Research Interests
Sami Romdhani

Human face images depend on several attributes: identity, illumination, pose, expression, age, etc. How cool would it be, if, given two face images or a sequence of images, we could, in an automatic and seamless manner, exchange one for the other: for instance, replacing the expression of one face image by the expression in another face image, leaving all other factors untouched; or changing the illumination of a face image, simulating the change of location in which the actor is photographed.

In a nutshell, my interest is to define a set of semantic attributes and operations that would be extracted from and applied to images or sequence of images of a specific class of objects, such as face images. This would require extracting some parameters of face images, modifying these parameters in some way and generating novel face images from the new parameters, as seen on this example:

Humans are expert at analyzing and facial images. Understanding this process, setting it in a unified formalism leading to automatic and tractable algorithms is my long term goal.

This work has applications in *biometrics* (extracting the identity parameter enables identification), contactless *man-machine interfaces*, *performance driven animation* for the movie industry (generation of a realistic looking facial sequence animated by another sequence), in ultra-low bandwidth *video conferencing*, and in automatic search and annotation of *visual databases*.

This endeavor has ramifications in **Computer Vision**, **Computer Graphics** and **Machine Learning**: Statistical learning algorithms, model selection, optimization, image segmentation, object detection and recognition, visual tracking, 3D graphics, animation and signal processing.

More specifically, a face model must be learned, in which the different sources of variation are factored out and allowed to change independently. The generative model must be accurate enough to be able to synthesize photo-realistic renderings. It must also be feasible to estimate the model parameters from a given input image efficiently, robustly and accurately.

Thomas Vetter, who supervised my PhD, introduced the idea of using Computer Graphics models as prior models for computer vision problems. This lead him to a natural modelling of face images in which, on one hand, pose and illumination are modeled using physical laws and, on the other hand, identity is statistically modeled. This is the only model that can generate photo-realistic images of faces of any individual at any pose and under any illumination. However, the inverse problem, that of optimizing the model parameters such as to match an input face image (estimating the 3D shape, RGB texture and imaging parameters from one image) is complex and prone to failure as it might get trapped in a local minima. In my PhD thesis, I introduced a novel parameter fitting algorithm that is more robust to the local minima problem. This was achieved by combining a top-down and a bottom-up approach and using more image information than previous algorithms. This approach led to state of the art

performance in the difficult problem of identification in the presence of combined pose and illumination variations.

This clearly constitutes a good foundation towards the goal of automatic facial image attributes transformation. However, to reach widespread use in the aforementioned applications, improvements are in order with respect to realism, automatic behavior and efficiency, which can be obtained by:

- Improving the illumination model with more realistic specular highlights, making use of environment maps and allowing cast shadows to be used to refine the 3D shape estimation.
- Leveraging successful computer vision algorithms (structure from motion, bundle adjustment, stereo vision, segmentation) using an accurate statistical prior 3D shape model.
- Integrating face detection and fitting of complex facial models in a uniform formalism.
- Understanding the impact of different facial image modalities on identification performance (high resolution still image vs multiple or sequence of images vs 3D image).
- Improving statistical 3D shape and texture models to handle attributes such as age, expression and emotion. These attributes should be represented by independent parameters and, as, for instance, two persons age differently, the correlation between identity and age should be correctly accounted for.

Certainly, addressing either of these issues would have an impact not only on face image processing but on various Computer Vision problems and applications.