A PRINCIPAL COMPONENT REGRESSION STRATEGY FOR ESTIMATING MOTION

Vania V. Estrela1, M.H. Da Silva Bassani2, and J. T. de Assis3

1 Scientific Computing Group (NCC), State University of Western Rio de Janeiro (UEZO)
Rua Manoel Caldeira de Alvarenga nº 1203, Campo Grande, RJ, Brazil, CEP 23070-120
vestrela@iprj.uerj.br

2 Mechanical Engineering Department, DEPES, CEFET-RJ
Av. Maracana 229, Rio de Janeiro, RJ, Brazil, CEP 20271-110
marcos@cefet-rj.br

3 State University of Rio de Janeiro (UERJ), Instituto Politécnico (IPRJ-UERJ)
CP 972825 – CEP 28601-970 – Nova Friburgo, RJ, Brazil
joaquim@iprj.uerj.br

ABSTRACT
In this paper, we derive a principal component regression (PCR) method for estimating the optical flow between frames of video sequences according to a pel-recursive manner. This is an easy alternative to dealing with mixtures of motion vectors due to the lack of too much prior information on their statistics (although they are supposed to be normal). The 2D motion vector estimation takes into consideration local image properties. The main advantage of the developed procedure is that no knowledge of the noise distribution is necessary. Preliminary experiments indicate that this approach provides robust estimates of the optical flow.

KEY WORDS
Motion estimation, principal component regression, and surveillance.

1. Introduction
Motion estimation algorithms have found a number of applications ranging from video coding to rendering of scenes in virtual reality environments. The evolution of an image sequence motion field can also help other image processing tasks in multimedia applications such as analysis, recognition, tracking, restoration, collision avoidance and segmentation of objects.

In coding applications, a block-based approach is often used for interpolation of lost information between key frames as shown by Tekalp [14]. The fixed rectangular partitioning of the image used by some block-based approaches often separates visually meaningful image features. Pel-recursive schemes ([6], [7], [13], [14]) can theoretically overcome some of the limitations associated with blocks by assigning a unique motion vector to each pixel. Intermediate frames are then constructed by resampling the image at locations determined by linear interpolation of the motion vectors. The pel-recursive approach can also manage motion with sub-pixel accuracy. The update of the motion estimate was based on the minimization of the displaced frame difference (DFD) at a pixel. In the absence of additional assumptions about the pixel motion, this estimation problem becomes “ill-posed” because of the following problems: a) occlusion; b) the solution to the 2D motion estimation problem is not unique (aperture problem); and c) the solution does not continuously depend on the data due to the fact that motion estimation is highly sensitive to the presence of observation noise in video images.

We propose to solve optical flow (OF) problems by means of a well-established technique for dimensionality reduction named principal component analysis (PCA) and is closely related to the EM framework presented in [4] and [15]. Such approach accounts better for the statistical properties of the errors present in the scenes than the solution proposed in previous works([1],[4) relying on the assumption that the contaminating noise has Gaussian distribution.

Most methods assume that there is little or no interference between the individual sample constituents or that all the constituents in the samples are known ahead of time. In real world samples, it is very unusual, if not entirely impossible to know the entire composition of a mixture sample. Sometimes, only the quantities of a few constituents in very complex mixtures of multiple constituents are of interest ([2],[13],[15]).

In this paper we propose a new method for estimating optical flow which incorporates a general PCR model. Our method is based on the works of Jollife [12] and Wold [16] and involves a much simpler computational
procedure than previous attempts at addressing mixtures such as the ones found in [2], [13] and [15].

In the next section, we will set up a model for our optical flow estimation problem, and in the following section present a brief review of previous work in this area. Section 2 states the simplest and most common form of regression: the ordinary least squares (OLS), as well as one of its extensions, the regularized least squares, RLS, ([5],[8],[9],[10],[11],[12],[14]). Section 3 introduces the principal component regression concepts, describing our proposed technique and shows some experiments used to access the performance of our proposed algorithm. Finally, a conclusion, a discussion of the results and future research plans are presented in Section 4.

2. Problem Characterization

The displacement of each picture element in each frame forms the displacement vector field (DVF) and its estimation can be done using at least two successive frames. The DVF is the 2D motion resulting from the apparent motion of the image brightness (OF). A vector is assigned to each point in the image.

A pixel belongs to a moving area if its intensity has changed between consecutive frames. Hence, our goal is to find the corresponding intensity value \( I(r) \) of the \( k \)-th frame at location \( r = [x, y]^T \), and \( d(r) = [d_x, d_y]^T \) the corresponding (true) displacement vector (DV) at the working point \( r \) in the current frame. Pel-recursive algorithms minimize the DFD function in a small area containing the working point assuming constant image intensity along the motion trajectory. The DFD is defined by

\[
\Delta(r; d(r)) = I_k(r) - I_{k-1}(r - d(r)), \tag{1}
\]

and the perfect registration of frames will result in \( I_k(r) = I_{k-1}(r - d(r)) \). The DFD represents the error due to the nonlinear temporal prediction of the intensity field through the DV. The relationship between the DVF and the intensity field is nonlinear. An estimate of \( d(r) \), is obtained by directly minimizing \( \Delta(r, d(r)) \) or by determining a linear relationship between these two variables through some model. This is accomplished by using the Taylor series expansion of \( I_k(r - d(r)) \) about the location \( (r - d(r)) \), where \( d(r) \) represents a prediction of \( d(r) \) in \( i \)-th step. This results in

\[
(r, r - d'(r)) = u^T \nabla I_{k-1} (r - d(r)) + e(r, d(r)), \tag{2}
\]

where the displacement update vector \( u = [u_x, u_y]^T = d(r) - d'(r) \), \( e(r, d(r)) \) represents the error resulting from the truncation of the higher order terms (linearization error) and \( \nabla = [\delta \delta_{x}, \delta \delta_{y}]^T \) represents the spatial gradient operator. Applying (2) to all points in a region \( R \) of pixels neighboring the current pixel being studied gives

\[
z = Gu + n, \tag{3}
\]

where the temporal gradients \( \Delta(r, r - d'(r)) \) have been stacked to form the \( N \times 1 \) observation vector \( z \) containing DFD information on all the pixels in a neighborhood \( R \), the \( N \times 2 \) matrix \( G \) is obtained by stacking the spatial gradient operators at each observation, and the error terms have formed the \( N \times 1 \) noise vector \( n \) which is assumed Gaussian with \( n ~ N(0, \sigma^2 I) \). Each row of \( G \) has entries \( [g_{xi}, g_{yi}]^T \), with \( i = 1, \ldots, N \). The spatial gradients of \( I_{k-1} \) are calculated through a bilinear interpolation scheme similar to what is done in [4] and [5].

Once we have stated the problem, we need to investigate possible ways of solving (3). This work concentrates its attention on regression-like methods.

The pseudo-inverse or ordinary least-squares (OLS) estimate of the update vector is

\[
u_{LS} = (G^T G)^{-1} G^T z, \tag{4}
\]

which is given by the minimizer of the functional

\[
J(u) = \|z - Gu\|^2 \tag{for more details, see [4], [5] and [7]).
\]

From now on, \( G \) will be analyzed as being an \( N \times p \) matrix in order to make the whole theoretical discussion easier. Since \( G \) may be very often ill-conditioned, the solution given by the previous expression will be usually unacceptable due to the noise amplification resulting from the calculation of the inverse matrix \( G^T G \). In other words, the data are erroneous or noisy. Therefore, one cannot expect an exact solution for the previous equation, but rather an approximation according to some procedure. The assumptions made about \( n \) for least squares estimation are \( E(n) = 0 \), and \( \text{Var}(n) = \sigma^2 I_n \), where \( E(n) \) is the expected value (mean) of \( n \), \( I_n \) is the identity matrix of order \( N \), and \( n^T \) is the transpose of \( n \).

The simplest way to stabilize the matrix inverse is to add a constant to the diagonal. This is a consequence of minimizing the functional resulting in the regularized solution obtained in [5] and [9]. The formal expression is

\[
u_{RLS} = (G^T G + \lambda I)^{-1} G^T z, \tag{5}
\]

where \( \lambda \) is the regularization coefficient. The resulting regression coefficients are biased as a result of the use of regularization. As their variance decrease, however, the model becomes more stable with respect to the prediction error. The variance-bias dilemma is a general problem in multivariate regression. The model bias can only be reduced at the expense of increased model variance and vice versa. The expected prediction error is the sum of two components:

\[
E(\hat{z} - z)^2 = (\text{model bias})^2 + (\text{model variance}). \tag{6}
\]
3. Principal Component Regression (PCR)

PCA is a useful method to solve problems including exploratory data analysis, classification, variable decorrelation prior to the use of neural networks, pattern recognition, data compression, and noise reduction, for example.

The formulation of PCA implies a Gaussian latent variable model and can easily lead to Bayesian models. This technique is used whenever uncorrelated linear combinations of variables are wanted, which reduces the dimensions of a set of variables by reconstructing them into uncorrelated combinations of another set of variables, called principal components (PC’s). The first PC accounts for the largest part of the variance. The second, for the next largest variance and so on.

PCR is based on principal component analysis (PCA) of the matrix $G$ and it is related to its second statistical moment, which is proportional to $G^T G$ according to the expression

$$G = T \lambda$$

where matrix $T$ contains the orthogonal eigenvectors of $G^T G$ (also called scores) ordered in descending order by their eigenvalues ([11], [12]). If $P$ has the same rank as $G$, i.e., $P$ contains the eigenvectors to all non-zero eigenvalues, then $T = GP$ is a rotation of $G$. As said before, PCA assumes a Gaussian variable model and it is simpler than other approaches such as the one in [2]. In essence, PCR is just multiple linear regression of PCA scores on $z$:

$$u_{PCR} = P(T^T T)^{-1} T^T z.$$  (7)

In PCR, the inverse matrix is stabilized in an altogether different way than in regularized least squares regression (RLS, as seen in [4] and [5]) regression. The main drawback of PCR is that the largest variation in $G$ might not correlate with $z$ and therefore the method may require the use of a more complex model. Some nice properties of the PCR are:

1) Using a complete set of PCs, PCR will produce the same results as the original OLS, but with possibly more accuracy if the original $G^T G$ matrix has inversion problems.

2) If $G^T G$ is nearly singular, a solution better than the one given by the OLS can be obtained. If $G^T G$ is singular, the vector associated with the zero root may point towards removing one or more of the original variables.

3) The regression coefficients will be uncorrelated and the amounts explained by each PC will be independent and, hence, additive so that the results may be reported in the form of an analysis of variance. If the PCs can be easily interpreted, the resultant regression equations may be more meaningful.

The division of objects into groups can be simplified in the space generated by the PCs. Then, the groups can characterizing by the DV’s corresponding to the center of each class. Two measures can be used to determine whether a sample belongs to a specific class or not: the Mahalanobis distance to the class center and the norm of the residual. If a set of observations is plotted with respect to the first two PCs, then one can easily apprehend that there is a strong suggestion of distinct groups on which convex hulls and ellipses have been drawn around them. It is likely that the clusters found would correspond to different types of displacement vectors.

PCR provides additional information about the data being analyzed. The eigenvalues of the correlation matrix of predictor variables play an important role in detecting multicollinearity and in analyzing its effects. The PCR estimates are biased, but may be more accurate than OLS estimates in terms of mean square error. One should first consider using OLS on a reduced set of variables, if it is unsatisfactory, then PCR can be used. Besides exploring the most obvious approach, it reduces the computer load.

Figure 1. An example of cluster analysis obtained by means of principal components.

Outliers should not be automatically removed, because they are not necessarily bad observations, since they can signal some change in the scene context or that the data is not Gaussian or that the model is not linear. If additional knowledge on the existence of borders is used, then one’s ability to predict the correct motion will increase. To classify a new observation, a “distance” of the observation from the hyperplane defined by the retained PCs is calculated for each group. If new observations are to be assigned to one and only one class,
then assignment is to the cluster from which the distance is minimized. As it is not close to any of the existing groups, it may be an outlier or come from a new group about which there is currently no information. Conversely, if the classes are not all well separated, some future observations may have small distances from more than one population.

At this time, we have done an analysis of the PCR for motion estimation purposes and the results for SNR=20 dB are shown in Fig. 1. In the meantime, they can function as a way of illustrating the performance obtained with our approach. The SNR is defined as $SNR = 10\log_{10} (\sigma^2/\sigma_c^2)$, where $\sigma^2$ is the variance of the original image and $\sigma_c^2$ is the variance of the noise corrupted image.

4. Conclusion

In this paper, we extended the PCR framework to the detection of motion fields. The algorithm developed combines regression and PCA which is primarily a data-analytic method that obtains linear transformations of a group of correlated variables such that certain optimal conditions are achieved. The resulting transformed variables are uncorrelated.

Unlike other works ([8],[11],[12]), we are not interested in reducing the dimensionality of the feature space describing the different behaviors inside a neighborhood surrounding a pixel. Instead, we use them in order to validate motion estimates. This can be seen as a simple alternative way of dealing with mixtures of motion displacement vectors.

We need to test the proposed algorithm with different noisy images and look closely at the discriminant analysis scheme performance. It is also necessary to incorporate more statistical information in our model and analyze if it will improve the outcome.

References


