UNIVERSIDAD NACIONAL DE COLOMBIA

Motion Analysis by Nonlinear Feature Extraction

Juliana Valencia Aguirre

Universidad Nacional de Colombia
Faculty of Engineering and Architecture
Department of Electric, Electronic and Computing Engineering
Manizales, Colombia
2011

Motion Analysis by Nonlinear Feature Extraction

Juliana Valencia Aguirre

Thesis submitted as a partial requirement to receive the grade of:
Master in Engineering - Industrial Automation

Advisor:
Ph.D. Germán Castellanos Domínguez

Co-advisor:
Ph.D. Genaro Daza Santacoloma

Academic Research Group:
Signal Processing and Recognition

Universidad Nacional de Colombia
Faculty of Engineering and Architecture
Department of Electric, Electronic and Computing Engineering
Manizales, Colombia
2011

Análisis de Movimiento Basado en Técnicas de Extracción no Lineal de Caraterísticas

Juliana Valencia Aguirre

Tesis presentada como requisito parcial para optar al título de:
Magister en Ingeniería - Automatización Industrial

Director:
Ph.D. Germán Castellanos Domínguez

Co-director:
Ph.D. Genaro Daza Santacoloma

Grupo de Investigación:
Control y Procesamiento Digital de Seãles

Universidad Nacional de Colombia
Facultad de Ingeniería y Arquitectura
Departamento de Ingeniería Eléctrica, Electrónica y Computación
Manizales, Colombia
2011

A mi papá y a mi mamá por su paciencia. Al grandulón por su apoyo y su ayuda incondicional. A mi amiguita del alma por estar siempre a mi lado. Y a Dios porque sin Él nada de esto hubiera sido posible.

Acknowledgement

Thanks God for for showing me the way, without Your guide I would be lost.

I would like to express my gratitude to my advisor Prof. Germán Castellanos and to the to the Prof. Carlos Daniel Acosta Medina for their guide and orientation during this research. I thank them for the funds and computational equipment that they arranged for me. Also, I would like to thank Prof. Genaro Daza-Santacoloma for his patience, suggestions and hours of academic discussion. I have learned a lot from him.

A very special thanks goes to my family. I know that is not easy to deal with me. I am privileged to have such a great family. And THANKS to my *Grandulón*, you were always by my side through good times and bad times. Your love always gave me hope. I can not forget my friends, who were always by my side to encourage me.

Thanks also goes out to the members of the research group Signal Processing and Recognition at Universidad Nacional de Colombia, not everyone has the opportunity to work with such extraordinary people.

I would like to thank to the projects 20201006570 funded by Universidad Nacional de Colombia, and the project 20201006594 funded by Universidad Nacional de Colombia and Universidad de Caldas.

Finally, I recognize that this research would not have been possible without the financial assistance provided by the scholarship *Beca de Estudiantes Sobresalientes de Posgrado* funded by Universidad Nacional de Colombia.

Abstract

In this work a comparative analysis of nonlinear feature extraction techniques is performed, in order to determine which method is suitable for motion analysis from image and video sequences. Each method is experimentally assessed on artificial and real-world data sets that allow us to visually and quantitatively confirm whether the embedding results were correctly calculated. Then, we propose a framework to find a low-dimensional embedding for data lying on multiple manifolds, which is a common situation in real-world applications such as video and image analysis. This approach is inspired on the Laplacian Eigenmaps - LEM algorithm, computing the relationships among samples of different datasets based on an intra manifold comparison to properly unfold the underlying data structure. The main advantage of this methodology is the possibility for analyzing videos with different conditions, e.g. different amount of frames and/or different image resolution. Finally, we present a methodology for automatic human action recognition, which is based on a new supervised feature extraction technique that includes class label information in the mapping process to enhance both the underlying data structure unfolding and the margin of separability among classes. The presented methodology shows the ability to recognize human actions on a benchmark database, and to be fast enough to be implemented in online applications.

Keywords: Motion analysis, Nonlinear feature extraction, Image Processing, Dimensionality Reduction, Image Processing, Pattern Recognition

Resumen

En este trabajo se realiza un análisis comparativo entre técnicas de extracción no lineal de características, con el fin de determinar cual de ellas es la más apropiada para realizar análisis de movimiento a partir de secuencias de imágenes y vídeo. Cada método fué evaluado experimentalmente en bases de datos artificiales y reales, las cuales permiten confirmar visual y cuantitativamente si las inmersiones fueron calculadas correctamente. Por otra parte, se propone una metodología para calcular un espacio de baja dimensón común para datos que provienen de variedades múltiples, lo cual es usual en aplicaciones de la vida real relacionadas con análisis de imágenes y vídeo. Dicha propuesta está basada en el algoritmo de extracción no lineal de características Laplacian Eigenmaps - LEM, y consiste en calcular las relaciones entre observaciones de diferentes conjuntos de datos a partir de una comparación con las relaciones entre los puntos que pertenecen a la misma variedad. La principal ventaja de la metodología propuesta es que permite encontrar un espacio de baja dimensión para datos heterogéneos, por ejemplo, vídeos que presenten diferente número de de frames y/o diferente resolución de entrada. Finalmente, se presenta una metodología para reconocimiento automático de acciones humanas a partir de secuencias de video, la cual esta basada en una novedosa técnica supervisada de extracción no lineal de características que incluye la información de clase en el proceso de reducción de dimensión, mejorando tanto la representación de los datos como el margen de separabilidad entre clases. La metodología propuesta tiene la capacidad de reconocer acciones humanas en una base de datos reconocida en el estado de arte, y es lo suficientemente rápida para ser utilizada en aplicaciones en línea.

Palabras clave: Análisis de movimiento, Extracción no lineal de características, Procesamiento de imágenes, Reducción de dimensión, Reconocimiento de patrones.

Contents

List of Figures

List of Tables

Part I.

Preliminary

1. Introduction

Vision-based motion analysis aims to detect, identify, and track objects, in other words, it seeks to understand objects behaviors from image sequences. Considering the growing presence of video data and its increasing importance in a wide range of real-world applications, such as, visual surveillance, human-machine interfaces, and sport event interpretation, it is becoming increasingly important to automatically analyze and understand object/subject motions from large amount of video data. Thence, this research area has received growing interest in recent years [1], particularly, the applications of human motion analysis are wide. For instance, researchers in the fields of biomechanics, medicine, sports, and rehabilitation study human locomotion for evaluating joint forces and moments that control motion and posture [2, 3]. Such studies are fundamental for understanding the mechanics of normal and pathological movement [4]. In computer vision, human motion analysis has emerged as an important area of research, motivated by the desire for improved man - machine interfaces [5]. Specifically, recognizing human activities from video is one of the most promising applications of computer vision [6].

Although there has been a lot of progress in the past decades, there are challenging problems that remain unsolved, e.g., robust object/subject detection and tracking, unconstrained object/subject activity recognition, etc. Recently, statistical machine learning algorithms, such as manifold learning [7, 8], have been successfully applied in this area for object tracking, activity modeling, and recognition. In this sense, this techniques have strong potential to further contribute to the development of robust and flexible vision systems. The process of improving the performance of vision systems has also brought new challenges to the field of machine learning and pattern recognition systems, taking into consideration that new machine learning algorithms are able to address more realistic problems in object motion analysis and understanding.

However, in computer vision systems, it is difficult to analyze the available information due to its complexity and the large amount number of features present in video and image sequences. Thereby, to apply techniques for motion modeling, tracking, or classification can be problematic, being necessary to calculate compact representations of the original data. That is, the dimensionality of the feature space should be reduced [9]. In this regard, finding low-dimensional representations of high-dimensional data, such as video and image sequences, is a common problem in data mining, pattern recognition and machine learning. The nonlinear feature extraction methods (also known as nonlinear dimensionality reduction methods) allow to discover the underlying data structure, diminishing irrelevant

or redundant information, and obtaining a graphical representation for visual analysis. However, traditional dimensionality reduction techniques compute linear transformations of the observations [10], and cannot correctly reveal the underlying structures of data that lie on nonlinear manifolds. Hence, it is necessary to explore some nonlinear alternative methods.

Therefore, nonlinear dimensionality reduction algorithms based on manifold learning are commonly used. These techniques aim to compute a low-dimensional representation of the input data, discovering a small number of factors that suitably model the studied phenomena. For example, suppose we have a collection of frames taken from a video of a rotating object. The dimensionality of the dataset is given by the number of pixels in a frame, which is generally very large. Nevertheless, the images are actually governed by only a couple of degrees of freedom (e.g. the angle of rotation and lighting). Such relationship, even though can be consider as nonlinear globally, it is often smooth and approximately linear in a local region [11, 12]. Thence, it is reasonable to presume that the high-dimensional data lie on a manifold, defined by a smaller set of characteristics than in the original feature space. Again, manifold learning techniques have been successfully applied as a tool to analyze video and image datasets [13–15]. There are different methods for nonlinear dimensionality reduction based on manifold learning, such as: Isometric Feature Mapping (ISOMAP) [16], Maximum Variance Unfolding (MVU) [17], Laplacian Eigenmaps (LEM) [18], Locally Linear Embedding (LLE)[12], among others.

Given the diversity of methods for nonlinear dimensionality reduction that have emerged, choosing a specific technique to analyze data that resides on manifold with nonlinear and complex structures has become a difficult task. It is necessary to study thoroughly each method to assess the effectiveness and performance. Motivated by the lack of a systematic comparison of nonlinear dimensionality reduction techniques, this work presents a comparative study of the traditional linear dimensionality reduction technique (Principal Component Analysis - PCA), and some nonlinear feature extraction methods. We tested the methods on synthetic data and real world-datasets for visualization purposes. In this sense, the analysis allow to establish wish of them are suitable to perform dimensionality reduction preserving simultaneously both global and local structure of data. For this purpose, a measure of the quality of the embedding is used and computational load is considered.

On the other hand, methods for nonlinear feature are constrained to single manifold datasets. Considering that in real world applications, like video and image analysis, datasets with multiple manifolds are common, we propose a framework to find a low-dimensional embedding for data lying on multiple manifolds. Our approach is inspired on the manifold learning algorithm LEM, computing the relationships among samples of different datasets based on an intra manifold comparison to unfold properly the data underlying structure. The main advantage of the proposed methodology is the possibility for analyzing dissimilar objects/subjects in appearance but with a common behavior. Moreover, it allows to deal with objects/subjects that have different input dimensions and number of samples among manifolds.