

INITIALIZATION - AN UNSOLVED PROBLEM IN MODEL-BASED CODING

Robert Forchheimer

Linköping University
Dept. of Electrical Engineering
Linköping, SWEDEN

ABSTRACT

This position paper discusses Model-Based Coding and the reasons why we still do not see a deployment of such codecs although real-time systems running on ordinary PC hardware have been demonstrated since several years. It is claimed that this is due to two main factors. First, the practical need for very low bit-rate video communication has diminished due to the wide availability of broadband IP networks. Second, some of the fundamental problems have not yet found sufficiently robust solutions. One of these problems is the initialization problem which will be discussed in particular.

1. INTRODUCTION

Model-based coding of face images is by now becoming a fairly mature research topic within the field of image and video coding. Since its inception in the early 1980s it experienced a rapid growth in popularity and attracted a lot of interest during the following 20 years. Since a few years, several real-time systems have been reported, the more recent ones having the ability to update texture adaptively, to handle difficult illumination, to model the skin of the face in a natural way and to synthesize highly realistic images.

But, although we have been blessed with a world-wide standard for the communication of 3-D face images according to the model-based approach, there are no encoders readily available with our web-cam software or network conferencing systems that naturally switches to model-based coding whenever a face appears in the image. Why is that? Well, one reason is quite likely that it is no longer necessary from a practical point of view to solve the remaining problems that could make model-based coding really work in the way it was once envisioned. Low bit-rate coding, in the order of a few kbits per second is no longer a necessity for video transmission. Even a long-distance Internet connection will yield bit-rates at least one or even two orders of magnitude higher, making it possible to rely on waveforms techniques such as H.264 coding.

Another concern is that we are still far from having achieved a system that is robust enough to be deployed on a large scale and expected to work under a wide variety of lighting conditions. What we have achieved to date is not much more than "demos" illustrating the basic concepts involved in the model-based coding approach. It is possibly fair to say that the receiver (synthesis) part is getting close to "gold quality" thanks to the tremendous growth in the computer graphic field. But the main solutions on the encoding (analysis) side such as face finding, feature detection, model fitting, motion tracking and texture updating are so far more rightly to be described as sophisticated "mock-ups". They will work whenever the researcher stands next to the system, gently poking it whenever it starts straying away in the wrong direction.

Now, given that we face all these difficulties. What would be the reason to continue spending time and effort to attack these problems when this is no longer necessary from a practical point of view? My humble answer is that these problems represent very basic computer vision tasks and that finding solutions will give us deeper insight into the vision process. Our brains are hardwired to detect and analyze faces. Searching for a face is most likely the first vision task performed by the newly born child as it opens its eyes to the world. A model-based coding system is a perfect platform for evaluating the performance of face analysis algorithms. Just by looking at the reconstructed images we are able to see minute discrepancies in face expressions, head movements, lip movements or to detect any unnatural movements.

2. THE INITIALIZATION PROBLEM

Here, I would like to focus on one problem that has been around since the beginning of model-based coding but which has just recently been investigated in some depth. This is the initialization problem or how to fit a generic face model onto a target face in the first video frame. In most model-based coding works the model-fitting is often omitted or handled manually. However, this stage is the

first in face motion estimation and without a good solution it is impossible to build a complete model-based coding system. Furthermore, re-initialization is often necessary during a transmission session. In a recent thesis by Zhengrong Yao at Umeå University [1] new light is shed on this problem. New solutions are proposed but it is also obvious that this is indeed a fundamental problem which still remains to be fully solved.

[1] Zhangrong Yao, "The Initialization Problem in Model-Based Coding" , Umeå University, Licentiate Thesis, ISBN 91-7305-478-X, June 2003.