CASCADED NONLINEAR PRINCIPAL COMPONENTS ANALYSIS: NA APPLICATION IN EXTRACTION OF HUMAN MOVEMENTS FROM VIDEO SEQUENCES

Matheus Figueiredo†
matheus@cpgei.cefetpr.br

Silvia Botelho*
silviacb@ee.furg.br

Rodrigo de Bem*
rodrigo@ee.furg.br

Tânia Mezzadri Centeno†
mezzadri@cpgei.cefetpr.br

William Lautenschlager*
pink@ee.furg.br

*FURG
Av. Itália km 8
CEP, Rio Grande, RS, Brazil

†CEFETPR
7 de setembro 3165
CEP, Curitiba, PR, Brazil

ABSTRACT

This article presents a methodology to extract principal components of large set of data, called C-NLPCA (Cascaded nonlinear principal component analysis), and evaluates its use in the extraction of main human movements in image series, aiming for the development of methodologies and techniques for skill transfer from humans to robotic/virtual agents. The C-NLPCA is an original data multivariate analysis method based on the NLPCA (Nonlinear Principal Component Analysis). This method has as main features the capability of taking principal variability components from a large set of data, considering the existence of possible nonlinear relations among them. The proposed method is used to extract principal movements of video sequence of human activities, which can be reconstructed in cybernetic and robotic contexts. Aiming for the validation of the method a human moving hand test is presented, where C-NLPCA is applied and the patterns of the movements obtained from it are confronted with linear traditional techniques.

KEYWORDS: neural networks, PCA, image processing, skill, transfer, robotic.

RESUMO

Este artigo tem como principal objetivo apresentar a metodologia empregada para o processamento do C-NLPCA (Cascaded Nonlinear Principal Component Analysis), e avaliar seu uso na extração de componentes principais de movimentos de series de imagens digitais 2D, visando ao desenvolvimento e metodologias e técnicas que permitam a execução de tal tarefa com eficiência e robustez. O método C-NLPCA é um método de análise multivariada de dados, baseado no NLPCA (Nonlinear Principal Component Analysis), que tem como principais características a capacidade de extrair componentes principais de variação de grande conjuntos de variáveis.

PALAVRAS-CHAVE: redes neurais, PCA, processamento de imagens, transferência de habilidades.

1 INTRODUÇÃO

The principal components analysis (PCA) is a statistic method used for data multivariate analysis (J. F. Hair et al., 1995). Such a method provides linear relations among the elements in a set of variables, giving as output the principal components (PC) which describe the variability patterns in such a set. However, when the variables have nonlinear relationships among then, classical PCA can not be applied. In this case, we have a set of alternatives methods (Lee, 2000; Diamantaras and Kung, 1996; Scholz and Vigrio, 2002). An efficient method is NLPCA (Nonlinear Principal Component Analysis) (Kramer, 1991). This approach takes advantage of the capacity to treat nonlinearities from the artificial neural networks (ANN). Nevertheless, due to internal saturation problems of the ANN (Malthouse, 1998; Hsieh, 2001), the application of the NLPCA method is restricted to the data analysis with limited number of variables involved (limited dimensional size).

The authors have been working for the extension of the NLPCA technique to allow treatment of higher dimension data. So, it is proposed a structure in layers, where NLPCAs are cascaded. Such a methodology, C-NLPCA (Cascaded Nonlinear Principal Components Analysis), was initially used for the determination of temporal series variability patterns of satellite images representing the ocean superficial temperature (Botelho et al., 2003). The results obtained were encouraging, motivating the interest of its use in other applications. For instance, one could capture frames of a flying bird, analyzing the animal's main movements. The C-NLPCA could provide the translation of the bird through the image, as the first PC (Principal Component), the movement of beating its wings, as the result of the second PC extraction and so on.

C-NLPCA in skill capture tasks - Recent advances in computer science have not been paralleled by corresponding advances in robot/virtual capabilities or the development of "intelligent (virtual) machines". This disparity is principally caused by the difficulty of formalizing intelligent human behavior and decision making processes into an algorithmic framework. Humans manage a set of tasks, such as manipulation, and mobility, with relative easiness and although humans are quite successful at executing these tasks, they are not successful to describe this process formally. Thus, in the literature, we find a set of works that have focused on learning computational models of human skill, aiming their transfer to robotic/virtual agents. For instance, one extracts Principal Components of movement in a sequence of human gestures, captured by a cyberglove (Lee, 2000). From joint sensor information, a humanoid Robot is guided by humans in a walking/dancing tasks, learning these skills (Tatani and Nakamura, 2003). These approaches works with a limited degree of freedom (two or three maximum), using a limited set of discrete sensors to extract principal components of human/robot movements. We are proposing an alternative methodology: to use, as environment perception tool, a sequence of video frames, which captures directly the human movement. This approach can bring together a set of advantages like the need of a simple camera used as sensor, the high level of details that a image frame can have, etc. However, there are many problems inherent to the image treatment, like high dimension of the data, occlusion problems, sensibility to lighting conditions, etc (Lin et al., 2000; Dorner, 1994).

We argue that the high dimension problem associated with skill capture (capture of the main movements of a certain human activity) from video sequence can be a good application for our Cascaded Nonlinear Principal Components Analysis. Thus, in this paper, it is studied the use of C-NLPCA for the determination of movement patterns made by a human hand obtained from images temporal series, captured by a camera. These patterns could be used to transfer skills between humans and robots, virtual elements, etc.

The first focus of this paper will be an explanation over multivariate analysis methods (PCA, NLPCA), aiming to illustrate their uses in correlated works. Later on, it will be done a specific discussion about peculiar details of the C-NLPCA, in order to show results of experiments as well as an analysis of those.

2 MULTIVARIATE ANALYSIS

Image Vector Independently of their nature, temporal data samples can be viewed as a vector. For instance, we can consider the dataset as a set of images, whose width and height associated are *w* and *h* pixels respectively. Thus, the number of components (pixels) of this vector will be $w*h$. Each pixel is coded by one vector component. The construction of this vector, called *image vector* $\vec{X}$, from an image is performed by a simple concatenation - the rows of the image are placed each beside one another (Romdhani et al., 1999).

Image Space The *image vector* belongs to a space, called *image space*, which is the space of all images whose dimension is $w*h$ pixels. Thus, when plotting the *image vectors* they tend to group together to form a narrow cluster in the common image space.

3 PRINCIPAL COMPONENT ANALYSES METHODS

3.1 Principal Component Analysis

Due to the cluster feature containing our *image vectors*, the full *image space* may be not an optimal space for our data description. This happens because we can have similarities (redundant information) between same components of different images. Thus, for several domains it can be interesting to build a new space, lower in size than the original. The base vectors of this new space are called the Principal Components. Thus, the goal of the Principal Component Analysis is to reduce the dimension of the original set or space so that the new basis better describes the typical "models" of the set. The new basis vectors (axes) will be constructed by a linear combination (thus they are essentially orthogonal), catching the total variance in the set of images.

Theory of PCA Let $\vec{X}(t)=[x_1,...,x_p]$ be a dataset, with dimension *p*, where each variable x_i, (*i* = 1,…, *p*) is a time series containing *n* observations. PCA transformation is given by a linear combination of the x_i, time function *u*, and an associated vector a such as:

$$u(t)=a*\vec{X}(t) \tag{1}$$

so that,

$$<<\| \vec{X}(t)-au(t) \|^2>> \tag{2}$$

is minimized (<< … >> denotes a sample or time mean). Here *u*, called the first principal component (PC), is a time series, while *a,* the first eigenvector of the data covariance matrix, often describes a spatial pattern. From the residual, $\vec{X}-au$, the second PCA mode (present variation in the residue in the first mode) can be obtained, and so on for higher modes (see (Hsieh,2001) for more details).

3.2 The theory of NLPCA

PCA only allows a linear mapping from $\vec{X}$ to *u*. On the other hand, NLPC is obtained using a multi layer Neural Network, see figure 1 (Kirby and Sirovich, 1990). To perform NLPCA, the ANN contains 3 hidden layers of neurons between the input and output layers. Hidden layers have nonlinear activation functions between the input and bottleneck layers and between the bottleneck and output layers. Hence, the network models a composition of functions. The five-layer NLPCA network has *p* nodes in the input layer, *r* nodes in the third (bottleneck) layer, and *p* in the output layer. Output layer must reproduce the input signals presented to the network. The nodes in layer 2 and 4 must have nonlinear activation functions, and the nodes in layer 1, 3 and 5 use linear activation functions. NLPCA network allows data compression/reduction because the *p-dimensional* inputs must pass through the *r-dimensional* bottleneck layer before reproducing the inputs. Once the network has been trained, the bottleneck node activation values give the temporal patterns (PCs).

Let $\mathbf{f}:\Re^p \rightarrow \Re^r$ denotes the function modeled by layers 1, 2 and 3, and let $\mathbf{s}: \Re^r \rightarrow \Re^p$ denotes the function modeled by layers 3, 4 and 5. Using this notation, the weights in the NLPCA network are determined under the following objective function:

$$\min \sum_{l=1}^{n} \| \vec{X}_l - \vec{X}_l^{'} \| \qquad (3)$$

where $\vec{X}^{'}$ is the output of the network and *n* is the number of training samples. The relation associated with *u* and $\vec{X}$ is now generalized to $u = f(\vec{X})$, where *f* can be any nonlinear function explained by a feed-forward ANN mapping from input layer to the bottleneck layer and instead of PCA, $\| \vec{X}_l - \vec{X}_l^{'} \|$ is minimized by nonlinear mapping functions, $\vec{X}^{'} = s(u)$.

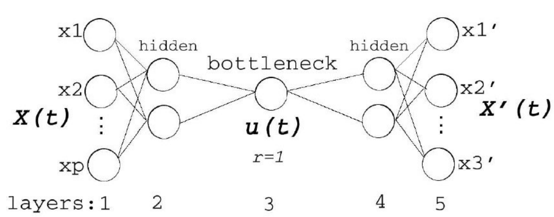


Figure 1: NLPCA – Neural Network to map Nonlinear Components.

The residual $\| \vec{X}_l - \vec{X}_l^{'} \|$, can be input into the same network structure to extract the second NLPCA mode, and so on for the higher modes (Monahan, 2000).

4 C-NLPCA: CASCADED NONLINEAR PRINCIPAL COMPONENT ANALYSES

When it is necessary to run NLPCA with large dimension datasets, like images, there is a noticeable increase of parameters (weights) associated with the neurons of the ANN. Thus, leading to the necessity to have a bigger number of temporal samples, so that this value can be near to the parameters of the ANN (Hsieh, 2001). It is known that sometimes it is not possible to attend this requirement even if one accepts the saturation and poor dimensional reduction risks. Moreover, the addition of more samples increases in an expressive way the computational overhead to conclude the analysis.

Thus, when the original dataset have many dimensions, several authors use a data filtering before the NLPCA analysis, as reduction techniques (Botelho et al., 2003; Hsieh, 2001). Using the former approach, the simplification introduced by the use of linear PCA analysis can lead to erroneous outputs or, at least, can produce coarser results.

Our C-NLPCA aims to allow the direct and totally nonlinear analysis of high dimension dataset, using a cascaded set of successive NLPCAs, see figure 2. The architecture is composed by two main stages: reduction and expansion. Images are decomposed into a set o small windows, which will be reduced and grouped by successive NLPCAs at **reduction stage**. A bottleneck NLPCA gives the final principal component. This value is expanded by the second stage (**expansion stage**), resulting in an output of the same dimension of the original dataset (special expansion).

Obtaining the C-NLPC - The Reduction Process

C-NLPCA assumes that $p^{'}$ (determined by observation) is the *ideal* dimension for the input data. The *ideal* concept is associated with the relationship between parameters number (weights) and the number of temporal samples. Thus, we divide the original input image with dimension p into smaller windows with dimension $p^{'}$. These windows are used directly as input of a first layer of NLPCAs. Each NLPCA, constrained by the saturation requirements, finds a local principal component (local reduction) of one window. The resulting patterns (reductions) are used as input to a new layer of C-NLPCA. This process is repeated until only one pattern is left, thus giving the final reduction of whole original dataset. Despite during the first step the windows are independently analyzed, in the second step the neighbor relations are considered, ensuing that the results are grouped in succession.

We consider the window $\vec{W}(t)$ as a subset of pixels of the *image vector*, $\vec{X}(t)$. For each component $\vec{W}_i(t)$ of $\vec{X}(t)$, we process the nonlinear analysis (section 3.2) with a standard NLPC network. Each network associated with $\vec{W}_i(t)$ is called *NLPCA_i*. They compose the first layer with $m1 = p / p^{'}$ networks. The bottleneck results $u_i(t)$ of each window $NLPC_i$ are grouped into a new second layer of NLPCAs. Notice that, the number of NLPCAs of the second layer is $m2 = m1 / p^{'}$ (we use sub-index 1,2 to describe, respectively the first and the second layer of the cascade). The process is repeated up to layer *n* composed by only one NLPCA network. The output of the bottleneck neuron of this only NLPCA is the nonlinear principal component, $u_n(t)$ of the original large dataset.

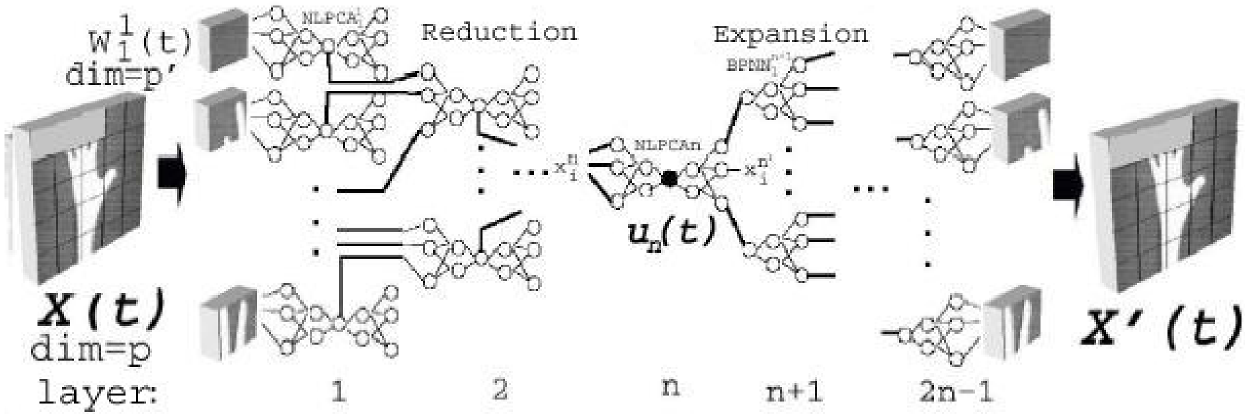


Figure 2: C-NLPCA – a set of layers grouped in Reduction and Expansion Stages. The pointed out neuron gives the pattern associated to the reduction of the original dataset.

Obtaining C-NLPCAs - The expansion Stage

The second role of the Principal Component Analysis, called expansion, is to obtain the data associated with each principal component reduction (PC, C-NLPC) in the original dimension of the image (PCA, CNLPCA). Moreover, each time when a principal component k is calculated and we desire to obtain the next (k+1) component, the expansion process is also necessary to calculate the residues associated with k, which will be the input to the (**k**+1) C-NLPC analysis.

Hence, due to the cascading process, we have lost the original dimension of the input image, it is then necessary a method to obtain the expansion of the reduction. Expansion Stage is trained to expand the nonlinear principal component, $u_n(t)$, in a set of nonlinear expanded images. In fact, we propose a bottleneck layered structure to obtain the reduction/expansion of NLPCs.

Expansion layers are symmetric with reduction layers, resulting in a total of $(2*n-1)$ layers. They are composed by simple backpropagation networks **BPNN** (without bottleneck neuron). The input of each **BPNN** is an output of the last layer. We use the original propagated image to train the desired outputs of **BPNN** networks.

Training BPNNs: The training process begins with the first **BPNNs** layer (layer n+1). To train each network $BPNN_i^{(n+1)}$ of this layer is used, as input, the resulting output $x_i^{'n}$ of the bottleneck NLPC layer n. For this training process, the desired output is the input data, $x_i^{'(n-1)}$, of the symmetric layer $(n-1)$ in reduction stage. This process is repeated up to $(2n-1)$ layer, which has as desired output each pixel of original image.

Thus, the expanded/reconstructed image represent the original input taking into account only the current principal component. We use all components of the original dataset, their neighbors relations and temporal variabilities. The method can be applied independently of the dataset dimension size. It also maintains the nonlinearity associated with ANN, avoiding the saturation restriction associated with them.

5 EXPERIMENTS AND RESULTS

The idea is to use the C-NLPCA to extract principal patterns of the variability for different movements, independent from its nature or modeling difficulty, presented in images temporal series. In this context CNLPCA can be used in an ample range of possible applications (see (de Bem, 2005) for a set of different application).

5.1 Obtain Nonlinear Principal Components from a set of Synthetic Images

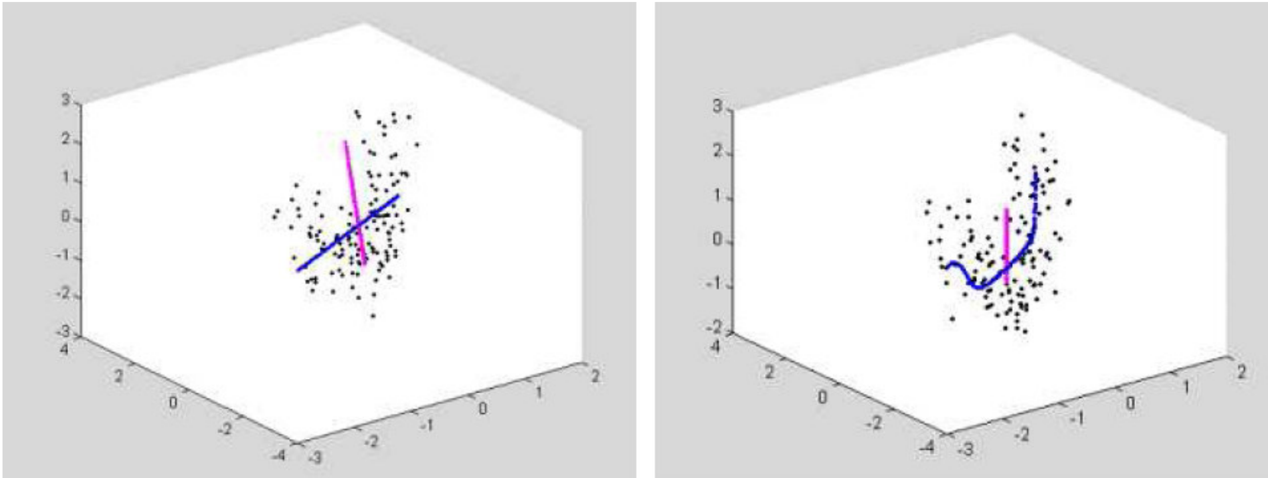
First, we intend to compare both methods, PCA and C-NLPCA, looking to show the advantages of C-NLPCA utilization. In high dimension data sets cases, the C-NLPCA method is applied directly in data, without the necessity of a dimension reduction like the one used at NLPCA method. This way, a set of synthetic imagens with 2 modes known variation ways was produced, likes the pre-set functions. It was increased the importance of the first mode looking for highlight the results.

The data set is formed by 128 points in time and images with 3.600 pixels (60 x 60), with three set of points (X1, X2 and X3, random dispersion) varying according to the pre-set functions showed in figure 3.

Figure 3: Synthesized produced images.

The first variation mode is described by:

X1= $10(t-0.3t^2)$, X2= $10(t+0.3t^3)$ e X3= $10(t^2)$,



(a) **(b)**

Figure 4: Principal components captured by (a) PCA and (b) C-NLPCA(first mode-dark; second mode light).

and the second variation mode is described by:

$X1' = s-0.3s^2$, $X2'=s-0.3s^3$ e $X3'=-s^4$,

Where the magnitude of the first mode is tem times the second mode.

The PCA method captured the two variations modes according to the method limitations, generating the principal components showed in figure 4(a). The application of the C-NLPCA method, applied directly on data, as expected, capture two variation modes. The first mode capture almost all the variability of the set, leaving only one residue in the second mode, showed on figure 4(b), possibly because of the difference of importance (10 times) imposed between the two modes.

In opposition to other works (Hsieh, 2001; Monahan, 2000), where a pre-filter stage is necessary to apply NLPCA method, here the C-NLPCA is able to extract the images principal components, being applied directly on the data. The method C-NLPCA is applied in high dimension data sets (3600 pixels, in this case) without only pre-processing, proving the method efficiency.

5.2 C-NLPCA for extraction of Principal Components of Movements

We present some experiments involving skill transference. More precisely, we are interested to identify principal movements made in a sequence of human gestures, captured through a static camera. These movements, once expanded, will be performed by a robotic or virtual hand.

In the experiments, in a 25 frames/second rate, a temporal series of 128 gray scale images were analyzed, each one with 60 pixels in width and height, resulting in an image vector with 3600 elements. The video was obtained in a real and usual light conditions, without any kind of specific treatment. Besides the proper C-NLPCA, the linear method, PCA, was also applied to these dataset, analyzing some results/features, according to:

- Temporal patterns (principal components - PC).
- Reconstructed Image: Obtained image as return of the methods, through the composition of spatial and temporal principal components, that aims to rebuild the samples. In the PCA it is the sum of the products between the spatial and temporal patterns. In the C-NLPCA, it is the sum of the expanded samples. Both sums have as many terms as principal components to be extracted.
- Sample Reconstructed Projection in the 3D space: evolution of the values of all pixels, in the samples succession, in a spatial graph.

The experiment made consisted of the oscillation of a human hand (left-right), as global movement, associated to the movement of opening and closing it, being that a local movement. In this case, three principal components were analyzed, being the first one responsible for explaining 94,2%, the second 1,5% and the third 1,15% of the total data variance. The figure 5 illustrates some samples of the series.

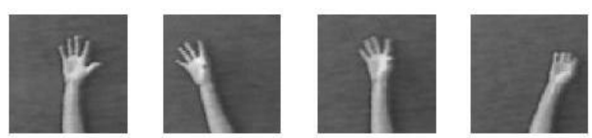


Figure 5: Some frames of the 128 analyzed images series.

For a better analysis we evaluate the nonlinear and linear spatial patterns, PCs (dark lines) and NLPCs (light lines), in the same figure 6 (a,b and c). One can notice that the first linear principal component (PC1, represented in the figure 6(a) in dark line), has a high frequency behavior which does not characterize the movement of the hand (occurring even when there is no meaningful movement of it). That “noise”, coming from the local lighting conditions,

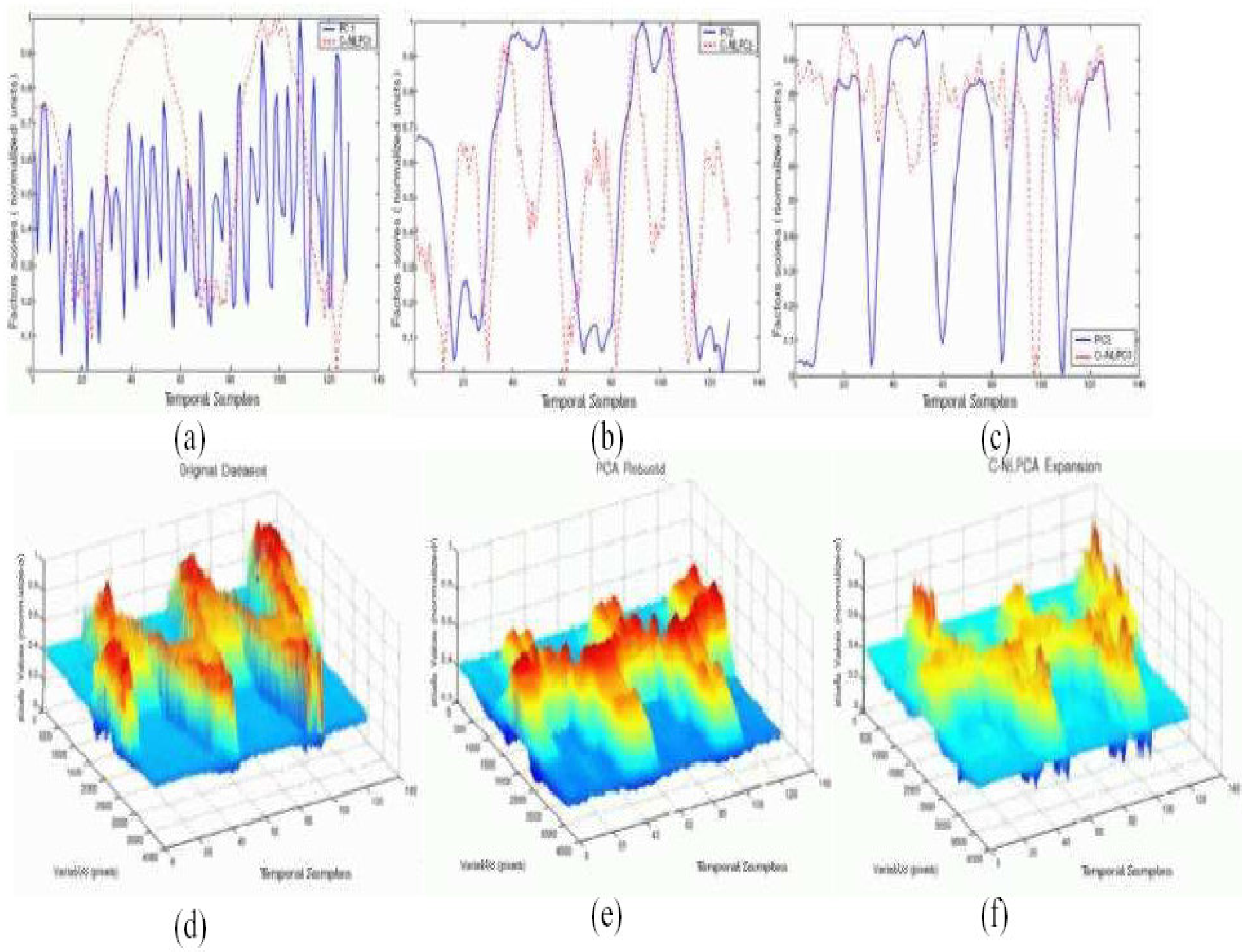


Figure 6: Comparison of nonlinear and linear principal components: (a) first component (b) second component (c) third component. Each pixel value evolution in the image series: (d) original (e) linear reconstruction (f) nonlinear reconstruction.

is a linear small variability of all pixels of each image. It does not occur in the C-NLPC1.

In the C-NLPC1, as well as in the C-NLPC2 (and even in the PC2), it is possible to notice clearly the oscillating patterns of the movement: in the C-NLPC1 the peaks represent the hand in the right side and valleys the hand in the left side (always from the observer's point of view). In the C-NLPC2 we can notice the influence of the shadows and light incidence in the hand due to its localization in the frame. The highest peaks are due to the hand in the right side. Intermediated peaks are the ones that assume the hand in the left side. The valleys show the hand in the center of the frame. The C-NLPC2 is capable even to represent another movement peculiarity: the fact of the closed hand to exceed the image limits in the movements to the right side, due to a fault in a capture process (observable in the local valleys between the highest peaks).

The opening and closing movement of the hand is represented by high frequency signals in C-NLPC3. Closing and opening movements are assumed by peaks and valleys in this series. We can notice that C-NLPC3 shows a different pattern from the others, showing more sensitivity to isolated cases: the fact of the closed hand to exceed the image limits is showed by a deep valley. As expected, the other components are not significant (the main movements of the sequence are the right-left and the opening-closing of the hand).

When observed the three dimensional graph that illustrates all the variability occurred in the series, pixel by pixel, the C-NLPCA proves its capacity to deal with nonlinearities when it does not cut, as the linear method did, the peaks presented in the original image graph (see figure 6 d, e and f).

The principal components of movements can be mapped to robot/virtual coordinated space (joint actuators, motor of wheels, etc...). Nowadays, we are using a multi-layer perceptron neural network to do this.

6 CONCLUSION

In the present study, we propose an original method, called C-NLPCA, to extract principal components of human movement captured by a camera (skill capture), aiming the skill transfer between humans and virtual/robotic agents. C-NLPCA reduces the dimension of large dataset, obtaining principal components of them. We use a cascaded Neural Network in a bottleneck structure to obtain dimension reduction, giving the principal components of the data variability. The same structure is also used to expand the data from obtained principal component.

The method was analyzed using a set of synthetic images. In these tests C-NLPCA allowed to extract nonlinear principal components, instead PCA found linear principal axes. From a video images sequence, the method was applied to obtain principal movements of a human hand. As results, the PCA finds as first mode the lighting conditions variability (not principal, but whole linear). On the other hand, the C-NLPCA network has demonstrated the capability of isolating the principal variability (hand translation) in its first PC. Moreover, closing/opening movements, and besides the fact of the closed hand to exceed the image limits were situations extracted of the three principal C-NLPCs obtained.

Despite of the problems inherent to the NLPCA (Hsieh, 2001), the mosaicking effect of the cascading (windows of frames) and the parameters adjustment difficulty, the technique was able to extract the main components (reduction) and reconstruct the data (expansion), what makes it applicable to movements analysis problems of different natures, presently solved through elaborated modeling, sensors use or pre-filtering by other methods. The experiments conducted to satisfactory results, signaling the C-NLPCA use in other applications, as for instance, the navigation in autonomous vehicles based on images, data compression, etc.

REFERENCES

Botelho, S., de Bem, R., Mata, M., and Almeida, I. (2003). Applying neural networks to study the mesoscale variability of oceanic boundary currents. *LNAI/LNCC*, 2871:684–688.

de Bem, R. (2005). C-nlpcamove - extração de componentes principais de movimento com o método c-nlpca. Technical report, FURG.

Diamantaras, K. I. and Kung, S. Y. (1996). Principal Component Neural Networks: Theory and Applications. Wiley.

Dorner, B. (1994). Chasing the colour glove: visual hand tracking. Master's thesis, Computing Science – Simon Frase University.

Hsieh, W. (2001). Nonlinear principal component analysis by neural network. *Tellus*, 53A:599–615.

J. F. Hair, J., Anderson, R. E., Tatham, R. L., and Black, W. C. (1995). Multivariate data analysis: with readings. Prentice Hall, 4th edition.

Kirby, M. and Sirovich, L. (1990). Application of karhunenloeve procedure for the characterization of human faces. *IEEE On pattern analysis and machine intelligence*.

Kramer, M. A. (1991). Nonlinear principal component analysis using autoassociative neural networks. *AlChe Journal*, 37:233–43.

Lee, C. (2000). *Learning Reduced-Dimension Models of Human Actions*. PhD thesis, The Robotics Institute - CMU.

Lin, J., Wu, Y., and Huang, T. S. (2000). Modeling the constraints of human hand motion. In *Proc of IEEE Human Motion Workshop*, pages 121–126.

Malthouse, E. C. (1998). Limitations of nonlinear PCA as performed with generic neural networks. *IEEE Trans. NN*, 9(1).

Monahan, A. (2000). *Nonlinear principal component analysis of climate data*. PhD thesis, University of British Columbia.

Romdhani, S., Psarrou, A., and Gong, S. (1999). Multiview nonlinear active shape model using kernel pca. In *Tenth British Machine Vision Conference*.

Scholz, M. and Vigrio, R. (2002). Nonlinear pca: a new hierarchical approach. In *ESANN'2002 proceedings - European Symposium on Artificial Neural Networks*, pages 439–444, Bruges(Belgium).

Tatani, K. and Nakamura, Y. (2003). Dimensionality reduction and reproduction with hierarchical nlpca neural networks - extracting common space of multiple humanoid motion patterns. In *IEEE ICRA*, pages 1927– 32.