MOTION ESTIMATION BY MEANS OF PRINCIPAL REGRESSION

Vania V. Estrela – vestrela@iprj.uerj.br
A. N. P. Boente – boente.uezo@faetec.rj.gov.br
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
Joaquim T. de Assis – joaquim@iprj.uerj.br
State University of Rio de Janeiro (UERJ), Instituto Politécnico
CP 972825 – CEP 28601-970 – Nova Friburgo, RJ, Brazil
Fernanda D. S. Alves – fds.rosa@yahoo.com.br

Abstract. *This paper presents an approach where the optical flow between frames of video sequences is estimated according to a pel-recursive strategy through the use of a principal component regression (PCR). This is a simple choice when it comes to treat mixtures of motion vectors due to the fact that it is not essential to have too much knowledge on their statistical properties (although they are supposed to be normal). Local image properties are taken into consideration in order to estimate the 2D motion. The main advantages of the developed procedure are: (i) the noise distribution is not used directly; and (ii) the types of motion present in neighborhood can be identified, so that a discriminant-type of test can be performed. Preliminary experiments indicate that this approach provides robust estimates of the optical flow.*

Keywords: *Motion estimation, principal component regression, computer vision, video sequences.*

1. INTRODUCTION

Motion estimation has been an active, interesting and important area of research since the 1970's. It has many applications including video coding, rendering of scenes in virtual reality environments, satellite remote sensing, surveillance, vehicle navigation and robot guidance. 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. It is important to recall that these areas of research have a close connection to and are not always easy to distinguish from each other, particularly in real-world applications.

Motion detection is employed to verify if there is motion between two video frames and the result is normally a region of interest indicating where motion occurs. An illustration of its use would be attempting to detect when changes in a scene are caused by real-world motion and not, for example, shadows caused by changes in illumination. Motion estimation is the process of identifying motion between at least two frames within a region of interest. In this situation, the output is normally the magnitude and direction of the motion vectors for the region of interest, either as a whole or on a per-pixel basis. Motion tracking is concerned with tracking an object across multiple frames, and it involves further processing (segmentation, identification, filtering) of the motion estimates. The motion tracking output would normally be the current position and velocity of the target object. Motion compensation is used in video coding standards to decrease the size of the compressed sequence by removing as much temporal redundancy as possible.

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. Without auxiliary assumptions about the motion of each pixel, 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 unravel optical flow (OF) problems through a well-stablished technique for dimensionality reduction called 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 offered in previous works ([1],[4) relying on the assumption that the contaminating noise has a normal distribution.

Most methods assume that there is little or no interference between the individual examplar constituents or that all the elements 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]).

This article proposes a new method for estimating optical flow which incorporates a general PCR model. This 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, stating 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 4 introduces the concepts of PCA, and principal component regression. Section 5 describes our proposed technique and shows some experiments used to access the performance of our proposed algorithm. Finally, a discussion of the results and future research plans can be found in Section 6.

2. PROBLEM FORMULATION

Displacement Estimation

The displacement of each pixel in each frame forms the displacement vector field (DVF) and its estimation can be done using at least two adjacent frames. The DVF is the 2D motion resulting from the apparent motion of the image intensities (OF), so that a specific displacement vector is assigned to each point in the image.

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

$$\Delta(\mathbf{r};\boldsymbol{d}(\mathbf{r}))=I_k(\mathbf{r})-I_{k-1}(\mathbf{r}-\boldsymbol{d}(\mathbf{r})), \tag{1}$$

and the perfect registration of frames will result in $I_k(\mathbf{r})=I_{k-1}(\mathbf{r}-\mathbf{d}(\mathbf{r}))$. 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 $\mathbf{d}(\mathbf{r})$, is obtained by directly minimizing $\Delta(\mathbf{r},\mathbf{d}(\mathbf{r}))$ or by determining a linear relationship between these two variables through some model. This is accomplished by using the Taylor series expansion of $I_{k-1}(\mathbf{r}-\mathbf{d}(\mathbf{r}))$ about the location $(\mathbf{r}-\mathbf{d}^i(\mathbf{r}))$, where $\mathbf{d}^i(\mathbf{r})$ represents a prediction of $\mathbf{d}(\mathbf{r})$ in i-th step. This results in

$$\Delta(\mathbf{r},\mathbf{r}-\mathbf{d}^i(\mathbf{r}))=-\mathbf{u}^T \nabla I_{k-1}(\mathbf{r}-\mathbf{d}^i(\mathbf{r}))+e(\mathbf{r}, \mathbf{d}(\mathbf{r})), \tag{2}$$

where the displacement update vector $\mathbf{u}=[u_x, u_y]^T = \mathbf{d}(\mathbf{r}) - \mathbf{d}^i(\mathbf{r})$, $e(\mathbf{r}, \mathbf{d}(\mathbf{r}))$ represents the error resulting from the truncation of the higher order terms (linearization error) and $\nabla=[\partial/\partial_x, \partial/\partial_y]^T$ represents the spatial gradient operator. Applying (2) to all points in a region $\mathcal{R}$ of pixels neighboring the current pixel being studied gives

$$\mathbf{z} = \mathbf{G}\mathbf{u}+\mathbf{n}, \tag{3}$$

where the temporal gradients $\Delta(\mathbf{r}, \mathbf{r}-\mathbf{d}^i(\mathbf{r}))$ have been stacked to form the N×1 observation vector $\mathbf{z}$ containing DFD information on all the pixels in a neighborhood $\mathcal{R}$, the N×2 matrix $\mathbf{G}$ is obtained by stacking the spatial gradient operators at each observation, and the error terms have formed the N×1 noise vector $\mathbf{n}$ which is assumed Gaussian with $\boldsymbol{n}\sim N(0, \sigma_n^2\mathbf{I})$. Each row of $\mathbf{G}$ has entries $[g_{xi}, g_{yi}]^T$, with i = 1, …, 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 the previous expression. This paper concentrates its attention on regression-like methods.

3. THE TRADITIONAL WAY: ORDINARY LEAST SQUARES

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

$$\mathbf{u}_{LS} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{z} \tag{4}$$

which is given by the minimizer of the functional $J(\mathbf{u})=\|\mathbf{z}-\mathbf{G}\mathbf{u}\|^2$ (for more details, see [4], [5] and [7]). From now on, $\mathbf{G}$ will be analyzed as being an N×p matrix in order to make the whole theoretical discussion easier.

Since $\mathbf{G}$ may be very often ill–conditioned, the solution given by (4) will be usually unacceptable due to the noise amplification resulting from the calculation of the inverse matrix $\mathbf{G}^T\mathbf{G}$ and the fact that its elements are approximation of derivatives. 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 $\mathbf{n}$ for least squares estimation are $\mathbf{E}(\mathbf{n})=\mathbf{0}$, and $\mathrm{Var}(\mathbf{n})=\mathbf{E}(\mathbf{n}\mathbf{n}^T)=\sigma^2\mathbf{I}_N$, where $\mathbf{E}(\mathbf{n})$ is the expected value (mean) of $\mathbf{n}$, $\mathbf{I}_N$, is the identity matrix of order N, and $\mathbf{n}^T$ is the transpose of $\mathbf{n}$.

The simplest way to stabilize the matrix inverse is to add a constant to the diagonal. This is a consequence of minimizing the functional

$$J_{\lambda}(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\,\|\mathbf{Q}\mathbf{x}\|_2^2,$$

resulting in the biased regularized solution obtained in [5] and [9] and shown in (5). Useful properties related to some prior knowledge about the problem is seized by the regularization operator $\mathbf{Q}$. In general, $\mathbf{Q}$ is selected so that the new result is smoother than the one given by the ordinary LS approach. This function is expected to have some regularity properties such as continuity and differentiability. The key point for obtaining sensible parameter estimates via regularization is the ability to use the additional qualitative information brought in by $\mathbf{Q}$ and to pick up the adequate regularization parameters. The formal expression is

$$\mathbf{u}_{\mathbf{RLS}} = (\mathbf{G}^{T}\mathbf{G} + \lambda\,\mathbf{I})^{-1}\mathbf{G}^{T}\,\mathbf{z}, \tag{5}$$

where λ 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 impasse 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:

$$\mathrm{E}((\hat{\mathbf{z}}\text{-}\mathbf{z})^2) = (\text{model bias})^2 + (\text{model variance}). \tag{6}$$

4. PRINCIPAL COMPONENT REGRESSION (PCR)

PCR relies upon the principal component analysis of the **G** data matrix. Several algorithms are available ([10],[11],[12]). They yield the same result, but differ in memory and computing time required.

The classical algorithms use the matrix $\mathbf{G}^{\mathbf{T}}\mathbf{G}$ in one way or another. This has dimensions p×p.

Principal Component Analysis (PCA)

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 approach is used whenever uncorrelated linear combinations of variables are wanted. It reduces the dimensions of a set of variables by reconstructing them into uncorrelated combinations. The variables that account for the largest part of the variance to form the first PC are combined. The second PC accounts for the next largest amount of variance, and so on, until the complete sample set variance is combined into progressively smaller uncorrelated component categories. Each successive component explains portions of the variance in the total sample. All of the components are uncorrelated with each other. PCA relates to the second statistical moment of **G**, which is proportional to $\mathbf{G}^{\mathbf{T}}\mathbf{G}$ and it partitions **G** into two matrices **T** and **P**, such that:

$$\mathbf{G} = \mathbf{T}\mathbf{P}^{T}. \tag{7}$$

Matrix **T** contains the eigenvectors of $\mathbf{G}^{\mathbf{T}}\mathbf{G}$ ordered by their eigenvalues with the largest first and in descending order. 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**. The first column of **P**, $\mathbf{p}_1$, gives the direction that minimizes the orthogonal distances from the samples to their projection onto this vector. This means that the first column of **T** represents the largest possible sum of squares as compared to any other direction in $\mathbb{R}^N$. It is customary to center the variables in matrix **G** prior to using PCA. This makes $\mathbf{G}^T\mathbf{G}$ proportional to the variance-covariance matrix. The first principal axis is then the direction in which the data have the largest spread. **T** and **P** can be found by means of singular value decomposition and, sometimes, are called scores and loadings respectively,. The number of components can be chosen via examination of the eigenvalues or, for instance, considering the residual error from cross-validation ([5].[12]). Due to the nature of our stated motion estimation problem, we will keep the PCs and use them to group displacement vectors inside a neighborhood. The resulting clusters give an idea about the mixture of motion vectors inside the mask.

Principal Component Regression

As its name implies, this method is closely related to principal components analysis. In essence, the method is just multiple linear regression of PCA scores on **z** using a suitable number of principal components. The formal solution may be written as:

$$\mathbf{u}_{\mathbf{PCR}} = \mathbf{P}(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{z}\ . \tag{8}$$

In principal components regression (PCR) the matrix inverse is stabilized in an altogether different way than in RLS regression. The scores vectors (columns in **T**) of different components are orthogonal. PCR uses a truncated inverse where only the scores corresponding to large eigenvalues are included. The main disadvantage 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 $\mathbf{G}^T\mathbf{G}$ matrix has inversion problems.
2) If $\mathbf{G}^T\mathbf{G}$ is nearly singular, a solution better than the one given by the OLS can be obtained by means of a reduced set of PC´s due to the calculated variances . If $\mathbf{G}^T\mathbf{G}$ is singular, the vector associated with the zero root may point towards removing one or more of the original variables.
3) Since the PCs are uncorrelated, straightforward significance tests may be employed that do not need be concerned with the order in which the PCs were entered into the regression model. The regression coefficients will be uncorrelated and the amounts explained by each PC are independent and hence additive so that the results may be reported in the form of an analysis of variance.
4) If the PCs can be easily interpreted, the resultant regression equations may be more meaningful.

The criteria for deciding when PCR estimators are superior to the OLS estimators depend on the values of the true regression coefficients in the model.

5. PROPOSED STRATEGY AND EXPERIMENTS

The division of objects into groups can be formulated as a mathematical problem consisting of finding the boundaries between the groups (clustering). This discrimination between objects can be highly nonlinear. An object is, or is not, a member of a class, depending on which side of the class boundary it lies.

Linear discriminant analysis is limited to situations where a sample belongs to exactly one of m classes. Sometimes the problem is such that a sample may belong to more than one class at the same time, or not belong to any class. In this method each class is modelled by a multivariate normal in the score space from PCA. Two measures are used to determine whether a sample belongs to a specific class or not. One is the leverage—the Mahalanobis distance to the center of the class, the class boundary being computable as an ellipse (please, see Fig. 1). The other is the norm of the residual, which must be lower than a critical value.

In Fig. 1, a set of observations is plotted with respect to the first two PCs. One can easily apprehend that there is a strong suggestion of four distinct groups on which convex hulls and ellipses have been drawn around the four suspected groups. It is likely that the four clusters shown correspond to four different types of displacement vectors. For a big neighborhood, it could happen that these vectors would not be readily distinguished using only one variable at a time, but the plot with respect to the two PCs clearly distinguishes the four populations.

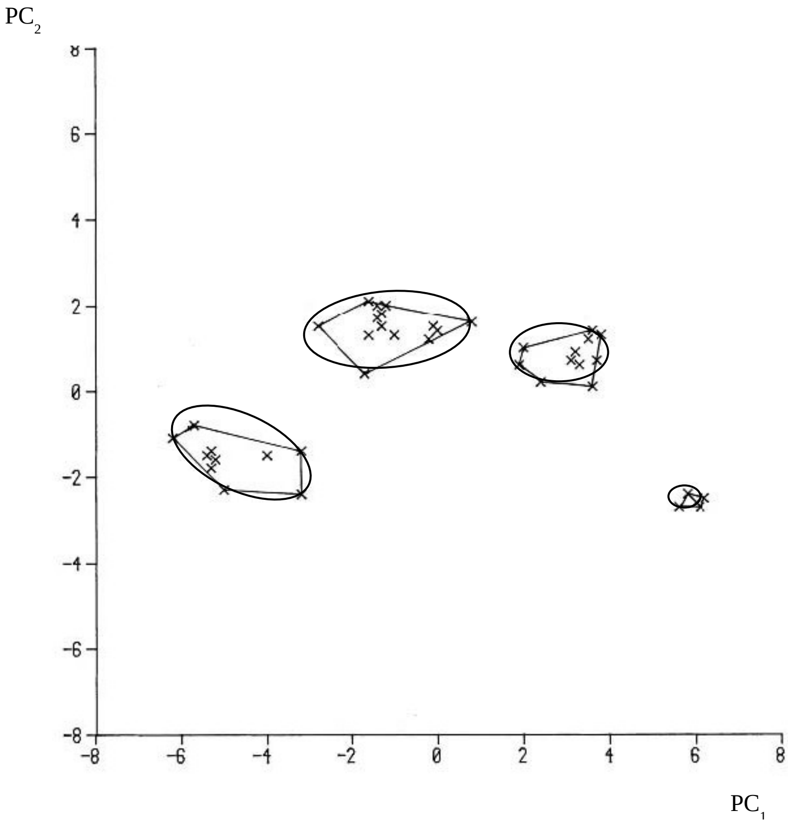


Figure 1 - An example of cluster analysis obtained by means of principal components

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. It is impossible to evaluate the gain in accuracy for a specific problem since a comparison of the two methods to OLS since for real video sequence, knowledge of the true values of the coefficients is required. Nevertheless, when severe multicollinearity is suspected, it is recommended that at least one set of estimates in addition to the OLS estimates be computed since these estimates may help interpreting the data in a different way. There is no strong theoretical justification for using principal components regression. We recommend that the methods be used in the presence of severe multicollinearity as a visual diagnostic tool for judging the suitability of the data for least squares analysis.

When principal component analysis reveals the instability of a particular data set, one should first consider using least squares regression on a reduced set of variables. If least squares regression is still unsatisfactory, only then should principal components be used. Besides exploring the most obvious approach, it reduces the computer load.

Outliers and other observations should not be automatically removed, because they are not necessarily bad observations. As a matter of fact, they can signal some change in the scene context and if they make sense according to the above-mentioned criteria, they may be the most informative points in the data. For example, they may indicate that the data did not come from a normal population or that the model is not linear.

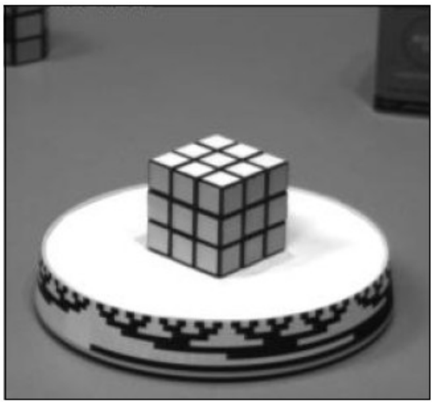
When cluster analysis is used for dissection, the aim of a two-dimensional plot with respect to the first two PCs will almost always be to verify that a given dissection 'looks' reasonable. Hence, the diagnosis of areas containing motion discontinuities can be significantly improved. If additional knowledge on the existence of borders is used, then ones ability to predict the correct motion will increase.

Principal components can be used in cluster (discriminant) analysis, given the links between regression and discrimination. The fact that separation between populations may be in the directions of the last few PCs does not mean that PCs should not be used at all in discriminant analysis. In regression, their uncorrelatedness implies that a linear discriminant analysis shows the contribution of each PC and that they can be assessed independently. This is an advantage compared to using the original variables, where the contribution of one of the variables depends on which other variables are also included in the analysis, unless all elements are uncorrelated. This latter approach holds for types of discriminant analysis in which the covariance structure is assumed to be the same for all populations. However, it is not always realistic to make this assumption, in which case some form of non-linear discriminant analysis may be necessary. The same type of quantity is also used for detecting outliers. 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. Alternatively, a tough decision may not be required and, if all the distances are large enough, the observation can be left unassigned. 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. In such cases, it may again be undesirable to decide on a single possible class; instead two or more groups may be listed as possible *loci* for the observation.

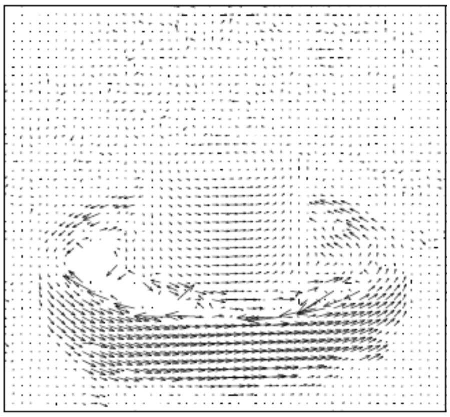
At this time, we have done an analysis of the PCR for motion estimation purposes and the results for SNR=20 dB are shown in Fig, 2. In the meantime, they can function as a way of illustrating the performance obtained with our approach. The SNR is defined in [4] and [5] as

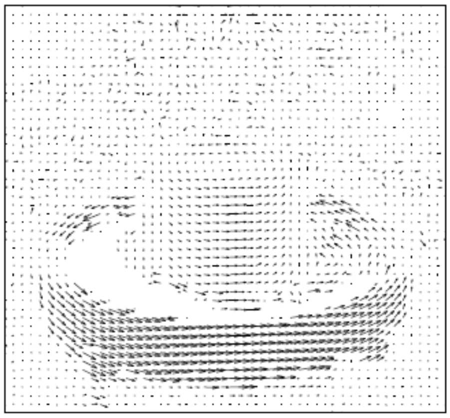
$$SNR = 10\log_{10} (\sigma^2/\sigma_c^2), \tag{9}$$

where σ^2 is the variance of the original image and σ_c^2 is the variance of the noise corrupted image.



(a)





(b) (c)

Figure 2 - Displacement field for the Rubik Cube sequence. (a) Frame of Rubik Cube Sequence; (b) Corresponding DVF for a 31×31 mask obtained by means of PCR with SNR=20 dB; and (c) PCR with discriminant analysis, 31×31 mask and edge information with SNR=20 dB.

6. CONCLUSIONS

One area that is of recent interest in motion estimation is robust techniques. They have been applied to many areas with some excellent results. In theory, it is desirable to have a robust technique for use in applications relying on motion detection. However, in practice a trade-off must be made between the computational cost and the performance increase.

Even using a dense set of optical flow measurements, least-squares (LS) strategies and analogous unweighted methods are sensitive to errors in the optical flow. Classically, the errors are assumed to be normally distributed with a zero mean and unknown standard deviation. Traditional techniques perform optimally when this model holds. However, real data does not often completely satisfy these assumptions, in particular data containing outliers. Two possibilities to solve this problem are outlier diagnostics and robust regression. Outlier diagnostics tries to identify and remove data outliers before applying regression while robust regression performs the regression while minimizing the effect of outliers on the result.

In this paper, the PCR framework was extended 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. As a result, the transformed variables are uncorrelated. Unlike other works ([8],[11],[12]), there is no interest in reducing the dimensionality of the feature space which describes 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.

The proposed algorithm has to undergo more tests with different types of motion and noise. The performance of the discriminant analysis scheme can also benefit from further studies. It is also necessary to incorporate more statistical information to our model and analyze if it will improve the outcome.

REFERENCES

[1] Biemond, J., Looijenga, L., Boekee, D. E., Plompen, R. H. J. M., A Pel-Recursive Wiener-Based Displacement Estimation Algorithm, Signal Proc., 13, December (1987), 399-412.

[2] Blekas, K., Likas, A., Galatsanos, N.P., Lagaris. I.E., A Spatially-Constrained Mixture Model for Image Segmentation, IEEE Trans. on Neural Networks,vol. 16, (2005) 494-498.

[3] Chattefuee, S.,. Hadi, A.S., Regression Analysis by Example, John Wiley & Sons, Inc., Hoboken, New Jersey, (2006)

[4] Estrela, V.V., da Silva Bassani, M.H., Expectation-Maximization Technique and Spatial Adaptation Applied to Pel-Recursive Motion Estimation, Proc. of the SCI 2004, Orlando, Florida, USA, July (2004), 204-209.

[5]Estrela, V.V., Galatsanos, N.P., Spatially-Adaptive Regularized Pel-Recursive Motion Estimation Based on Cross-Validation, Proc. of ICIP-98 (IEEE International Conference on Image Processing), Vol. III, Chicago, IL, USA, October (1998), 200-203.

[6]Franz, M.O. and Krapp, H.G. , Wide-Field, Motion-Sensitive Neurons and Matched Filters for Optic Flow Fields. Biological Cybernetics 83, (2000)185-197.

[7]Franz, M.O., Chahl, J.S. , Linear Combinations of Optic Flow Vectors for Estimating Self-Motion - a Real-World Test of a Neural Model. Advances in Neural Information Processing Systems 15, 1343-1350. (Eds.) Becker, S., S. Thrun and K. Obermayer, MIT Press, Cambridge, MA (2003).

[8]Fukunaga, K., Introduction to Statistical Pattern Recognition, Computer Science and Scientific Computing. Academic Press, San Diego, 2 ed. (1990).

[9]Galatsanos, N.P., Katsaggelos, A.K., Methods for Choosing the Regularization Parameter and Estimating the Noise Variance in Image Restoration and Their Relation, IEEE Trans. Image Processing, Vol No , July (1992), 322-336.

[10]Haykin, S., Neural Networks: A Comprehensive Foundation. Prentice Hall, London, 2 ed. (1999)

[11]Jackson, J.E., A User's Guide to Principal Components, John Wiley & Sons, Inc., (1991).

[12]Jolliffe, I., Principal Component Analysis. New York: Springer-Verlag, (1986).

[13]Kim, K. I., Franz, M., Schölkopf, B., Iterative Kernel Principal Component Analysis for Image Modeling. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(9), (2005) 1351-1366.

[14]Tekalp, A. M., Digital Video Processing, Prentice-Hall, New Jersey (1995).

[15]Tipping, M.E. and Bishop, C.M.. Probabilistic Principal Component Analysis. J. R. Statist. Soc. B, 61, (1999), 611–622.

[16]Wold, S., Albano, C., Dunn, W.J., Esbensen, K., Hellberg, S., et. al., Pattern Recognition: Finding and Using Regularities in Multivariate Data, Food Research and Data Analysis, eds. H. Martens and H. Russwurm, London, Applied Science Publishers, (1983), 147–188.