An Improved Management Model for Tracking Missing Features in Computer Vision Long Image Sequences

RAQUEL R. PINHO[1], JOÃO MANUEL R. S. TAVARES[2], MIGUEL V. CORREIA[3]
[1] FEUP - Faculdade de Engenharia da Universidade do Porto, INEGI - Instituto de Engenharia Mecânica e Gestão Industrial, LOME - Lab. de Óptica e Mecânica Experimental
[2] FEUP (DEMEGI), INEGI, LOME
[3] FEUP (DEEC), INEB – Instituto de Engenharia Biomédica, LSI – Lab. de Sinal e Imagem
[1, 2, 3] Rua Dr. Roberto Frias, s/n, 4200-465 Porto
{rpinho, tavares, mcorreia}@fe.up.pt

Abstract: - In this paper we present a management model to deal with the problem of tracking missing features during long image sequences using Computational Vision. Some usual difficulties related with missing features are that they may be temporarily occluded or might even have disappeared definitively, and the computational cost involved should always be reduced to the strictly necessary. The proposed Net Present Value (NPV) model, based on the economic Theory of Capital, considers the tracking of each missing feature as an investment. Thus, using the NPV criterion, with adequate receipt and outlay functions, each occluded feature may be kept on tracking or it may be excluded of the tracking process depending on its historical behavior. This approach may be applied to any tracking system as long as the tracking results may be evaluated in each temporal step, and it can deal with the appearance, occlusion and disappearance of features especially useful for tracking in long image sequences. Experimental results, both on synthetic and real image sequences, which validate our model, will be also presented.

Key-Words: - Tracking, Occlusion, Management, Net Present Value (NPV), Motion Estimation

1 Introduction

In Computational Vision tracking features is a complex problem which has evolved considerably in the last years [1]. Usually many tracking applications require the tracking of large number of features simultaneously, and new features can rise at any time of the sequence, but they may also disappear during some instances of time or even definitively. Therefore, when features are not visible in some temporal step it should be decided if it is necessary to keep on tracking them, due to temporary (partial or total) occlusion, or if they have disappeared definitively from the image scene and so their tracking should cease. This decision is of great importance if a large number of features is being tracked, if the image sequence is long, if the images are noisy, etc.

In this paper we propose a new features management model which can deal with the appearance, occlusion and disappearance of features especially useful for tracking in long image sequences. This model can be applied to any tracking process as long as the quality of the previous tracks can be evaluated.

1.1 Related Work

Many approaches have been proposed to track features in image sequences, but a large number of them suppose that features are permanent and only situations of temporary occlusion are considered.

In [2] human motion is captured with infra-red computer aided gait analysis systems, and accurate estimation of makers positions plays an important role in estimating the different joint angles for gait analysis. During occlusion the 3D coordinates of each marker are predicted and interpolated in each frame using Radial Basis Function Neural Networks. Multiple humans occlusion is solved in [3] considering the Extended Kalman Filter (EKF) for trajectory estimation, but the system expects that merged blobs will eventually separate after awhile. The approach proposed in [4] also builds on the idea of object permanence, and long period tracking is performed at two levels: in the region level, a customized Genetic Algorithm (GA) is used to search for optimal region tracks; at the object level, each object is located based on adaptive appearance models, spatial distributions and inter-occlusion relationships.

Few tracking procedures consider the case of definitive disappearance of features of the image scene, although it is a common problem; for instance, in a surveillance system people may move on to another compartment and thereby disappear from the image field.

To our knowledge, existing approaches that deal with permanent occlusion, keep tracking features during some previously defined number of frames or simply discard them if they are not visible. In [5] when features are occluded, it is assumed that they follow the prediction of the Kalman filter, and if they are occluded for a determined number of frames they will be simply discarded. Similarly, in [6] a prediction model is used to track during temporary occlusion to increase the robustness of tracking people in video sequences, and if features disappear longer than a defined threshold then their tracking is ceased. A two layer solution is used in [7] to deal with total occlusions as well as group merging and splitting. In that work, the first layer produces a set of spatio-temporal strokes based on low level operations; and the second layer performs a consistent labeling of detected segments using a statistical model based on Bayesian Networks (BN), and thus if an object is not visible during some instances of time then that label ceases. In all these approaches, the number of instances during which the tracking of missing features is maintained is user defined, but no guarantee is given on the correctness or adequateness of the number selected. In [8] the Pfinder system tracks people without explicit reasoning the occlusion problem: blobs are simply deleted and added during and after occlusions, respectively. In this approach, data is discarded whenever a feature is not detected; as a result, in the case of noisy or cluttered images, this approach can lead to the loss of valuable information.

These usual approaches discard features independently of their previous behavior, and the total computational cost is not contemplated (as the number of features being tracked is not considered). To overcome such difficulties, we propose a new management model to deal with the problem of missing tracked features in long image sequences.

1.2 Our Approach

Our approach is based on a management model of the Theory of Capital, as many resemblances may be found between missing features in image sequences and investment projects in economics. Accordingly, to deal with the missing features in image sequences we propose the investment evaluation Net Present Value (NPV) method: by using adequate investment outlay and receipt functions, each feature without any new measurement data in an image frame (e.g. that is not visible) is considered as a project that requires an investment decision.

In our previous work [9, 10] we had already noticed the importance of an adequate features management model to keep the computational cost to the strictly necessary, by continuing to track features during some frames during which they could be temporarily occluded. But in our former approach, the only criterion used to keep on tracking the missing features was the number of frames during which each feature had been visible or not. Thus, in this previous management model, we associated a confidence value to each tracked feature in a frame. While tracking a feature if its predicted state had been updated with a new measurement the confidence value would increase, otherwise it would decrease. For a determined lower threshold confidence value, it would be as if the feature definitively disappeared and so its tracking would cease. Thereafter, if the same feature reappeared it would be considered as a new feature to track, and its tracking would be initialized. On the other hand, confidence values could increase to an upper threshold, which would be maintained if the features prediction continued to be updated with new measurements. So, if features disappeared during a short period of time the process continued without losing track of them, and if they had been tracked during a long period of time then they took longer to be discarded. The choice of the upper and lower thresholds would be made according to each application, and would be the same for all tracked features during the entire tracking process.

Instead, the new features management model proposed in this paper manages the decision to keep on tracking each feature, taking into account its previous behavior. The decision to keep on tracking a missing feature depends on the built investment outlay and receipt functions. So the Theory of Capital validates the choice of cost functions to deal with the missing features tracking problem. Thus, features which keep on appearing in the image scene will obviously continue being tracked; but if no measurement data has been introduced for a feature in the previous frames, then its tracking might cease. To decide whether to continue tracking a missing feature or not we take into account the number of tracked features and the quality of the previous matching results. These items are combined in the outlay and receipt functions, which are used to evaluate if a missing feature should be kept for tracking or not. This approach does not require any user input besides a measure of how vulnerable the management model should be to tracking features without any need of rejecting them incorrectly. Then the approach evaluates each missing features tracking performance individually, although it is flexible to contemplate the total computational performance as well.

The simplicity of our approach allows good tracking results with the involved computational cost reduced to the strictly necessary.

1.3 Paper Overview

This paper is organized as follows. In the next section we will give an overview of economic evaluation of investment proposals, and describe the NPV method with its application to tracking missing features in image sequences using Computational Vision. Then some experimental results are shown on synthetic and real image sequences. In the last section some conclusions will be held.

2 Investment Proposals and Missing Features Tracking

Several resemblances between investment proposals in economics and missing features tracking in image sequences may be found: projects are managed attending to their own specificities and to the global market state, while missing features can be managed attending to each features behavior and to the global tracking system performance; projects have to be approved for investment, while missing features should be considered for tracking; several projects can be undertaken at the same time just like missing features tracking.

When making capital and investment decisions, capital budgeting must be taken into account. That is why there must be done a selection of projects, the timing of each investment should be evaluated and it also should be determined the invested amount [11],[12].

To evaluate each possible investment some ideas should be kept [11],[12]

- the replacement of assets is equivalent to a new investment and so it has to be chosen on a solid basis;
- the historical cost is irrelevant as undertaken investments should more than pay for themselves;
- the previous assets should be taken into consideration;
- the production or acquisition of assets in the investment should be evaluated.

An investment decision requires the comparison of alternatives, and it is the fact that money can always earn a positive return that lends importance to the time dimension of the typical capital investment. The profitability of a proposed investment can be evaluated in several senses:

- the payout period – the number of years required to accumulate earnings sufficient to cover the investment costs;
- the internal rate of return (marginal efficiency of investment) – rate of interest/return which would render the net present value of all cash flow equal to zero;
- the NPV – the difference between the investments revenues and its costs.

The payout period approach does not take into account that the equipment may keep on working much longer than the payout period, and the time patterns of receipt are completely ignored. Also, the time value of money is not considered [11],[12]

The marginal efficiency also neglects the timing of receipts and is inappropriate for cash flows that change over life of projects [11],[12]

The NPV obtains optimal results when evaluating the cost of capital across years; it compares the difference between the investments revenues and its costs:

$$NPV = \sum_{t=1}^{n} \frac{S_t}{(1+r)^t} - I_0 , \qquad (1)$$

where S_t is the expected net cash receipt at the end of year t, I_0 is the initial investment outlay, r the discount rate and n the projects lifetime in years. If a projects NPV is positive, then it should be accepted, otherwise it should be rejected [11],[12]

2.1 The NPV-Investment Model for Features Tracking in Image Sequences

Applying the NPV method on the management of missing features, the tracking revenue and cost functions should be evaluated. To do so, two basic ideas ought to be considered:

1) the number of features – if the computational resources are overloaded with the tracking process of many features, then it would be a good idea to discard useless data as soon as possible; however, if a small number of features are been tracked, then it may not be necessary to discarded them so easily;

2) the quality of the previous estimates – if little certainty is given on a features state, then it may not be necessary to keep its tracking as we do not have reliable results; on the other hand, if the obtained estimates are very reliable, then it may be convenient to keep tracking.

Thus, using the NPV method, if I_{0i} corresponds to a medium cost of tracking a feature, then we would only choose to keep tracking it while missing if the actual revenue is higher than the medium cost.

To employ the NPV method, the user should specify the pretended discount rate. In the Theory of Capital the discount rate corresponds to the return rate that an investor is willing to accept in a risky investment, instead of a certain return in a safer investment in the financial market. Consequently, for tracking features it would correspond to a measure, defined by the user, of how vulnerable the management model should be to tracking features without any need of rejecting them incorrectly.

3 Experimental Results

In this section we exemplify the application of the NPV method to the Kalman filter results, when features are matched using the optimization of the Mahalanobis distances [13]. As a result, the quality of the previous tracks in the NPV method can be evaluated in this process by:

- the predicted state uncertainty given by the used stochastic filter – this measures the Kalman filter certainty on the predicted position for each missing feature;
- the Mahalanobis distance of the previous matches – if a feature has diverged largely from the previous estimates, then it may be keeping this divergence behaviour and keeping its tracking may be a safe way of preserving data.

This means that the initial investment outlay for each feature, i, is evaluated as:

$$I_{0i} = \frac{\sum_{t=0}^{m} k_t * c_{ti} * d_{ti}}{m}, \qquad (2)$$

where k_t is the number of tracked features in the instance of time t, c_{ti} is the sum of the main entries in the filters prediction reduced covariance matrix, d_{ti} is the Mahalanobis distance used to match that feature i, and m is the number of instances during which the feature has been successfully tracked. For the expected net cash receipt of that feature i in instance of time t we considered:

$$S_{ti} = k_t * c_{ti} * d'_{ti}, \qquad (3)$$

where d'_{it} is the Mahalanobis distance of the last instance in which that feature was successfully matched.

For the first example consider 10 synthetic sequences with 2 blobs: one blob does not move while the other has a horizontal movement of 5 pixels to the right in each frame. For each sequence the moving blob will disappear in a different frame. In the first sequence it disappears in the 3rd frame and onwards (figure 1), in the next sequence it disappears in the 4th frame, etc.

Figure 2 shows the NPV values associated to the tracking process of the missing blob whether it initially disappears in the 3rd, 4th, 5th,…, 12th frame. Therefore, with a discount rate of 20%, if the moving blob disappears in the 3rd or 4th frame, it only is tracked in the next frame and then its tracking ceases. However, if it disappears in the 5th frame, it will be tracked during the next 2 frames. Analyzing figure 2, it can be noticed that as the missing blob disappears later on in the image sequence, longer it will be tracked. This could be explained with the quality of the previous estimates; in this case, if a feature is tracked longer then the covariance values and Mahalanobis distances tend to decrease (because of the undergoing rigid movement), table 1.

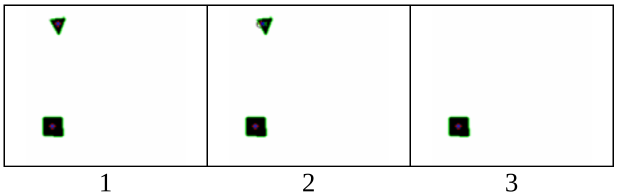

Figure 1: First frames of one of the sequences used in the first example: the moving blob disappears in the 3rd frame and onwards.

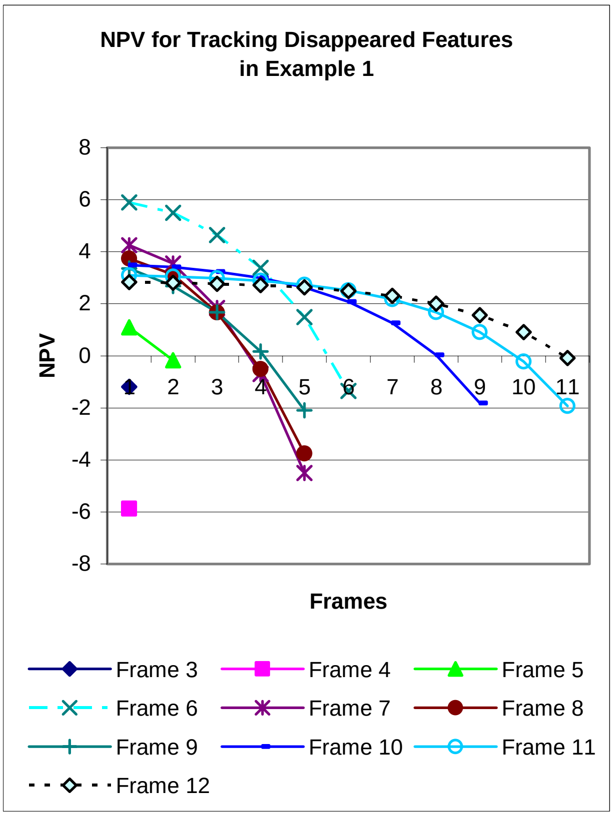

Figure 2: NPV values for tracking the missing blob in the first example (figure 1) if it disappears from the ith frame onwards.

For the next example consider a sequence of 10 frames with ten blobs (figure 3). As in the previous example the quadrangular and triangular blobs move 5 pixels to the right in each frame. The circular blobs move in a non-linear manner. In the 6th frame a quadrangle as well as a circle disappears. The quadrangular blob is tracked during the next four frames, but the circular blob is only maintained during three frames. The obtained results evidences the influence of the Kalman filter tracker and the type of undergoing movement, as the filter gets better estimates on rigid movement, and so the covariance values are smaller, as well as the Mahalanobis distances; on the other hand when non-linear movement is tracked, not so good estimates are built and the covariance values as well as the Mahalanobis distances are higher, which leads to smaller NPV values and tracking is not maintained for so long.

Frames	c_{ti}	d'_{ti}	I_{0i}
3	6.5	0.157204	3.295367
4	10.48118	0.348229	7.605962
5	17.11023	0.203616	8.222628
6	23.22009	0.022807	6.406252
7	20.98316	0.045731	5.369932
8	13.38963	0.039209	4.665639
9	14.59085	0.018318	4.07358
10	14.22746	0.004471	3.578984
11	13.06511	0.000161	3.181758
12	12.23145	0.000712	2.865281

Table 1: Values used for the NPV calculation in the 1st example when the feature disappears in frame *i*.

In this example, and comparing to the results in figure 2, it can also be noticed the influence of the number of tracked features in NPV calculation: the quadrangular blob has the same movement as those in the previous example, and it disappears in the 6th frame, but its tracking is not kept during 6 frames, as indicated in figure 3. So, the greater the number of features to track, easier it will be to discard missing features to attenuate the computational cost.

For the next example consider the real images obtained from a surveillance system of a shopping center, figure 3 (images from [14]). In the 2nd frame one of the previously tracked persons starts to enter a store and so the management model will cease tracking him in the 6th frame of figure 3. If he would have come out quickly the management model could maintain its previous tracking data, but in this case if he comes out the store our approach would initialize its tracking as a new feature. This example shows our

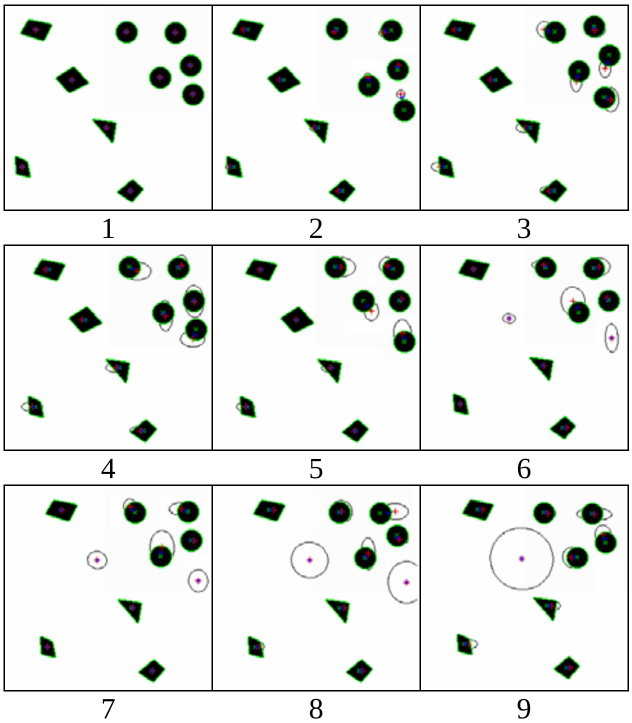

Figure 3: Tracking 10 blobs - A quadrangular blob with rigid movement and a circular blob with non-linear movement disappear in frame 6 and tracking continues until frame 9 or 8, according to their movement.

Figure 4: Tracking persons in a shopping center: our proposed management model can discard missing features for computational resources relieve.

approach is consistent in long image sequences such as those of surveillance systems, because it is essential to discard features whenever they might be unnecessary.

For the last experimental example consider a sequence of real images of an outdoor campus scene from the PETS (Performance Evaluation of Tracking and Surveillance) 2001 datasets. In the presented sequence a person is partially occluded by a car. As they approach each other, the used processing techniques only detect one region of motion and the only captured measurement is attributed to the car. However, the persons’ tracking is maintained during 3 frames if an internal rate of 0.02 is used (for higher internal rates it would be tracked longer, for instance a internal rate of 4.0 would keep tracking during 6 frames). When the person and the car split into different regions of motion the tracking of the person is initialized. This example shows the influence of the internal rate of return in tracking missing features with the NPV model.

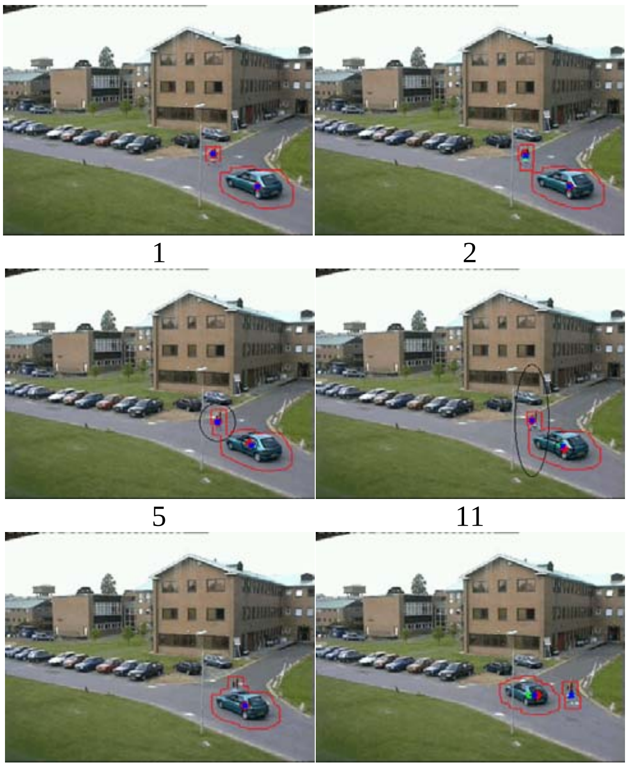

Figure 5. Tracking a person and a car during partial occlusion: until frame (2) the person and car are tracked separately; during frames (5)-(11) their motion regions overlap, and the only captured measurement is assigned to the car, so the persons tracking is done with higher uncertainty until it is ceased; after frame (58) features are tracked independently.

4 Conclusion

When tracking features in image sequences, as they could disappear from the image scene or not be visible during some instances of time, one has to decide to keep on tracking each missing feature or simply cease its tracking process. The criterion used in our new management model comes from the Theory of Capital, and considers each missing feature as an investment.

According to the NPV method, a feature should continue to be tracked if its revenues are higher than its costs, for a user determined discount rate which depends on the application. The receipt/outlay functions considered in our management model for tracking features are based on the results obtained from the used tracking approach, and may be applied to any tracking process as long as the quality of tracks may be evaluated.

The experimental results hereby presented use a probabilistic tracking approach, and so the exemplified NPV outlay and receipt functions consider the filters uncertainty as well as the cost of the correspondence between features. These results show that the number of frames during which a missing feature tracking is maintained, depends on the certainty of the used tracking process and on the complexity of the images.

Acknowledgments:

This work was partially done in the scope of the project “Segmentation, Tracking and Motion Analysis of Deformable (2D/3D) Objects using Physical Principles”, reference POSC/EEA-SRI/55386/2004, financially supported by FCT - Fundação para a Ciência e a Tecnologia in Portugal.

The first author would like to thank the support of the PhD grant SFRH / BD / 12834 / 2003 of the FCT.

References:

[1] A. Davison, J. Deutscher, I. Reid, "Markerless motion Capture of Complex Full-Body Movement for Character Animation," Proceedings Eurographics Workshop on Animation and Simulation, Manchester, UK, 2001.

[2] H. Lakany, G. Hayes, M. Hazlewood, S. Hillman, "Human Walking: Tracking and Analysis," IEEE Electronic & Communications, Colloquium on Motion Analysis and Tracking, 1999.

[3] R. Rosales, S. Sclaroff, "Improved Tracking of Multiple Humans with Trajectory Prediction and Occlusion Modelling," Proceedings IEEE Conference on Computer Vision and Pattern Recognition, Workshop on the Interpretation of Visual Motion, Santa Barbara, USA, 1998.

[4] Y. Huang, I. Essa, "Tracking Multiple Objects through Occlusions," IEEE Conference on Computer Vision and Pattern Recognition 2005 (CVPR 2005), San Diego, USA, 2005.
[5] T. S. Zhao, R. Nevatia, "Tracking Multiple Humans in Complex Situations," *IEEE Transactions on Image Processing*, vol. 26 (9), 2004, pp. 1208/1221.
[6] R. Plänkers, P. Fua, "Tracking and Modeling People in Video Sequences," *Computer Vision and Image Understanding*, vol. 81 (3), 2001, pp. 285/302.
[7] P. Jorge, A. Abrantes, J. Marques, "On-line Tracking Groups of Pedestrians with Bayesian Networks," Workshop PETS, Prague, Czech Republic, 2004.
[8] C. Wren, A. Azarbayejani, T. Darrell, A. Pentland, "Pfinder: Real-Time Tracking of the Human Body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19 (7), 1997, pp. 780/785.
[9] J. Tavares, A. Padilha, "Matching Lines in Image Sequences with Geometric Constraints," 7º Congresso Português de Reconhecimento de Padrões, Aveiro, Portugal, 1995.
[10] R. Pinho, J. Tavares, M. Correia, "A Movement Tracking Management Model with Kalman Filtering, Global Optimization Techniques and Mahalanobis Distance," ICCMSE 2005 - International Conference on Computational Methods in Science and Engineering, Loutraki, Greece, 2005.
[11] W. Baumol, *Economic Theory and Operations Research*: Prentice-Hall, 1985.
[12] H. Levy, *Capital Investment and Financial Decisions*: Prentice-Hall, 1994.
[13] R. Pinho, J. Tavares, M. Correia, "Human Movement Tracking and Analysis with Kalman Filtering and Global Optimization Techniques," II International Conference On Computational Bioengineering, Lisbon, Portugal, 2005.
[14] EC-Funded-CAVIAR-project, I. 2001-37540, "OneShopOneWait1cor.mpg, http://homepages.inf.ed.ac.uk/rbf/CAVIAR/," 2004.