A temporal belief filter improving human action recognition in videos

Emmanuel Ramasso, Michèle Rombaut, Denis Pellerin

A TEMPORAL BELIEF FILTER IMPROVING HUMAN ACTION RECOGNITION IN VIDEOS

Emmanuel Ramasso, Michèle Rombaut, Denis Pellerin

Laboratory of Images and Signals, 46 av. Félix Viallet, 38031 Grenoble, France.
first_name.family_name@lis.inpg.fr

ABSTRACT

In the context of human action recognition in video sequences, a temporal belief filter based on the Transferable Belief Model is proposed. It ensures a consistency in the temporal belief evolution. The filter is useful to cope with varying video quality and experiment conditions by smoothing belief on actions and solving conflict due to contradictory parameters. The proposed approach is validated on real video sequences with moving camera under several view angles.

1. INTRODUCTION

Human motion analysis is an important topic of interest in Computer Vision and Video Processing [1]. Research in these domains is motivated by the diversity of applications such as video indexing and retrieval, automatic surveillance and human-computer interaction. One scientific challenge in human motion analysis is to recognize the human behavior from observations coming from multimedia features such as video, audio and text. The main problem is to link the real world, which has intrinsically an analogical nature, to the human world which is symbolic [2].

Many methods have been proposed for action recognition [1] notably based on *classification*, *template matching* and *neural network*. Generally, the methods are based on the *Bayesian framework* [3] with *Hidden Markov Models* (HMM) and *Dynamic Bayesian Networks* (DBN) [4]. Other methods are developed in Artificial Intelligence community notably *Petri Nets* [5, 6].

In [7], we have proposed an architecture for human action recognition using the *Transferable Belief Model* (TBM) [8] which is based on belief theory. The TBM is well-suited for action recognition notably because (i) *doubtful* transitions between actions are explicitly modelled, (ii) *conflict* between parameters reflects the need to improve the fusion process and (iii) *reliability* of parameters depends on the context and can be included in the system.

Action recognition is a preliminar step for activity recognition based on a sequence of actions. Thus, belief on actions have to be reliable despite varying video quality and experiment conditions. A temporal belief filter is proposed for this purpose and notably, it ensures a consistency in the belief on actions during time. Furthermore, it allows to solve conflict appearing when combining several parameters in the TBM framework. In order to illustrate the proposed approach, athletics meeting video sequences acquired with a moving camera and under an unknown view angle are analyzed to detect and recognize actions performed by one athlete.

The organization of the paper is as follows. An overview of our recognition architecture and the action recognition process based on the Transferable Belief Model is presented Section 2. A temporal belief filter is described Section 3. Section 4 deals with experimental results. Finally, we conclude and propose future work.

2. BASIC BELIEF ON ACTIONS OBTAINED BY FUSION

The video stream is analyzed by means of image processing and a semantic concerning the trueness of actions is computed. The system presented in figure 1 provides a weighted opinion, which is also called a belief, as well as uncertainty concerning reality of actions. A detailed description of this part is given in our previous work [7] and the main points are recalled here.

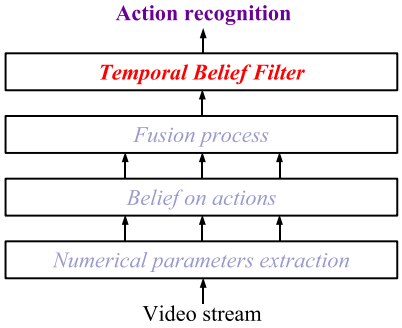

Fig. 1. Architecture for human action recognition in videos based on the Transferable Belief Model.

2.1. Numerical parameters

Relevant numerical parameters are extracted at each frame from the video stream. Parameters are generally application dependant and here the choice of the parameters is based on the following two main assumptions: first, the human is tracked by the cameraman because he is the center of interest and, second, the trajectories of human's head, center of gravity and one end of leg give information on actions. The chosen parameters are: the camera motion parameters estimated from two successive frames and which are the horizontal, vertical and divergence. Moreover, three human major points are tracked: the center of gravity, the head and the end of one leg. Points coordinates are analytically combined to obtain more advanced parameters which are the angle made by the humain main axis and the horizon, the variation of the center of gravity and the alternance of legs.

2.2. Transferable Belief Model fusion process

The numerical parameters values are converted into belief concerning the trueness of actions. A belief on actions is generated at each frame for each parameter. Belief of several parameters are then combined in the axiomatically well-founded Transferable Belief Model (TBM) framework proposed by Smets and Kennes [8] to obtain a belief which takes all parameters into account.

2.2.1. From numerical parameters to belief on actions

An action A is described by two hypotheses gathered in the frame of discernment (FoD) $\Omega_A = \{R_A, F_A\}$ with R_A (resp. F_A) stands for "action A is right" (resp. "A is false"). In the sequel, an hypothesis concerning an action is called a state, e.g. "the current state of A is R_A".

The goal of the fusion process is to obtain the belief of each action A according to several numerical parameters. A basic belief assignment (BBA) on an action A according to a parameter P is defined on the set of propositions $2^{\Omega_A} = \{\emptyset, R_A, F_A, R_A \cup F_A\}$ (where $R_A \cup F_A$ is the explicit doubt between hypotheses R_A and F_A) by $m_P^{\Omega_A} : 2^{\Omega_A} \rightarrow [0,1]$, $X \rightarrow m_P^{\Omega_A}(X)$ and by construction $m_P^{\Omega_A}(\emptyset) = 0$, and $\sum_{X \subseteq \Omega_A} m_P^{\Omega_A}(X) = 1$. A value $m_P^{\Omega_A}(X)$ is a basic belief mass which expresses a confidence in proposition $X \subseteq \Omega_A$ according to parameter P but does not imply any additionnal claims regarding subsets of X. It is a fundamental difference with probability theory. A fuzzy-set inspired method is used to convert each numerical parameter into sources of belief.

2.2.2. Fusion process

Rules of combination are then applied to obtain a belief which takes the belief of all parameters into account. The fusion process is performed frame by frame for each action independently. Given two distinct BBAs $m_{P_1}^{\Omega_A}$ and $m_{P_2}^{\Omega_A}$ defined on the same FoD Ω_A then their combination is defined as:

$$m_{P_1}^{\Omega_A} \textcircled{\triangle} m_{P_2}^{\Omega_A}(E) \quad = \quad \sum_{C \triangle D = E} m_{P_1}^{\Omega_A}(C).m_{P_2}^{\Omega_A}(D) \qquad (1)$$

with $\triangle = \cap$ (resp. $\cup$) for the conjunctive (resp. disjunctive) rule of combination. The rules of combination can be used in logical rules such as "*if* ... AND ... OR ... *then* ..." for describing actions by means of parameters states. These logical rules are then translated into belief combinations where the logical AND is replaced by the $\textcircled{\cap}$-rule and the logical OR by the $\textcircled{\cup}$-rule assuming the same FoD [9].

3. TEMPORAL BELIEF FILTER

A temporal belief filter is proposed to ensure (i) *a temporal consistency*: the belief on action can not vary abruptly between two successive frames of the video, (ii) *a consistency between parameters*: the conjunctive rule of combination used in the TBM fusion process may emphasize a conflict between parameters which has to be solved, (iii) *an exclusivity*: the temporal belief filter ensures that only one hypothesis concerning action (either R_A or F_A) is true at each frame.

The general principle (fig. 2) consists in assuming that the BBA at frame f is close to the BBA at frame $f-1$. Based on this assumption, a model of evolution predicts the current BBA taking the BBA at the previous frame into account. One model is defined for each of the two states of an action A (R_A and F_A) and a conflict-based criteria embedded in a CUSUM process is proposed for model change detection.

The temporal belief filter process works at each frame f and consists in four steps: (i) prediction, (ii) fusion, (iii) detection of conflict and (iv) model change if required.

3.1. Prediction of the current BBA

If an action state was R_A (resp. F_A) at frame $f-1$ then it would be partially R_A (resp. F_A) at frame f. This is an implication rule

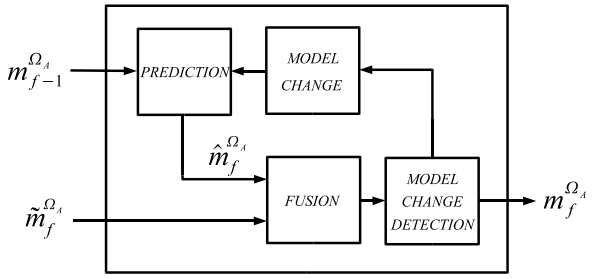

Fig. 2. The temporal belief filter principle.

$\mathcal{R}$ (resp. $\mathcal{F}$) which can be weighted by a confidence value of $\gamma_{\mathcal{R}} \in [0,1]$ (resp. $\gamma_{\mathcal{F}} \in [0,1]$) such as:

$$\begin{array}{ll} \text{Rule } \mathcal{R}: & \text{If } R_A \text{ at } f-1 \text{ then } R_A \text{ at } f \text{ with belief of } \gamma_R \\ \text{Rule } \mathcal{F}: & \text{If } F_A \text{ at } f-1 \text{ then } F_A \text{ at } f \text{ with belief of } \gamma_F \end{array} \qquad (2)$$

In the case concerned in this paper, the premise of the rule is not crisp but is a BBA. Implication rules are well managed in the TBM framework and details on their formalization as well as an application to target identification are described by Ristic and Smets in [9]. We have interpreted implication rules $\mathcal{R}$ and $\mathcal{F}$ (eq. 2) as models of evolution denoted $\mathcal{M} \in \{\mathcal{R}, \mathcal{F}\}$. Each one focuses on one hypothesis of the FoD of an action A which is either R_A or F_A.

In the sequel, the following vector notation of a BBA defined on a FoD Ω_A is used:

$$m^{\Omega_A} = [m^{\Omega_A}(\emptyset) \quad m^{\Omega_A}(R_A) \quad m^{\Omega_A}(F_A) \quad m^{\Omega_A}(R_A \cup F_A)]^T$$

Thus, a model of evolution can be interpreted as a BBA defined as:

$$m_{\mathcal{R}}^{\Omega_A} = \left[\begin{array}{cccc} 0 & \gamma_{\mathcal{R}} & 0 & 1-\gamma_{\mathcal{R}} \end{array} \right]^T \qquad (3)$$

$$m_{\mathcal{F}}^{\Omega_A} = \left[\begin{array}{cccc} 0 & 0 & \gamma_{\mathcal{F}} & 1-\gamma_{\mathcal{F}} \end{array} \right]^T \qquad (4)$$

The previous BBA $m_{f-1}^{\Omega_A}$ on action A required for the computation of the prediction $\hat{m}_{f,\mathcal{M}}^{\Omega_A}$ has the property to be defined with only two focal sets[1] depending on the current model $\mathcal{M}$: either R_A and $R_A \cup F_A$ if the model is $\mathcal{R}$, or F_A and $R_A \cup F_A$ if the model is $\mathcal{F}$. The disjunctive rule of combination (eq. 1) is used for computing the prediction from the previous BBA and the model of evolution:

$$\hat{m}_{f,\mathcal{M}}^{\Omega_A} = m_{\mathcal{M}}^{\Omega_A} \textcircled{\cup} m_{f-1}^{\Omega_A} \qquad (5)$$

The $\textcircled{\cup}$-rule is well-suited for model change detection under uncertainty because it allows to never assign more belief to an hypothesis than does the previous BBA as shown in the following expressions (eq. 6-7) of the prediction:

$$\hat{m}_{f,\mathcal{R}}^{\Omega_A} = \left[\begin{array}{c} 0 \\ \gamma_{\mathcal{R}} \times m_{f-1}^{\Omega_A}(R_A) \\ 0 \\ (1-\gamma_{\mathcal{R}}) \times m_{f-1}^{\Omega_A}(R_A) + m_{f-1}^{\Omega_A}(R_A \cup F_A) \end{array} \right] \qquad (6)$$

$$\hat{m}_{f,\mathcal{F}}^{\Omega_A} = \left[\begin{array}{c} 0 \\ 0 \\ \gamma_{\mathcal{F}} \times m_{f-1}^{\Omega_A}(F_A) \\ (1-\gamma_{\mathcal{F}}) \times m_{f-1}^{\Omega_A}(F_A) + m_{f-1}^{\Omega_A}(R_A \cup F_A) \end{array} \right] \qquad (7)$$

[1]The explanation will be given later but this remark is necessary to give the analytic expression of the prediction.

When $\gamma_{\mathcal{M}} = 1$, the prediction equals the previous BBA reflecting a total confidence in the current state of action A. When $\gamma_{\mathcal{M}} = 0$, the model expresses a total ignorance about the prediction of the current BBA on the action.

3.2. Fusion of prediction and measure

Prediction $\hat{m}^{\Omega_A}_{f,\mathcal{M}}$ and measure $\tilde{m}^{\Omega_A}_{f}$ represent two distinct pieces of information concerning the actual state of action A. The conjunctive combination (eq. 1) of their associated BBA leads to a new BBA whose conflict value ϵ_f (eq. 8) is relevant for model change requirement:

$$\epsilon_f = (\hat{m}^{\Omega_A}_{f,\mathcal{M}} \textcircled{\cap} \tilde{m}^{\Omega_A}_{f})(\emptyset) \tag{8}$$

The conflict analysis is thus required to know whether the current model is no longer valid. The CUSUM process of the conflict is well adapted for solving problems concerning *abrupt and short changes* or *gradual and long changes* in the conflict value because it allows to sum up conflict during time.

3.3. Detection of model change by a CUSUM process

The CUSUM is the cumulative sum during time of the error between a prediction and a measure. In the case concerned, the error is the conflict value. The initial CUSUM process works as follows [10]: when the CUSUM value becomes greater than a warning threshold $\mathcal{T}_w$ then the frame is stored as f_w and the model is *kept as valid*. As soon as the CUSUM value becomes greater than a stop threshold $\mathcal{T}_s$ (at frame f_s) then the model is *changed* and the new model is applied from f_s.

When a conflict appears between prediction and measure, as it could be the case in interval $[f_w, f_s]$, it was chosen to *trust the model of evolution*. Thus, the prediction is kept instead of an erroneous measurement and it avoids propagating conflict which is absorptive by the $\textcircled{\cap}$-rule:

$$m^{\Omega_A}_{f} = \begin{cases} \hat{m}^{\Omega_A}_{f,\mathcal{M}} \textcircled{\cap} \tilde{m}^{\Omega_A}_{f} & \text{if } \epsilon_f = 0 \\ \hat{m}^{\Omega_A}_{f,\mathcal{M}} & \text{otherwise} \end{cases} \tag{9}$$

This accounts for the fact that the BBA $m^{\Omega_A}_{f-1}$ can have only two focal sets (eq. (6)-(7)) depending on the current model $\mathcal{M}$. Therefore, the output of the belief filter is a BBA without conflict and with only one hypothesis whose belief is not null. The interest of the $\textcircled{\cup}$-rule is emphasized when there is often conflict because it allows to obtain $m^{\Omega_A}_{f\rightarrow\infty}(R_A \cup F_A) = 1$ which reflects total ignorance of the system.

To cope with low conflict during a long time, a *fadding memory* process has been embedded which allows to forget gradually past event. The fadding memory process requires a coefficient nicknamed *fadder*, and denoted as λ, which works on the current CUSUM $\mathbf{CS}(f)$ as follows:

$$\mathbf{CS}(f) \leftarrow \mathbf{CS}(f-1) \times \lambda + \epsilon_f \tag{10}$$

The fadder is here chosen as a constant and is applied at each frame.

3.4. Model change process

The two models ($\mathcal{R}$ and $\mathcal{F}$) are set once and one model is applied while it is valid otherwise, it is changed by the other. When $\mathcal{T}_s$ is reached, the interval of frames $\mathbf{I_T} = [f_w, min(f_s, f_w + \mathcal{W})]$ is interpreted as an interval of transition between two action states. The parameter $\mathcal{W}$ limits the size of the transition. The *vacuous* BBA is assigned to the frames belonging to $\mathbf{I_T}$ to well represent ignorance:

$$m^{\Omega_A}_{\mathbf{I_T}}(R_A \cup F_A) = 1 \tag{11}$$

After a model change, the new model is applied from the upper bound of the interval of transition $\mathbf{I_T}$ and the CUSUM is reset.

Remark concerning the initialization procedure: The temporal belief filter is an online process. To initialize the system, it is required to determine which is the better model fitting the first data. For that, the CUSUM process is applied on an interval of frames for all models and the chosen one minimizes the CUSUM.

4. EXPERIMENTS

Database description: The proposed system is used to distinguish between *running*, *jumping* and *falling* actions in two activities namely high jump and pole vault. The database is composed of 34 videos acquired with a moving camera. There are 22 pole vaults and 12 high jumps equivalent to 5318 frames (2573 of running action, 1552 of jumping and 1193 of falling). The database is characterized by its heterogeneity (fig. 3) with a panel of view angles as well as environments and athletes (out/indoor, male, female, other moving people).

Fig. 3. Heterogeneous database used for testing.

Settings: The temporal belief filter parameters are *set once for each action* and only one setting is provided *for each type of activity*. The parameters are set by expert knowledge.

Decision-taking based on belief: The recognition results have been kept as belief masses because the real purpose of the proposed action recognition is to go on with activities as a sequence of actions still based on the TBM. However, to assess the method, it is required to know whether an action is true. It was considered that if action A is credible at frame f, i.e. $m^{\Omega_A}_{f}(R_A) > 0$, then it is true.

Evaluation criteria: Recall and precision indexes, noted $\mathcal{R}$ and $\mathcal{P}$ respectively, are used for the evaluation and are computed as $\mathcal{R} = \frac{C \cap R}{C}$ and $\mathcal{P} = \frac{C \cap R}{R}$, where C is the reference set obtained by expert annotations, R is the set of retrieved frames provided by the recognition module by using the credibility-based criteria, and $C \cap R$ is the number of correctly retrieved frames.

Illustration and analysis: Table 1 gathers the recall and precision indexes for running, jumping and falling actions. The last line represents their mean over all videos. A wide gain is obtain on the mean recall with almost 19% for pole vaults and 8% for high jumps. A loss is observed on precision but quite low. The gain on recall is due to three effects of the temporal belief filter: First, uncertainty is reduced which is due to the specialization process involved by eq. (9); Second, the model change detection allows to reduce even annihilates abrupt changes from a non null belief to a null belief; Third, the filter solves conflict between two different actions states by astutely converting it into ignorance. The low loss on precision is mainly due to delays (fig. 4) added by the filter's settings.

Table 1. Recall ($\mathcal{R}$), precision ($\mathcal{P}$) in % before and after filtering for actions in pole vault and high jump. The last column is the gain/loss on recall/precision after filtering.

POLE VAULT	before		after		gain	
running	83.8	71.0	91.7	67.9	(+)7.90	(-)3.1
jumping	40.2	94.7	78.4	95.3	(+)38.2	(+)0.6
falling	51.5	87.0	67.5	83.8	(+)16.0	(-)3.2
mean	63.4	77.3	82.3	77.0	(+)18.9	(-)0.3
HIGH JUMP	**before**		**after**		**gain**	
running	89.5	86.1	99.8	84.1	(+)10.3	(-)2.0
jumping	80.1	79.7	84.2	79.0	(+)4.10	(-)0.7
falling	95.5	91.0	97.8	90.1	(+)2.30	(-)0.9
mean	88.2	85.6	95.8	84.0	(+)7.60	(-)1.6

The illustration depicted in figure 4 shows the efficiency of the approach to solve the problems defined in the beginning of Section 3. Belief on propositions $\emptyset$ (contradictory parameters concerning A), R_A (A is right), F_A (A is false) and $R_A \cup F_A$ (A is right or false) are represented. The illustration concerns a jumping action in a high jump sequence. The setting of the filter are $\lambda = 0.9$, $\gamma_{\mathcal{R}} = \gamma_{\mathcal{F}} = 0.95$, $\mathcal{T}_w = 2.5$, $\mathcal{T}_s = 3.8$ and $\mathcal{W} = 5$. High value of $\mathcal{T}_s$ involves a long time to be reached and is required when there are many abrupt changes on belief. To prevent from a too large back on data, $\mathcal{T}_w = 2.5$ is quite close to $\mathcal{T}_s$ thus, in this case, the sensitivity of the detection w.r.t $\mathcal{W}$ is low.

5. CONCLUSION

A temporal belief filter was proposed to smooth belief and solve conflict on human actions in real video sequences detected by means of a Transferable Belief Model fusion process. The disturbances on belief are mainly due to bad quality videos and varying experiment conditions and the proposed temporal belief filter gives one solution to solve those problems ensuring consistency. Notably, the transition between action states are emphasized. The method was applied on 34 real video sequences acquired with a moving camera where the purpose was to recognize *running*, *jumping* and *falling* actions in high jumps and pole vaults. The evaluation process based on action credibility showed the efficiency of the method.

Work is under progress for action sequencing still based on the Transferable Belief Model for activity recognition which is a higher level of interpretation of videos.

6. ACKNOWLEDGEMENT

This research is partially supported by SIMILAR European excellence network. The authors would like to thank the Computer Science Department of the University of Crete for the data exchange.

7. REFERENCES

[1] L. Wang, W. Hu, and T. Tan, "Recent developments in human motion analysis," *Pattern Recognition*, vol. 36, pp. 585–601, 2003.

[2] M. Lew, N. Sebe, and J. Eakins, "Challenges in image and video retrieval," *Lecture notes in Computer Science, Int. Conf. in Image and Video Retrieval*, vol. 2383, pp. 1–6, 2002.

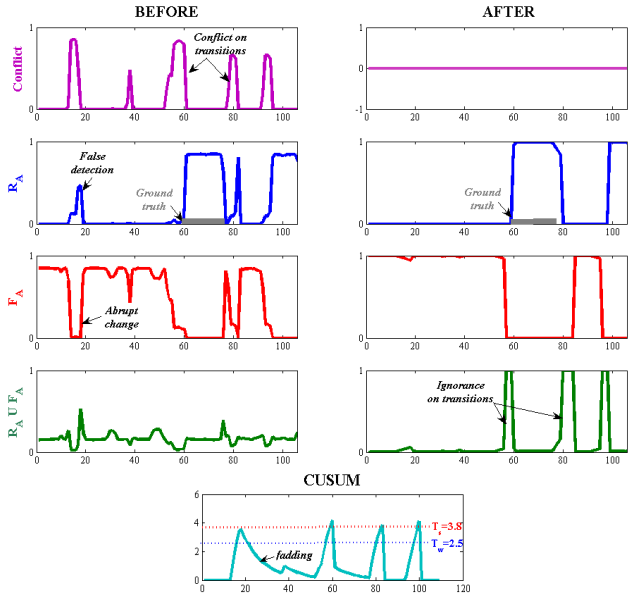

Fig. 4. Result of the temporal belief filter on the trueness on jumping action in a high jump sequence. Left (1-4): Before temporal belief filtering. Right (1-4): After temporal belief filtering. Bottom: CUSUM evolution. The disturbances due to conflict, false detections and abrupt changes, like in frames [12-19] and [75-83] are smoothed by the filter. Intervals of frames with conflict between two action states like in [52-61] and [77-81] are converted into transitions between these states. A specialization process is performed with the effect to reduce uncertainty. One can notice the effect of the memory fadding (exponential decreasing) on the CUSUM on frames [20-38].

[3] S. Hongeng, R. Nevatia, and F. Bremond, "Video-based event recognition and probabilistic recognition methods," *Computer Vision and Image Understanding*, vol. 96, pp. 129–162, 2004.

[4] T. Xiang and S. Gong, "Discovering bayesian causality among visual events in a complex outdoor scene," *IEEE Advanced Video and Signal based Surveillance*, pp. 177–182, 2003.

[5] M. Rombaut, I. Jarkass, and T. Denoeux, "State recognition in discrete dynamical systems using petri nets and evidence theory," in *Symbolic and quantitative approaches to reasoning and uncertainty*, June 1999, pp. 352–361.

[6] Z. Ding, H. Bunke, M. Schneider, and A. Kandel, "Fuzzy timed petri net : Definitions, properties, applications," *Mathematical and Computer Modelling*, vol. 41, pp. 345–360, 2005.

[7] E. Ramasso, D. Pellerin, C. Panagiotakis, M. Rombaut, G. Tziritas, and W. Lim, "Spatio-temporal information fusion for human action recognition in videos," in *13th European Signal Processing Conf.*, Antalya, Turkey, Sept. 2005.

[8] P. Smets and R. Kennes, "The Transferable Belief Model," *Artificial Intelligence*, vol. 66, pp. 191–234, 1994.

[9] B. Ristic and P. Smets, "Target identification using belief functions and implication rules," *IEEE Trans. Aerospace and Electronic Systems*, vol. 41, pp. 1097–1102, July 2005.

[10] S. Charbonnier, "On line extraction of temporal episodes from icu high-frequency data: A visual support for signal interpretation," *Computer Methods and Programs in Biomedicine*, vol. 78, pp. 115–132, May 2005.