LINEAR AND NON-LINEAR FILTERS FOR BLOCK BASED MOTION ESTIMATION

Virginie F. Ruiz
Department of Cybernetics, The University of Reading
Whiteknights, Reading RG6 6AY, UK
vfr@cuber.reading.ac.uk

ABSTRACT

Many techniques are currently used for motion estimation. In the block-based approaches the most common procedure applied is the block-matching based on various algorithms. To refine the motion estimates resulting from the full search or any coarse search algorithm, one can find few applications of Kalman filtering, mainly in the intraframe scheme. This paper presents an 8x8-block based motion estimation which uses the Kalman filtering technique to improve the motion estimates resulting from both the three step algorithm and the 16x16-block based Kalman application of [9]. In the interframe scheme, due to discontinuities in the dynamic behaviour of the motion vectors, we propose the filtering by approximated densities [10]. This application uses a simple form involving statistical characteristics of multi-modal distributions.

1. INTRODUCTION

In the field of motion estimation for video compression many techniques have been applied [1-5]. It is now quite common to see the Kalman filtering technique and some of its extensions used for the estimation of motion within image sequences. Particularly in the pixel-recursive approaches, which suit very much the Kalman formulation, one finds various ways of applying this estimation technique both in the time and frequency domains. On a very general perspective, we find use of Kalman filter (KF), the extended Kalman filter (EKF) and the parallel extended Kalman filter (PEKF) [6-8].

In the block-based motion-compensated prediction approaches, the most common procedure is the block-matching technique. There are several well known algorithms that perform the block matching motion estimation, among them being the full search algorithm (FSA) [3-5] that determines the motion vector of a macroblock by computing the MAE at each location in the search area. This is the simplest method, it provides the best performance, but at a very expensive computational cost.

To reduce this computational requirements, several heuristic search strategies have been developed, as for example the two-dimensional logarithmic search, the parallel one-dimensional search, etc [3-5]. These are often referred to as fast search algorithms.

Lately, some new fast strategies as well as motion estimation have been proposed. But, very few applications are available of Kalman filtering for the estimation of motion vectors resulting from a 16x16-block based approach (16x16-KF), [9].

Section 2, we propose an 8x8-block based motion estimation using Kalman filtering (8x8-KF) to improve the motion vector estimates resulting from both the conventional three step algorithm (TSA) [3-5] and the 16x16-KF proposed in [9]. Section 2.1 introduces the state-space representation for the motion vector, the Kalman equations based on this later are given in section 2.2. The comparative results obtained for different classes of video sequences are presented in section 2.3.

Section 3 considers the interframe situation. The problem with the use of Kalman filtering is that its conventional modelling is not appropriate when discontinuities in the dynamic behaviour appear. Therefore, the filtering by approximated densities FAD is proposed in order to improve the motion vector estimates resulting from both the conventional FSA and TSA algorithms. The FAD [10] is a non-linear, adaptive filtering technique. It uses a maximum entropy principle under linear constraints. The method is essentially based on the development of a logarithm for the computation of *a priori* and *a posteriori* probability density functions as linear combinations of several functions chosen according to some specific criterion. Section 3.1, we elaborate functions of an exponential type for the definition of probability density that characterise the block-motion vector. The non-linear filter is then implemented in section 3.2. In section 3.3, the results are given and the superior performance of the filter on class A, B and other well known video sequences is demonstrated.

2. INTRAFRAME ESTIMATION

2.1 State representation

The scanning in a frame is from the top left to the bottom right. The motion vector of a macroblock can be predicted from that of its left spatial neighbour. The measured motion vectors are obtained through a conventional three step procedure. In the same manner as in [9], the intraframe motion estimation process is modelled through an auto-regressive model which produces the state-space equations.

We define the 8x8-block based representation as follows: each 16x16-block yields a zig-zag sequence of four 8x8-blocks. This corresponds to a conventional pixel decimation for block matching in an 8x8 bock. The motion vector of these sub-blocks is defined as $V^t = (x,y)$, where x and y denote the horizontal and vertical components, respectively. These two components are assumed independent.

The motion vector of an 8x8-block can be predicted from the one of the previous 8x8-block according to the time index k of the zig-zag order using the state equation:

$$V(k+1)=F(k)V(k)+W(k) \quad (1)$$

where $W(k)$ is the state noise vector, with covariance matrix Q. The state noise components w_x, w_y are assumed independent and Gaussian distributed with zero-mean and same variance q. The matrix $F(k)$ is the transition matrix which describes the dynamic behaviour of the motion vector from one 8x8-block to the next. As the motion vector components are independent the matrix is diagonal.

The Kalman filter updates the motion estimate $V(k/k\text{-}1)$ given the previous measurement at time index k-1, according to the new measured motion vector $Z(k)=z(k)$ and the measurement equation:

$$Z(k)=V(k)+N(k) \quad (2)$$

where $N(k)$ is the measurement noise vector with covariance matrix R. The noise components n_x, n_y are assumed independent and Gaussian distributed with zero-mean and same variance r.

It is observed that the measured motion vectors are actually obtained from the TSA run on a 16x16-block basis that yields the zig-zag sequence of four measured motion vectors $Z(k)$ with same values on the 8x8-block basis. By this mean the Kalman filter has four measurements instead of one to adjust the motion estimate $V(k/k)$ when the assumption of smooth changes is not strictly valid. As a result, it is expected that we have a better motion estimate for the 8x8-motion compensation procedure.

2.2 Linear filtering

From the above state-space model (eq. 1, 2) the consecutive motion vectors $V(k/k\text{-}1)$ and $V(k/k)$, with error covariance matrix P, are recursively estimated given the past measurement $Z(k\text{-}1)$ and present measurement $Z(k)$ through the Kalman filter.

Based upon this very basic state-space representation for the motion, the conventional Kalman equations are implemented as follows.

- *Update with the new measurement $Z(k)$:*

$$K(k) = P(k/k\text{-}1).\{P(k/k\text{-}1)+R(k)\}^{-1} \quad (3)$$

$$V(k/k)=V(k/k\text{-}1)+K(k)\{Z(k)-V(k/k\text{-}1)\} \quad (4)$$

$$P(k/k)=\{I-K(k)\}.P(k/k\text{-}1) \quad (5)$$

- *Prediction:*

$$V(k+1/k)=F(k)V(k/k) \quad (6)$$

$$P(k+1/k) = F(k)P(k/k)F^t(k)+Q(k) \quad (7)$$

2.3 Results

For comparison purposes the FSA, the TSA, the 16x16-KF presented in [9] and the proposed 8x8-KF are run for different classes of video sequences. We use 50 frames of the following sequences: 'Alkistis' class A, 'Carphone', 'Foreman' and a sub-sampled sequence of 'News'.

It is observed that all the above listed techniques can be qualified as sub-optimal techniques. Hence, in any case the results expected are very close in terms of average PSNR:

Table 1. Comparative results on the average PSNR

	Average PSNR dB			
Sequences	FSA	TSA	16x16-KL	8x8-KL
Alkistis	30.769	30.664	30.670	30.821
Carphone	33.793	33.578	33.585	33.672
Foreman	32.649	32.235	32.240	32.342
News	27.686	27.246	27.254	27.414

As expected the 8x8-block based proposed approach results in an even greater PSNR better than both the TSA and 16x16-KF.

For the particular state-space representation used, when the real motion corroborate the assumptions made regarding, only translational motion with very smooth changes, the 8x8-KF is even better than the one resulting from the FSA, as observed when monitoring the average PSNR values.

3. INTERFRAME ESTIMATION

3.1 Block motion characterisation

Before implementing FAD, we must define an *a priori* statistical distribution characterising the motion vector. A statistical study has been conducted using the full search algorithm for each 16x16-block of a frame over 100 frames of 'Missa' and 'Alkistis', 130 frames of 'Mother and Daughter', and 50 frames of 'Carphone' and 'Foreman'. We keep the general assumption of displacement x and y independence despite that it is rarely true. Hence, in the following we will consider only the developments for the x-component of the motion vector. In a frame at time instant k, a block $B(l)$, (l=0,…,98) is characterised by its motion vector. This motion vector is predicted from the previous frame using the block-matching technique. Thus, the displacement x takes numerical values within the range -7, ..., +7, at integer pixel accuracy.

On a frame to frame basis, the x-component (resp. y) of block $B(l)$ is rarely long-lasting stationary on one or more of the above possible values. What more likely describes the dynamic behaviour of x (resp. y) is jumps from one value to another with very short-lasting stationarity on one or more particular values.

From now on, we consider the multi-model situations. A pair (M, x) defines the state where the available models M, belong to a finite set $\{md_1,\ldots,md_N\}$. A Gaussian probability distribution for $x(k)$ (resp. $y(k)$) yields a probability on each model M. Hence at frame k+1, the state $x(k+1)$ (resp. $y(k+1)$) has several probability distributions that are conditional upon each model. We define the state space model,

$$a(k),b(k) \quad Markov\ \ chain$$
$$x(k+1) = a(k)x(k) + b(k) + w(k) \qquad (8)$$
$$z(k) = x(k) + v(k)$$

where state noise $w(k)$ and measurement noise $v(k)$ are Gaussian distributed with zero-mean and variance Q and R, respectively. The coefficients (a,b), are taken to be a N-state Markov chain, $(a(k),b(k))$. Due to the product $a(k)x(k)$, the x-state distribution loses its unimodal Gaussian feature. For our application we consider tri-modal and five-modal approximations:

- *Tri-modal Gaussian approximation*

In this case we consider that $x(k)$ (resp. $y(k)$) takes values with respect to the available models: $md_1=\{x(k)<0\}$, $md_2=\{x(k)=0\}$, $md_3=\{x(k)>0\}$ and the corresponding Markov coefficients (a_1,b_1), $(a_2,\ b_2)$, $(a_3,\ b_3)$. These coefficients and the state equation (8) without noise define the three possible (*limit*) Gaussian distributions for $x(k)$ (resp. $y(k)$) with mean and variance: $m_i=b_i/(1-a_i)$ and $\sigma_i^2=Q/(1-a_i^2)$ $i=1,2,3$.

- *Five-modal Gaussian approximation*

In the same manner the five possible models are: $md_1=\{x(k)<1\}$, $md_2=\{x(k)=-1\}$, $md_3=\{x(k)=0\}$, $md_4=\{x(k)=+1\}$, $md_5=\{x(k)>+1\}$ with the corresponding Markov coefficients $(a_i,\ b_i)$, $i=1,2,...,5$ which define with the state equation (8) the five possible Gaussian distributions for $x(k)$ (resp. $y(k)$) with mean and variance m_i and σ_i^2 , $(i=1,2,...,5)$.

According to our concept, we now define the density functions relative to the non-linear state-space model (8) where the coefficients $(a(k),b(k))$ are subject to random variations in time. The distribution taken for $(a(k),b(k))$ is a measure such that

$$\sum_{i=1}^{N} P[(a(k),\ b(k)) = (a_i, b_i)]\delta_i(a,b) \qquad \text{with } N=3 \text{ or } 5.$$

Thus, $(a(k),b(k))$ randomly assumes the values, (a_i,b_i), with respect to the transition probability

$$\pi_{il} = P[(a(k+1),b(k+1)) = (a_l,b_l) / (a(k),b(k)) = (a_i,b_i)] \text{ with } i,l=1,...,N. \qquad (9)$$

These transition probabilities have been obtained through the statistical study mentioned above.

We have to express the *a priori* densities for x (resp. y). A Gaussian distribution is of exponential type. The logarithm of its density function is linearly developed from the basis functions: $\varphi_0(x)=1$, $\varphi_1(x)=x$, $\varphi_2(x)=x^2$, (resp. y).

We define the linear constraints l_j, $(j=0,1,2)$ as the expected values of the functions φ_j, $(j=0,1,2)$. Hence, these constraints correspond to moments the Gaussian distribution. Thus, the density for each possible model $(i=1,...,N)$ is defined by

$$^i f(x) = exp(^i\lambda_0 + {}^i\lambda_1.x + {}^i\lambda_2.x^2), \qquad (10)$$

where the Lagrange multipliers $^i\lambda_j$, $(i=1,...,N;\ j=0,1,2)$ are bijectively obtained from the linear constraints, i.e. the moments of the Gaussian distribution.

Finally, the density function for the state $(a(k),b(k),x(k))$ can be written as follows

$$\sum_{i=1}^{N} P\left[(a(k),b(k)) = (a_i,b_i)\right]\delta_i(a,b).{}^i f(x) = \sum_{i=1}^{N} p^i \delta_i.{}^i f(x) \qquad (11)$$
with $\int {}^i f(x)dx = 1.$

3.2 Non-linear filtering

In a frame k, for each block $B(l)$, $(l=0,\ldots,98)$ the filtering by approximated densities updates and predicts the conditional density functions given the new measured motion $z(k)=\eta$, resulting from the TSA.

The update uses the measurement equation (2) and the Bayesian formula. That yields the Lagrange multipliers $^i\lambda_j$, $(j=0,1,2)$ of the density $^i f(x)$ conditional upon the measured displacement

$$^i\lambda_0(k/k) = {}^i\lambda_0(k/k\text{-}1) + \mu_0 + \mu_1\eta + \mu_2\eta^2$$
$$^i\lambda_1(k/k) = {}^i\lambda_1(k/k\text{-}1) - \mu_1 - 2\mu_2\eta$$
$$^i\lambda_2(k/k) = {}^i\lambda_2(k/k\text{-}1) + \mu_2$$

where μ_i, $(i=0,1,2)$ are Lagrange multipliers of the measurement noise density. The normalisation introduces the external coefficients $^i\alpha_j = \int {}^i f(x)dx$, such that the updated probability of the Markov coefficients (i.e. models) is defined by

$$p^i(k/k) = \frac{p^i(k/k-1).{}^i\alpha_j}{\sum_{l=1}^{N} p^l(k/k-1).{}^l\alpha_j} \qquad i=1,\ldots,N.$$

The displacement estimate $x(k/k)$ used in the motion compensation procedure, is hence evaluated through the following expression in terms of the *a posteriori* probabilities and means:

$$x(k/k) = \sum_{i=1}^{N} p^i.m_i(k/k). \qquad (12)$$

Its numerical value is taken at half-pixel accuracy.

The prediction of the distribution for the block under consideration in frame k+1 conditional upon the observation in frame k, is now left. The probabilities of the coefficients $(a(k+1),b(k+1))$ are predicted from the above ones and the transition probabilities (9):

$$p^l(k+1/k) = \pi_{il}.p^i(k/k) \quad i,l=1,\ldots,N.$$

The prediction of the densities $^i f(x)$, $(i=1,\ldots N)$ uses the state equation (3) to compute the linear constraints $^i l_j(k+1/k)$,

$$^i l_0 = 1$$
$$^i l_1 = {}^i l_0(\ a_i m_i(k/k) + b_i)$$

$$^{i}l_2 = {}^{i}l_0(Q+a_i^2(\sigma_i^2(k/k)+m_i^2(k/k))+b_i^2+2a_ib_im_i(k/k))$$

where Q is state noise variance. The values of constant coefficients, a_i and b_i, depend on the model $\{md_1,\ldots,md_N\}$.

The predicted distribution which maximises the entropy under the linear constraints is the distribution for which the Lagrange multipliers of the density solve the non linear system: $^{i}l_j(k+1/k) = E[^{i}\varphi_j(x(k+1/k))]$. Thus, the Lagrange multipliers of the density are defined in terms of mean and variance resulting from the constraints as follows:

$$^{i}\lambda_2(k+1/k) = -1/2\sigma_i^2(k+1/k)$$
$$^{i}\lambda_1(k+1/k) = m_i(k+1/k)/\sigma_i^2(k+1/k)$$
$$^{i}\lambda_0(k+1/k) = -0.5\ln(2\pi\sigma_i^2(k+1/k)) + \ln(^{i}l_0) - m_i(k+1/k)/2\sigma_i^2(k+1/k)$$

Finally, the predicted distribution conditional upon the observed motion through TSA is characterised by its exponential density function with Lagrange multipliers predicted in accordance with the FAD-concept.

3.3 Results

The filter has been implemented for 100 frames of 'Missa' and 'Alkistis', 130 frames of 'Mother and Daughter', and finally 50 frames of 'Carphone' and 'Foreman' sequences. In all cases, the 3-modal distribution is characterised by the means m_1=-3.5, m_2=0, m_3=3.5 and standard deviations $\sigma_1=\sigma_3$=1, $3\sigma_2$=0.5; for the 5-modal case we have m_1=-4, m_2=-1, m_3=0, m_4=1, m_5=4 and $3\sigma_1=3\sigma_5$=2.5, $3\sigma_2=3\sigma_3=3\sigma_4$=0.5. These values are valid for both components x and y of the motion vector of each block within a frame. The filter is initialised once and for all as the first frame is loaded. It is observed that we do not need any refreshment or motion detection as it is often the case when using the Kalman filter.

The results presented in the Table 2 in terms of average PSNR demonstrate the performance of the technique applied to interframe motion estimation.

Table 2. Comparative results on the average PSNR

		AVERAGE PSNR dB							
		FSA	TSA	FILTERING BY APPROXIMATED DENSITIES					
				3-modal	5-modal	3-modal	5-modal	3-modal	5-modal
				R=0.06		R=0.25		R=1	
Missa	Q=0.01	41.069	40.971	*41.373*	***41.380***	*41.177*	*41.273*	40.977	40.979
	Q=0.027			***41.250***	40.983	*41.131*	*41.317*	40.976	40.981
Mother & Daughter	Q=0.01	36.139	36.046	36.132	***36.143***	36.097	36.100	36.066	36.063
	Q=0.027			36.104	36.063	36.072	36.091	36.055	36.054
Alkistis	Q=0.01	30.393	30.154	30.228	30.214	30.277	***30.432***	30.246	30.280
	Q=0.027			30.357	30.191	30.308	***30.432***	30.264	30.293
Carphone	Q=0.01	33.793	33.539	33.781	*33.884*	33.658	***33.904***	33.622	33.691
	Q=0.027			*33.939*	33.571	33.656	***33.957***	33.580	33.658
Foreman	Q=0.01	32.649	32.222	32.592	32.588	32.570	***32.656***	32.559	32.537
	Q=0.027			32.431	32.244	32.413	32.511	32.401	32.424

For different state and measurement noises the average PSNR is always greater than the one resulting from the TSA. For each sequence, we find one or more situations where the average PSNR resulting from FAD is even greater than the one resulting from the FSA.

4. CONCLUSION

For the intraframe motion estimation these are encouraging results, in the sense that with the appropriate state model and *a priori* assumptions appropriate to a closer real motion vector behaviour, we would be able through Kalman filtering to have a greater PSNR than the full search for any frame of the sequence.

Regarding the applicability of the concept of filtering by approximated densities in the area of motion estimation the results are also very encouraging. It would be of interest to see if a 7-modal approximation would give better results, and of course what the results of this approach are in the case of intraframe motion estimation.

Regarding the cost for using such an approach, it is observed that in all cases, three step plus FAD is still faster than the full search or three step plus Kalman.

5. REFERENCES

[1] Musmann H. G., Pirsch P. and Grallert H.-J., 'Advances in picture coding', *Proceedings IEEE*, Vol. 73, No. 4, pp. 523-548, 1985.

[2] Aksu I., Ildiz F. and Burl J. B., 'A comparison of the performance of image motion operating on low signal to noise ratio images', *34th Midwest Symposium on Circuits and Systems*, Vol. 2, pp. 1124-1128, NY 1992.

[3] A. Murat Tekalp, 'Digital video processing', Prentice-Hall, 1995.

[4] Bashkaran V. and Konstantinides K., 'Image and video compression standards: algorithms and architectures', Kluwer Academic Publishers, 1995.

[5] Rao K. R. and Hwang J. J., 'Techniques and standards for image, video and audio Coding', Prentice-Hall, 1996.

[6] Tziritas G., 'Motion analysis for image sequence coding', Elsevier Science 1994.

[7] Namazi N. M., Penafiel P. and Fan C. M., 'Nonuniform image motion estimation using Kalman filtering', *IEEE Transactions on Image Processing*, Vol. 3, No. 5, pp. 678-683, 1994.

[8] Burl J. B., 'A reduced order extended Kalman filter for sequential images containing a moving object', *IEEE Transactions on Image Processing*, Vol. 2, No. 3, pp. 285-295, 1993.

[9] Kuo C.-M., Hsieh C.-H., Jou Y.-D., Lin H.-C., Lu P.-C., 'Motion estimation for video compression using Kalman filtering', *IEEE Transactions on Broadcasting*, Vol. 42, No. 2, pp. 110-116, 1996

[10] Ruiz V. 'Estimation et prédiction d'un système évoluant de façon non-linéaire. Filtrage par densités approchées', *PhD-thesis, University of Rouen*, France, 19 March 1993.

[11] V. Ruiz, A. N. Skodras, 'Motion estimation through approximated densities', *13th International Conference on Digital Signal Processing*, Santorini, Greece, Vol. 2, pp. 805-808, 1997.