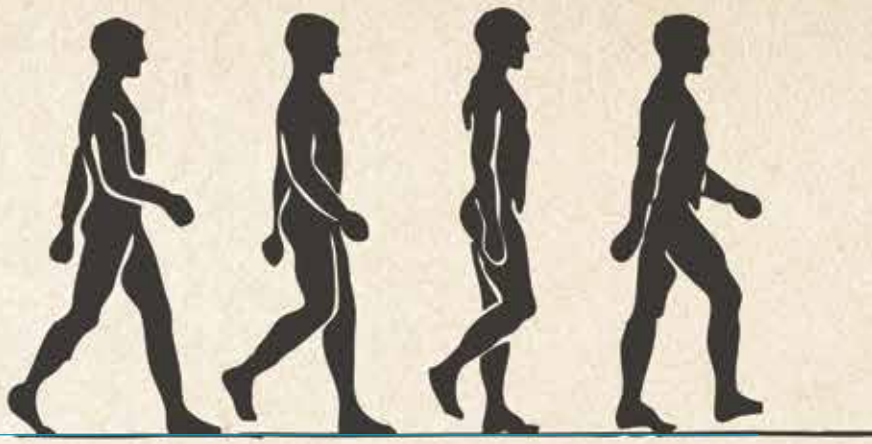
ASSOCIATE PROFESSOR AHMED ELGAMMAL

The shapes of motion



Associate Professor Ahmed Elgammal provides context to his research into generalised separation of style and content on nonlinear manifolds and how they possess application to human motion analysis

To begin, could you outline the motivation for your study of human motion? In what way do you hope to improve understanding?

In the last two decades, extensive research in the computer vision community has focused on the analysis and understanding of human motion in images and videos. This wide interest emanated from various potential real-world applications such as visual surveillance, human-machine interface, video archival and retrieval, computer graphics animation, autonomous driving and virtual reality. Humans are typically the most important subjects in the images and videos of these applications. Researchers have looked at a wide range of problems including detection of humans and their motion, locating faces in images, tracking people and their limbs, recovering body posture, extracting various biometrics, analysing facial expression and hand gestures.

As the human body moves through the 3D world, motion is constrained by body dynamics and projected by lenses to form the visual input we capture through our cameras. Therefore, the changes (deformation) in appearance (texture, contours, edges) in the visual input (images and videos) corresponding to performing certain actions are well constrained by the three dimensional (3D) body structure and the dynamics of the action being performed. Researchers have always tried to explicitly or implicitly exploit such kinematic and dynamic constraints in their models to recover the body configuration.

Despite the high dimensionality of the configuration space, many human motions intrinsically lie on low-dimensional manifolds. This is true for the kinematics of the body, as well as for the observed motion through image sequences. For example, the silhouette (occluding contour) of a human walking is an example of a dynamic shape, where the shape deforms over time based on the action being performed. These deformations are restricted by the physical body and the temporal constraints posed by the action being performed. Given the spatial and the temporal constraints as points in a high-dimensional visual input space, these silhouettes are expected to lie on a low-dimensional manifold. Intuitively, the gait is a 1D manifold that is embedded in a high-dimensional visual space.

What is the principal aim of this investigation?

The main goal of this research project is to have a better understanding of the role of manifold learning as a tool that can mathematically exploit the kinematic and dynamic constraints of the body in motion analysis.

The main contribution of this project is in developing a computational framework for learning models that can explicitly factorise the intrinsic body configuration, as a function of time, from the various appearance factors. The learned models support tasks such as synthesis and body configuration recovery, as well as the recovery of other aspects such as viewpoint, person style parameters, etc.

How will computational models help to aid understanding in the human cognitive science field?

The computational models developed and tested in this project highlight the role that a low-dimensional manifold representation can play in understanding biological motions. They also highlight that an intrinsic representation of motion, invariant of viewpoint and people shapes, can be learned only from visual input.

Will other fields benefit from your mathematical models?

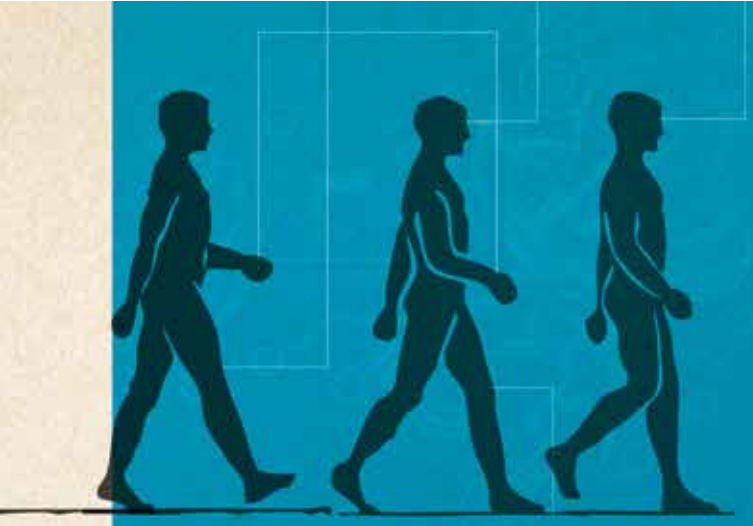
We have applied the models that we developed on various applications of human motion analysis including gait tracking, extracting gait biometrics, analysis and synthesis of facial expression, and analysis of complex motion, such as ballet motion. We are also investigating other applications of the mathematical models in different problems, including object recognition and viewpoint estimation.

What would you say has been your greatest accomplishment since you embarked on this project?

The greatest accomplishment is highlighting that an explicit low-dimensional representation of human motion can effectively help solve the challenging posture estimation problem. Several researchers have followed our lead in investigating manifold-based representations for different human motion analysis problems

Finally, have there been any major challenges to overcome in your mission to understand human motion perception? How have you tackled these issues?

The major challenge is to factor out variability due to viewpoint variations; such as different people shapes from the intrinsic motion. The main remaining challenge is to find a way to break down complex human motion to simple motion primitives or units, where each of these primitives can be modelled as a manifold effectively.



Human motion **analysis**

Keen to explore the drivers of motion, researchers at **Rutgers University** are studying how to separate the style and content on a manifold that represents a dynamic object

IN RECENT YEARS, human activity recognition has garnered much attention in both academic and industrial circles. Driven by advancements in technology, such as semiconductors and microelectromechanical systems, experts have been able to build systems from miniature sensing equipment to recognise human activity and comprehend human behaviour like never before.

Analysing human motion is complex for a number of reasons, not least because the human body is made up of many joints and has many degrees of freedom. Body configuration space is high dimensional and in addition to the coupling of an individual's body style and choice in clothing, poses significant challenges for those wishing to track the limbs of people. Furthermore, due to such factors as posture and perspective, some limbs can obstruct others and make the process of analysis relatively difficult.

In a bid to overcome current challenges, a project unfolding at Rutgers University seeks to establish the role of manifold learning as a tool that can mathematically exploit kinematic and dynamic constraints of the body in motion analysis. The project is considering the visual space of biological motion (manifold) and investigating how to separate the style and content on a manifold representing a dynamic object. Key to the study is an understanding of motion – being the 'content' – as well as its appearance as a function that is affected by style variables.

Heading the research is Ahmed Elgammal, Associate Professor in the Department of Computer Science, whose interest in computer vision research has focused on deciphering information about the three dimensional (3D) world from 2D images which, at times, has proved to be a highly complicated task. A prevalent factor in computational vision, for instance, is object and scene representation. Experts have been exploring different representations, ranging from object-orientated 3D geometric representations to viewer-orientated representations which are highly problematic when delineating dynamic (articulated and deformable) objects. In the context of human motion analysis, Elgammal wants to explain the visual appearance of humans without the need for explicit 3D geometric models and understand representations for their shape and appearance. Such information can then be used to support tasks such as synthesis, pose recovery, reconstruction and tracking.

THE SEPARATION OF STYLE AND CONTENT

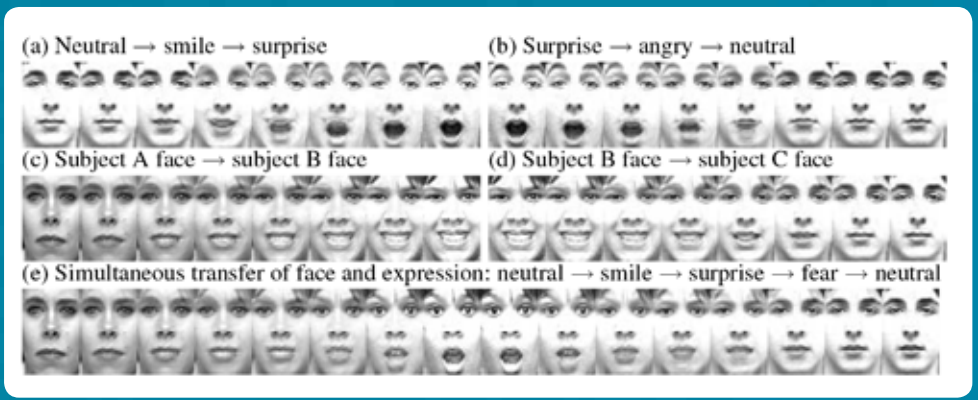
So far, bilinear and multilinear models have proved successful in breaking down static images into orthogonal sources of variations such as in the separation of style and content. Take the appearance of human motion for instance, such as facial expression or gestures; these activities generally result in nonlinear manifolds in the image space and have conventionally been hard to delineate. To combat the seperation of style and content, the team has formulated a mathematical framework that decouples intrinsic body configuration from other sources that affect the visual input. In describing the challenge, Elgammal admits: "The problem is an essential element of visual perception and a fundamental mystery of perception. While the role of manifold representations in perception is still unclear, it is clear that images of the same object lie on a low-dimensional manifold in the visual space defined by the retinal array". Conversely, neurophysiologists have discovered that neural population firing is a normal function of some variables, suggesting that population activities also reside on low-dimensional manifolds.

Generally speaking, a dynamic object's appearance is a function of the intrinsic body configuration or the content. All other factors are referred to as style factors. Given all these factors, Elgammal emphasises that the combined appearance manifold is very challenging to model: "The solution we use utilises the fact that the underline motion manifold, invariant to all other factors, is low in dimensionality. Therefore, the motion manifold can be explicitly modelled, while all the other factors that cause deformation on that manifold can be parameterised".

A FRAMEWORK FOR SEPARATION

Experiments conducted by Elgammal and his team have led to the introduction of a framework for separating style and content on manifolds representing dynamic objects, based on the factorisation of style variables in the space of nonlinear functions. Different methods have been established to acquire a unified content manifold, embedding: first, via unsupervised nonlinear dimensionality of visual data and manifold warping; second, through supervised conceptual embedded representation of the manifold; and third, through nonlinear dimensionality reduction of auxiliary data that is invariant to the visual variability.

The computational models created over the course of the study emphasise the function that a low-dimensional manifold



Facial expression synthesis. **First row**: transition between expressions: (a) from neutral to 50-50 smile-surprise to 100 per cent surprise (at peak); (b) from 100 per cent surprise to 50-50 surprise-anger to neutral. **Second row** (c,d): morphing between different faces during smile expressions. **Third row**: morphing between different faces and different expressions.

GENERALIZED SEPARATION OF STYLE AND CONTENT ON NONLINEAR MANIFOLDS WITH APPLICATION TO HUMAN MOTION ANALYSIS

OBJECTIVES

- To investigate learning of a unified invariant content manifold representation from various style variations on the same manifold
- To investigate factorised generative models for the data-given representation of one or more of the underlying manifolds
- To study representation of the underlying manifold and how that can be used to select discriminative features in the visual input
- To apply the findings towards the recovery of intrinsic body configuration

KEY COLLABORATORS

Assistant Professor Chan-Su Lee, Department of Electronic Engineering, Yeungnam University

FUNDING

National Science Foundation – award no. 0546372

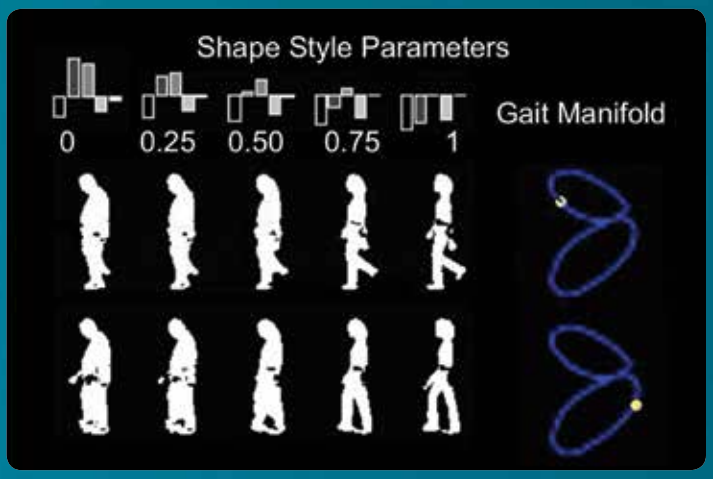
CONTACT

Dr Ahmed Elgammal
Principal Investigator

Department of Computer Science
Rutgers, the State University of New Jersey
110 Frelinghuysen Road
Piscataway
New Jersey 08854-8019
USA

T +1 848 445 8316
E elgammal@cs.rutgers.edu

AHMED ELGAMMAL has been Associate Professor at the Department of Computer Science, Rutgers University since 2002. Elgammal is also a member of the Center for Computational Biomedicine Imaging and Modeling (CBIM). Elgammal received his PhD from the University of Maryland, College Park, in 2002. His primary research interest is computer vision with a focus on human motion analysis.



Interpolation between two shapes (thin female to the right and big male to the left) through a walking motion. Moving along the curve (the gait manifold) generates the walking motion, while changing the style parameters generates different body shapes.

representation has in the understanding of biological motions, as well as explaining that an intrinsic representation of motion can be learned from visual input.

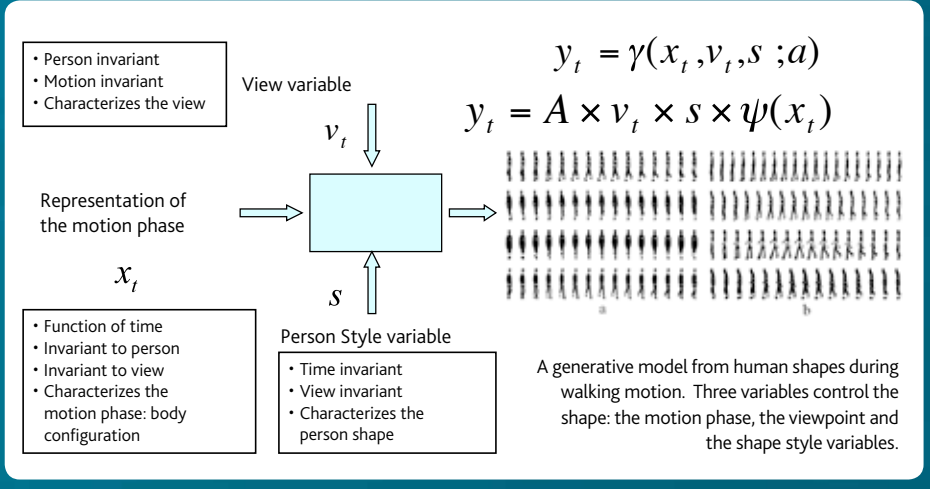
MANIFOLD MODELS FOR MOTION

The project can be separated into several distinct lines of investigation. For modelling view and posture manifolds, the project considers data based on multiple continuous manifolds. Specifically, a shape manifold is taken of a person performing an action. This is recorded from various angles along a view circle at a fixed height. In order to unite the body configuration (kinematics) manifold with the visual (observations) manifold, a model is introduced to allow tracking of the 3D configuration with ongoing view variability. The model takes advantage of the low-dimensional nature of both the body configuration manifold and the view manifold, each being singularly represented. The resulting representation is then employed for tracking complicated actions and motions. Data is analysed within a Bayesian context, where the model not only represents a low dimensional state, but also a constrained dynamic model for kinematic and observation variations.

Torus tracking is used to model visual patterns of a regular articulated action, such as walking, but seen from any perspective. A torus is a geometric object describing a surface of revolution. The study has shown that visual manifolds of such actions as walking can be modelled by mapping the manifold to a toroid topological structure: a torus manifold, in the case of a single view circle, or a family of tori, for the entire view sphere. Unlike classic manifold learning approaches, which are unsupervised and aimed at finding a low-dimensional embedding of the data, this approach is based on learning the visual observation manifold in a supervised manner.

With multi-factor models for style separation, the researchers have been investigating how to model numerous style factors by employing multilinear analysis in the space of nonlinear basis functions. This approach allows different sources of same-motion style variations to be separated. For instance, from different angles, this model can generate data for people's varying walking stances or facial expressions. In reverse, using an input pattern, the researchers integrated this with an optimisation process which can account for the various factors that produce a pattern. For example, in the case of a single shape, the body configuration, viewpoint and person's shape can be recovered. Likewise, with a single face image, facial expression and identity can be recovered.

The project has made – and will continue to make – great progress in understanding human motion, activities and behaviours on an abstract level. Although their are still a few kinks to iron out, its real-world applicability will unquestionably help thousands of people in the near future.



A generative model from human shapes during walking motion. Three variables control the shape: the motion phase, the viewpoint and the shape style variables.