Facial Feature Tracking via Evolutionary Multiobjective Optimization

Eric C. Larson, Oklahoma State University, USA
Gary G. Yen, Oklahoma State University, USA

ABSTRACT

Facial feature tracking for model–based coding has evolved over the past decades. Of particular interest is its application in very low bit rate coding in which optimization is used to analyze head and shoulder sequences. We present the results of a computational experiment in which we apply a combination of non-dominated sorting genetic algorithm and a deterministic search to find optimal facial animation parameters at many bandwidths simultaneously. As objective functions are concerned, peak signal-to-noise ratio is maximized while the total number of facial animation parameters is minimized. Particularly, the algorithm is tested for efficiency and reliability. The results show that the overall methodology works effectively, but that a better error assessment function is needed for future study.

Keywords: Cyclic Optimization, Genetic Algorithm, Facial Feature Tracking, Model-Based Coding, Multiobjective Optimization, NSGA-II

INTRODUCTION

Facial feature tracking has received considerable attention in recent decades for its applications in video gaming, man-machine interaction, model based coding, and numerous other disciplines. Analyzing a head and shoulders sequence remains to be one of the most needed, yet challenging problems in video processing today. Even so, years of research has reduced the problem of facial analysis to a numerical optimization in which classical search methods are often used. See Pearson (1995) for an excellent review of the evolution of model based coding and Pandizic et al. (2003) for the current state of the art.

A dominant amount of attention for facial feature tracking has been spent applying it towards model-based facial coding (MBFC). In model-based coding, the idea is to analyze an object (i.e., the face) and send high level information about the animation of the object like movement and rotation, instead of sending raw video. The decoder uses knowledge of the model and renders the appropriate animation. The face, however, is a particularly difficult object to model because it requires not only rigid body motion, but also deformations to create...
expression. Modern facial models can recreate a human likeness with remarkable clarity. Figure 1 shows three reconstructed facial models and the original frames.

Using a model to transmit video opens the possibility for transmission using dynamic bandwidths (i.e., sending video at different bandwidths dynamically depending on the conditions of the transmission channel). In transform based coding, dynamic bandwidth means increasing or decreasing the compression thresholds, reducing video quality when the compression is high. In MBFC, dynamic bandwidth means sending a dynamic number of parameters of the face, resulting in less movement of the model at high compression but not necessarily less quality.

This application promises to revolutionize video telephony and teleconferencing by drastically reducing the bandwidth required for transmission (Eisert, 2003). But there is no free lunch. The bandwidth reductions in MBFC must be paid for in computational analysis. Despite the advances in facial analysis, current work has been limited in several aspects. Before MBFC can be adequately implemented, the following factors need addressing:

1. **The limitations of gradient-based optimization.** This type of analysis, while showing promise for real-time implementation, inherently relies upon a gradient approximation. This approximation limits the problem scope to facial sequences involving movements built into the gradient approximation training (i.e., small head movements).

2. **The use of static facial parameters.** Current algorithms use a hand selected set of animation parameters on the face, or use all facial animation parameters. It is unclear if bandwidth could be further reduced using a dynamic set of animation parameters, or what the ideal set of parameters is that adequately represents all facial animation sequences.

3. **The prohibitive use of computationally complex algorithms.** The use of direct methods to analyze head and shoulder sequences in real time has been completed only by reducing the number of animation parameters optimized. The resultant frames are not considered high enough quality to be realistically rendered.

In this article, we address all of the stated concerns by formulating the analysis and synthesis into a multiobjective optimization problem (MOP). In short, we want to design a facial coder that simultaneously finds a high

---

*Figure 1. Selected frames of a model-based video. Top: the original video frames. Bottom: frames synthesized from a model based coder. The frames are generated using a facial model; they are rendered sequences derived from stretching a facial “texture” over the three dimensional facial wireframe. See (Eisert, 2000).*
Related Content

What Organizes the World: N-Relational Entities
www.igi-global.com/chapter/organizes-world-relational-entities/28313?camid=4v1a

Research on the Cluster-Level Architectural Knowledge: Exploration of Six Content Dimensions of Architectural Knowledge based on a Case Study
www.igi-global.com/article/research-cluster-level-architectural-knowledge/75332?camid=4v1a

Dynamically Reconfigurable Hardware for Evolving Bio-Inspired Architectures
www.igi-global.com/chapter/dynamically-reconfigurable-hardware-evolving-bio/38448?camid=4v1a