant depends on the previous ω instants: the exponential of the time distance part is the cooling function of probability, which reinforces the near instants' effect and gives less importance to the far ones[10]. Generally, a primitive state is a continuous process which lasts seconds or minutes. The window size parameter depends on the domain application, and it should fit the minimum time interval of the person's primitive state.

$$PROB_{k,i,e}^{temp} = \frac{PROB_{k,i,e}^{inst} + M}{\sum_{t=k-w}^{t=k-1} e^{-(k-t)}} \quad (4)$$

$$M = \sum_{t=k-w}^{t=k-1} [e^{-(k-t)}(PROB_{k,i,e}^{temp} - PROB_{k,i,e}^{inst})] \quad (5)$$

The temporal reliability of each sensor for each target event is the input of the D-S evidence uncertainty reasoning. The reason why we choose this method in the fusion framework consists in its filter effect of instantaneous likelihood's uncertainty caused by segmentation and tracking errors during video processing. It should be noticed that ω is flexibly configurable parameter based on domain events.

3.3.3 Data fusion based on D-S Evidence Theory

The fusion architecture is illustrated in Figure 2. The event fusion engine serves as information process center for two sensors at each time, and the architecture can be used in an iterative way to fuse more than two sensors.

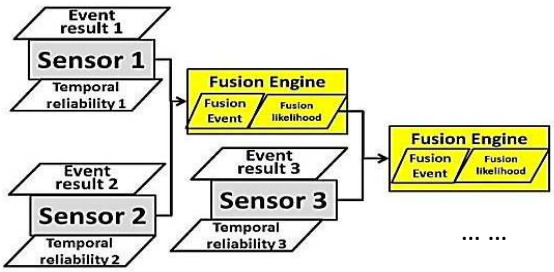

Figure 2. Architecture of event fusion process

The synchronization of each sensor uses their event timestamps and the "leave-accord-out" logic is adopted. Therefore, the engine only deals with conflicting events, as there is no need to make any fusion when the two sensors' events agree.

Proposed by Dempster [21] and improved by Shafer [32], the Evidence theory extends the Bayesian inference's application by allowing the uncertainty reasoning based on incomplete information. The support also comes from the possibility of distributing the imprecision to the combination of propositions such as "The person could be sitting and be in the zone chair". The evidence theory is used here to process uncertainty reasoning and to output the final event recognition result. In the next subsection, Dempster-Shafer evidence theory is briefly introduced, and then comes the fusion strategy in the proposed framework.

3.3.3.1 Dempster-Shafer sensor fusion algorithm

D-S theory is also called evidence theory which is statistical inference method used for modeling uncertainty. The evidence theory is based on a pre-defined set Θ, called the frame of discernment, which contains the group of all the possible mutually exclusive evidences or hypothesis of interest:

$$\Theta = \{A, B, \cdots\}$$

The function $m: 2^{\Theta} \rightarrow [0,1]$ related to a proposition satisfying:

$$m(\emptyset) = 0$$
$$\sum_{A \in \Theta} m(A) = 1$$

is defined as basic probability assignment (BPA). For any $A \in 2^{\Theta}$, m(A) is considered as the subjective confidence level on the event A. Accordingly, the whole body of evidence of one sensor is the set of all the BPAs greater than 0 under one frame of discernment.

The combination of multiple evidences defined on the same frame of discernment is the combination of the corresponding confidence level values based on BPAs (e.g., pre-defined by experts). Given two sensors (1 and 2), where each sensor has its body of evidence (m1 and m2), which are the corresponding BPA functions of the frame of discernment. We can combine two bodies of evidences into a new one by applying the following combination rule:

$$(m_1 \oplus m_2)(\mathrm{A}) = \frac{\sum_{M \cap N = A} m_1(M) m_2(N)}{1 - \sum_{M \cap N = \emptyset} m_1(M) m_2(N)} \quad (6)$$

where the sum in denominator is called the conflict factor of two bodies of evidence, and $\emptyset$ means the conflict between two propositions. This combination rule can be used iteratively because of its commutativity and associativity. Thus, the data fusion of more than two bodies of evidence is done by iterative pairwise process

Dempster-Shafter evidence theory was chosen as our data fusion basic idea, because compared to Bayesian theory, its reasoning process resembles the human perception. Most importantly, based on our likelihood computation steps in the previous subsections, we can use those statistical-based results to assign the unknown parameters in D-S evidence theory, and thus deal with the data fusion problem by a more reliable uncertainty reasoning method.

3.3.3.2 Modified combination rule dealing with conflicts

The classical D-S combination rule can be implemented to fuse data from two sensors, but it can lead to illogical results when the conflict factor approaches 1. For example, Doctor A and doctor B's judgement of a patient's disease: ***a*** or ***b*** or ***c***. Doctor A has 99% confidence on disease **a**, and 1% on ***b***; doctor B has 99% on ***c*** and 1% on ***b***. The classical combination rule gives the unexpected conclusion that the patient's disease is ***b***. In strong contradictory situations like this example, the reason for the unrealistic result of D-S evidence theory is that it distributes the uncertainty of global conflict to the common evidence of the two bodies. [25]

Varies alternatives or improvements of D-S combination rule have been put forward like: (Yager 1987), (Dubois *et al*. 1988), (Deng *et al*. 2004). These approaches consist in distributing the conflicting evidence probability to the whole set of propositions (Θ) in the frame of discernment. But the limitations are mainly falls in the following two aspects:

- When there are a great number of propositions in the frame of discernment, the weighting factor in the improved rules has to be computed for every subset or combination of evidences, and this increases the arithmetic operations.
- The associativity of the rule is not satisfied, which is vital for the local distributed algorithm structure in the reality applications.

In our framework, the data fusion is mainly carried out and verified with the following strategies:

For posture recognition conflict management, a new combination rule proposed and verified by Ali *et al*. [31] is used for primitive state recognition's data fusion. As written in the Equation 7 and 8, this newly proposed rule has been demonstrated to be efficient for combining the evidences from two or more sensors, and can be extended to the application of more primitive states and complex events.

$$(m_{RGB} \oplus m_{RGBD})(\mathrm{Sitting}) = \frac{1 - (1 - m_{RGB}(Sitting)) \times (1 - m_{RGBD}(Sitting))}{1 + (1 - m_{RGB}(Sitting)) \times (1 - m_{RGBD}(Sitting))} \quad (7)$$

$$(m_{RGB} \oplus m_{RGBD})(\mathrm{Standing}) = \frac{1 - (1 - m_{RGB}(\mathrm{Standing})) \times (1 - m_{RGBD}(\mathrm{Standing}))}{1 + (1 - m_{RGB}(\mathrm{Standing})) \times (1 - m_{RGBD}(\mathrm{Standing}))} \quad (8)$$

4 Results and Evaluation

The event uncertainty for "being sitting" and "being standing" is computed for each instant and each sensor based on the described training process: mean and variance of person height during evaluated postures. Results of the fusion approach are compared to the result of individual cameras (fixed-threshold for standing/sitting discrimination) to show the improvements brought on event detection performance.

A leave-one out cross-validation is adopted to evaluate the proposed approach. Briefly, the evaluation takes 9 videos out of the database for training while the tenth is used for test, and this is repeated until every video in the database is tested as validation data.

A priori knowledge like event model files are shared by both sensors with some new created events: sitting-fusion, standing-fusion, sitting-agreed, standing-agreed, which represents the different sensors situations (conflicting or consistent). For each instant of the validation video, the temporal reliability of each possible event result is computed based on a 5 seconds window size of historical reliabilities. Then, they are assigned to data fusion engine as BPAs to compute all the possible propositions' uncertainty after simple normalization for each sensor. At last, the output is the final decision of person's posture recognition result by comparing the fused uncertainty of every possible result.

The recognition performance of the proposed framework is measured in respect to the video annotation of experts, where TP (True positive) is assigned to the system evaluation when the system's recognition result is equal to the events annotated by experts; FP (False positive) is assigned when an event that doesn't occur in the annotated is detected; and FN (False negative) when the system misses an event that occurs in the annotated Groundtruth.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{10}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \tag{11}$$

Table 1. Comparison between mono-sensor using the threshold method and the fusion framework method: performance evaluation

Activity	Sitting		Standing	
Sensor	Precision	Sensitivity	Precision	Sensitivity
2D camera	84.29%	69.41%	79.82%	91.58%
Kinect	100.00%	36.47%	86.92%	97.89%
Fusion method	**82.35%**	**91.30%**	**91.04%**	**95.31%**

Table 1 presents the comparison of the proposed framework with the individual recognition performance of each video camera. The precision and the sensitivity indicate the best of the ten test sets during the cross-validation. Figure 3 (Detection view of RGB camera) shows a "standing" detection example of the RGB camera which is fixed at a corner of the experimental room, and Figure 4 shows the detection view of the RGB-D camera at the same time instant.

Figure 3. RGB cam. view

Figure 4. RGB-D cam. view

5 Conclusion and Future work

This paper proposes a framework for event fusion in activity recognition based on statistical learning and uncertainty reasoning using Dempster-Shafer theory. The decision-making process is implemented in an event recognition system to evaluate the applicability of the framework in real data. The proposed framework allows adding more sensors into the event recognition system by designing a proposition set and recomputing the weights. When it comes to a complex event, more primitive state computation can be implemented into the uncertainty reasoning process.

The experiments show that the fusion approach has a higher average sensitivity 93.31% than that of 2D camera and RGB-D camera. The averaged precision of 86.7% shows an improvement compared to the RGB camera, but a decline in respect of RGB-D camera. The reason may involve the fact that parameter factors like fixed window-size value used in the framework still need to be optimized. The framework is not able to recover when segmentation and tracking errors happen, where sometimes the person not detected or something else is mis-tracked as a person.

Future work includes evaluating different window sizes for the temporal reliability computation, adding optimised weighting factors in the combination rule for each camera in respect to their recognition performance, and also the combination of more sensors in the recognition of more postures like lying and bending, and eventually complex events.

References

[1] N. Zouba, F. Bremond, & M. Thonnat, An activity monitoring system for real elderly at home: Validation study, *7th IEEE International Conference on Advanced*

Video and Signal-Based Surveillance AVSS10, Boston, Etats-Unis, Aug, 2010.

[2] A. Avanzi, F. Brémond, C. Tornieri, & M. Thonnat, Design and assessment of an intelligent activity monitoring platform, EURASIP J. Appl. Signal Process, 2005

[3] Hongeng. S, Nevatia. R, Multi-agent event recognition, Computer Vision, *Proc. ICCV 2001Eighth IEEE International Conference vol.2,* 84-91 2001.

[4] D. J. Moore, I. A. Essa, & M. H. Hayes, Exploiting human actions and object context for recognition tasks , *Computer Vision, 1999. Proc. the Seventh IEEE International Conference, vol.1,* 80-86, 1999.

[5] M. Barnard, J. Odobez, & M. Bengio, Multimodal audio-visual event recognition for football analysis, *IEEE Workshop on Neural Networks for Signal Processing (NNSP),* 2003.

[6] T. Duong, H. Bui, D. Phung, & S. Venkatesh, Activity recognition and abnormality detection with the switching hidden semi-markov model, *IEEE computer society Conference on Computer Vision and Pattern Recognition CVPR'05,* 2005.

[7] O. Chomat, & J. Crowley, Probabilistic recognition of activity using local appearance, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'99),* 2:2104,1999.

[8] M. Ghallab, On chronicles: Representation, online recognition and learning, *5th International Conference on Principles of Knowledge Representation and Reasoning (KR'96), 5(8):*597-606, 1996.

[9] T. Vu, F. Bremond, & M. Thonnat, Automatic video interpretation : A novel algorithm for temporal scenario recognition, *The 18th International Joint Conference on Artificial Intelligence (IJCAI'* [9]*03),* 2003.

[10] R. Romdhane, F. Bremond, & M. Thonnat, A framework dealing with uncertainty for complex event recognition, *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on,* 29 2010-sept. 1 2010, pp. 392-399.

[11] N.Zouba, F.Bremond & M.Thonnat, Multisensor Fusion for Monitoring Elderly Activities at Home, *The 6th IEEE International Conference ON Advanced Video and Signal Based Surveillance (AVSS 09),* Genoa, Italy, September 2-4, 2009.

[12] A. Fleury, N. Noury, M. Vacher , Introducing knowledge in the process of supervised classification of activities of Daily Living in Health Smart Homes, *e-Health Networking Applications and Services (Healthcom), 2010 12th IEEE International Conference, 322-329, 1-3 July 2010*

[13] S. Zhang, S. McClean, B. Scotney, X. Hong, C. Nugent, & M. Mulvenna , Decision Support for Alzheimer's Patients in Smart Homes, *Computer-Based Medical Systems, 2008. CBMS '08. 21st IEEE International Symposium,* 236-241, 17-19 June 2008.

[14] C. F. Crispim-Junior, F. Bremond, & V. Joumier, A multi-sensor approach for activity recognition in older patients, *The Second International Conference on Ambient Computing, Applications, Services and Technologies - AMBIENT 2012, IARIA. Barcelona, Espagne:XPS/ThinkMindTM Digital Library,* Sep. 2012, in press.

[15] D.Malan, T.Fulford-Jones, M.Welsh, & S.Moulton, CodeBlue: An Ad Hoc Sensor Network Infrastructure for Emergency Medical Care, *MobySys 2004 Workshop on Applications of Mobile Embedded Systems (WAMES 2004),* June, 2004.

[16] C.Lombriser, D.Roggen, M.Stager, & G.Troster, Titan : A tiny Task Network for Dynamically Reconfigurable Heterogeneous Sensor Networks, *Verteilten Systemen (KiVS 2007), Bern, Switzerland,* Feb26-Mar2, 2007.

[17] W. Li, J. Bao, X. Fu, G. Fortino, & S. Galzarano, Human Postures Recognition Based on D-S Evidence Theory and Multi-sensor Data Fusion, *Proc. 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (ccgrid 2012) (CCGRID '12). IEEE Computer Society,* Washington, DC, USA, 912-917.

[18] T. B. Moeslund, E. Granum, A Survey of Computer Vision-Based Human Motion Capture, *Computer Vision and Image Understanding, Volume 81,* Issue 3, March 2001, 231-268.

[19] D. Hall, *Mathematical Techniques in Multisensor Data Fusion*, Boston, MA: Artech House, 1992.

[20] G. R. Iverson, *Bayesian Statistical Inference*. Beverly Hills,CA: Sage, 1984.

[21] A. P. Dempster, Generalization of Bayesian inference, *J.Royal Statist. Soc.*, vol. 30, 205–247, 1968.

[22] J. D. Lowrance, & T. D. Garvey, Evidential reasoning: A developing concept, *Proc. Int. Conf. on Cybern. and Soc.* Oct. 1982.

[23] O. Basir, & X. Yuan, Engine fault diagnosis based on multi-sensor information fusion using Dempster–Shafer evidence theory, *Information Fusion, Volume 8,* Issue 4, 379-386, October 2007.

[24] H. Wu, M. Siegel, R. Stiefelhagen, & J. Yang, Sensor fusion using Dempster-Shafer theory [for context-aware HCI], *Proc. Instrumentation and Measurement Technology Conference,* vol.1, 7- 12, 2002.

[25] R.R. Yager, On the Dempster–Shafer framework and new combination rules, Information Sciences 41 (1987) 93–138.

[26] D. Dubois, H. Prade, Representation and combination of uncertainty with belief functions and possibility measures, Computational Intelligence 4 (1998) 244–264.

[27] P. Smets, The combination of evidence in the transferable belief model, *IEEE Transactions on Pattern Analysis and Machine Intelligence 12* (5) 447–458,1990.

[28] C.K. Murphy, Combining belief functions when evidence conflicts, *Decision Support Systems 29* ,1–9, 2000.

[29] E. Lefevre, O. Colot, & P. Vannoorenberghe, Belief function combination and conflict management, *Information Fusion, Volume 3,* Issue 2, 149-162, June 2002,

[30] Y. Deng, W. Shi, Z. Zhu, Q. Liu, Combining belief functions based on distance of evidence, *Decision Support Systems, Volume 38*, Issue 3, 489-493, December 2004.

[31] T.Ali, P.Dutta, & H.Boruah, A new combination rule rule for conflict problem of Dempster-Shafer evidence theory, *International Journal of Energy, Information and Communications, Vol.3,* Issue.1, Feb 2012.

[32] G.Shafer, *A mathematical theory of evidence*: Princeton University Press Princeton, NJ, 1976