3D-assisted Facial Texture Super-Resolution

Pouria Mortazavian
p.mortazavian@surrey.ac.uk
Josef Kittler
j.kittler@surrey.ac.uk
William Christmas
w.christmas@surrey.ac.uk

Centre for Vision, Speech and Signal Processing
University of Surrey
Guildford, UK
http://www.ee.surrey.ac.uk/cvssp

In this paper we propose a new framework for super-resolving facial images under arbitrary pose. While example-based super-resolution methods have demonstrated impressive results for face super-resolution under given pose and imaging conditions, they have limitations dealing with different poses and illuminations. Due to these limitations their application to face super-resolution in generalised situations is either impractical or sub-optimal.

In example-based face super-resolution, the quality of the super-resolved face depends crucially on how representative the training set is and how well a specific super-resolution method can generalise it and utilise the available information. Due to these issues, example-based methods are limited in handling variations in the subject's pose or other imaging conditions. Most approaches are either limited to one specific pose (*e.g.* [1] or [4]) or a number of pre-defined poses for which training data is available (*e.g.* [3]).

A powerful tool introduced by Blanz and Vetter [2] which can describe and synthesise human faces under a large range of poses and imaging conditions is the 3D morphable model. A 3D morphable model is a vector space representation of 3D faces. Given a single 2D face image as input and a set of landmarks, the parameters of the 3D morphable model can be estimated such that they represent the 3D shape and texture of the input face. This process is called model fitting which estimates the model parameters together with a set of rendering parameters such that rendering the model with the estimated parameters will produce an image which resembles the input face image. Model fitting can be formulated as a *maximum a posteriori* (MAP) estimation of the model and rendering parameters given the input face image and the landmarks. Assuming independence between some parameters:

$$\alpha^*, \beta^*, \rho^* = \underset{\alpha,\beta,\rho}{\operatorname{argmax}}\, P(\alpha,\beta,\rho|f,L)$$
$$= \underset{\alpha,\beta,\rho}{\operatorname{argmax}}\, P(f|\alpha,\beta,\rho)P(L|\alpha,\beta,\rho)P(\alpha)P(\beta)P(\rho) \quad (1)$$

where f is the input face image, L is a set of landmarks marked on f, α and β are the model shape and texture parameters respectively, and ρ is the set of rendering parameters.

In order to use the real facial texture from the input image (as opposed to the texture *estimated* by model fitting), the shape and rendering parameters estimated during the model fitting can be used to extract the facial texture from the input image (where available) and map it to a predefined texture coordinate frame. This coordinate frame is independent of the initial subject pose in the input image. The texture mapping process yields a pose- and shape-normalised 2D facial texture.

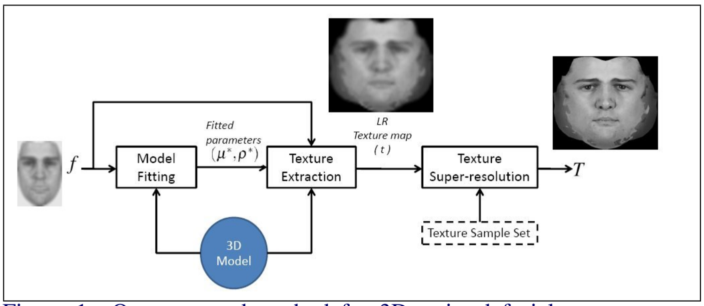

Figure 1: Our proposed method for 3D-assisted facial texture super-resolution

It is on this texture image that we propose to perform example-based super-resolution. That is, we first build a sample set by fitting a 3D morphable model on our training facial images and extracting texture from them. Then, given an input LR image, we super-resolve its texture by fitting the model on the input LR image, extracting the texture and super-resolving the extracted texture. Figure 1 illustrates this process. The super-resolved texture can then be used together with the estimated shape parameters and arbitrary rendering parameters to render a new HR version of the face in the same or a different pose.

We formulate the task as:

$$T^* = \underset{T}{\operatorname{argmax}} \sum_{\mu,\rho} P(T,\mu,\rho|f) = \underset{T}{\operatorname{argmax}} \sum_{\mu,\rho} \{P(T|\mu,\rho,f)P(\mu,\rho|f)\} \quad (2)$$

where μ is the model parameters and T^* is the sought HR texture. Although more appropriate ways of finding T^* from equation 2 are possible, in this paper we make some simplifying assumptions using which we estimate T^* as:

$$T^* = \underset{T}{\operatorname{argmax}}\, P(T|\mu^*,\rho^*,f) = \underset{T}{\operatorname{argmax}}\, P(T|t) = \underset{T}{\operatorname{argmax}}\, P(t|T)P(T) \quad (3)$$

where t is the LR texture extracted from the LR image by fitting the 3D model on the LR input image:

$$t = \text{TEXTURE\_EXTRACT}(\mu^*, \rho^*, f) \quad (4)$$

We use an approach similar to that of Baker and Kanade's face hallucination method [1] with a minor modification in order to obtain T^* in equation 3.

We show that even in this oversimplified case, our framework yields impressive results both in frontal and non-frontal poses. We compare our results in the frontal case with those obtained using Baker and Kanade's well known face hallucination method [1] both visually and in a face recognition experiment. The added value of our framework is that it can be used to super-resolve facial images of arbitrary pose using only frontal training samples.

We conclude that although the model fitting step can add additional error, the final super-resolved face has visually acceptable quality and it can also be used to enhance face recognition with low resolution inputs.

[1] Simon Baker and Takeo Kanade. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):1167 – 1183, September 2002.

[2] Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3d faces. In Alyn Rockwood, editor, *Siggraph 1999, Computer Graphics Proceedings*, pages 187–194, Los Angeles, 1999. Addison Wesley Longman.

[3] Kui Jia and Shaogang Gong. Generalized face super-resolution. *IEEE Transactions on Image Processing*, 17(6):873–886, 2008.

[4] Xiaogang Wang and Xiaoou Tang. Hallucinating face by eigentransformation. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, 35(3):425–434, 2005.