Super-resolution facial images from single input images based on discrete wavelet transform

Ali Mohammed Darvish
Dept. Applied Physics and Electronics
Umeå University
SE-90187, Umeå, Sweden
Email: ali.darvish@gmail.com

Haibo Li
School of Computer Science and Comm.
Royal Institution of Technology (KTH)
SE-10044, Stockholm, Sweden
Email: haiboli@kth.se

Ulrik Söderström
Dept. Applied Physics and Electronics
Umeå University
SE-90187, Umeå, Sweden
Email: ulrik.soderstrom@umu.se

***Abstract*—In this work, we are presenting a technique that allows for accurate estimation of frequencies in higher dimensions than the original image content. This technique uses asymmetrical Principal Component Analysis together with Discrete Wavelet Transform (aPCA-DWT). For example, high quality content can be generated from low quality cameras since the necessary frequencies can be estimated through reliable methods. Within our research, we build models for interpreting facial images where super-resolution versions of human faces can be created. We have worked on several different experiments, extracting the frequency content in order to create models with aPCA-DWT. The results are presented along with experiments of deblurring and zooming beyond the original image resolution. For example, when an image is enlarged 16 times in decoding, the proposed technique outperforms interpolation with more than 7 dB on average.**

Keywords*—Super Resolution, Image generation, Principal Component Analysis, Discrete Wavelet Transform.

I. Introduction and related work

Super-resolution (SR) images are images that are reconstructed with a higher resolution than the original version of the image but still have a high visual quality. It is easy to increase the resolution and at the same time reduce the visual quality, i.e. zooming. SR images have an increased resolution without reducing the visual quality in a serious way. This kind of technique can overcome issues with poor image qualities from low-cost image devices such as surveillance cameras and inexpensive mobile phones. Even if modern smartphone cameras can produce relatively high resolution images, most of them do not have optical zoom. This means that they only facilitate digital zoom, which reduces the resolution in the enlarged area. This problem with loss of image quality can be resolved with SR image reconstruction. A rather conventional, and suitable, approach to creating SR images is to fuse several similar low-resolution images of a particular scene into a high-resolution version of the same. There are solutions that use only a single image to create a SR version. We create SR images from a single image so the related research we review are such works. Information that is not important for human perception, or masked out by other information, can be removed while information that is important is retained, based on knowledge about human perception. In this work, we will use a frequency content model but not to remove redundant information but instead to add relevant frequency information to the existing content of an image. This will allow higher frequencies to be added to a low frequency image, i.e. an image with low resolution, in order to create a high resolution version of the image. Hence, it is possible to improve the resolution of an image, i.e. create a SR version of the image, and allow zoom beyond the original resolution of the image.

Another possible application is deblurring of images. A simple, and widely used, method to remove noise is to filter the image. Filtering can remove noise but may also cause loss of sharpness when high frequency content is removed. With our model it is possible to add frequencies after filtering, yielding a high frequency image without noise. There are many applications where it would be desirable to improve the resolution and/or deblur an image. With our model it is possible to transmit low resolution video content and estimate a high resolution version at the decoder side, meaning that low bandwidth for transmission can be used while the end result has a high resolution.

In this article we will show how to estimate frequencies of a higher resolution from an existing image in a reliable fashion. Encoding can be done with low quality images allowing the use of low cost devices while decoding of high quality content is possible. Even if the original quality of the video is high it can be beneficial to create SR versions of it. This enables the content to be highly compressed and the transmission cost can be lowered while high quality images can be decoded. We will also show how the same technique can be used to deblur images. In a blurry image, the low frequency information is the same as in a sharp image. To create a blurry image you can simply low-pass filter a sharp image. Removal of high frequency content will produce a low-resolution image. From the low-resolution image it is possible to create a sharp image with high resolution by estimating the high frequencies. It is easy to remove frequencies, the problem is how to add relevant frequencies. In this work we focus on a model for facial images and it is only such images that can be improved with our model. We are not describing a general model that works for any kind of image.

A. Related work

A lot of research has been aimed at SR imaging and it is a field that interests many academic researchers as well as companies. Most solutions make use of several images with low resolution and combine them in various ways to create high-resolution results. Image fusion of multiple, aligned images

1051-4651/14 $31.00 © 2014 IEEE
DOI 10.1109/ICPR.2014.155

are used in several solutions and interpolation of images are used in several other approaches. The solutions that are most interesting are those that work with a single image to improve its resolution beyond the original quality. Our work is focused on improving the resolution in a single image, therefore we will not list research where working on multiple images is the main focus. We have video sequences and could have used multiple images to improve the resolution. Our focus, however, is to improve single image resolution.

Yang et al. have come up with a solution that is similar to ours [1]. They use pairs of low-resolution and high-resolution image patches whereby they can estimate a high-resolution image from the content of a low-resolution version of the image. Their idea builds on the fact that image patches can be represented as a sparse linear combination from an over-complete dictionary. They relate the low-resolution and high-resolution image patches into pairs and use the learned image pair as representation, meaning that the amount of data and computational cost are reduced compared to their previous solution [2]. We have compared some of our results with first method. Glasner et al. propose a method which combines the use of correspondence between low and high-resolution image patches and combination of aligned images [3]. Their solution can work on single images and create SR versions of these single images. Their use of image patches is similar to the work presented by Yang et al. Kim and Kwon have a solution based on sparse regression [4]. Their idea is also very similar to ours since they build a mapping of low-resolution images to high-resolution images. This mapping is learned through regression and they have a sparse solution for this to overcome the problem with computational complexity associated with regression. Chang et al. make use of local linear embedding (LLE) to estimate SR content from a low-resolution version [5]. Since small image patches have a similar manifold in both low and high resolution, LLE can model how a SR patch should be modeled by its low resolution neighbors; thus enabling creation of SR content from low resolution content which is similar to it. To improve the visual result further they also employ smoothness constraints and local compatibility in the constructed SR images. Freeman et al. have a solution called example-based SR [6]. In their work they predict the high frequency information in images using only the next lower frequency octave. They also use cubic-spline interpolation for their prediction. The solution is a general one and they only need one generic training set for any kind of image. However, this model needs to be specified to accurately produce the results that the viewer expects, thus creating a need for specific models. An approach that is similar to ours, which is aimed at human faces.

II. Theory and method

In this section we will describe the different techniques we have used and how they are combined.

A. Asymmetrical principal component analysis

Asymmetrical Principal Component Analysis (aPCA) is an extension to Principal Component Analysis (PCA) [7] that connects a subspace with another subspace or the entire eigenspace. A detailed explanation of how aPCA works can be found in [8]. In this work we extend aPCA to the frequency domain; working with Discrete Wavelet Transform (DWT). We will first give a short introduction to how aPCA functions in the spatial domain and then explain how it is used with DWT.

Image parts can be estimated based on other image parts because there is a correspondence between pixels in the different image parts. If there is no correspondence this estimation is not possible. In the spatial domain it is rather easy to see that pixels close to each other have a correspondence, e.g. if the right eye is opened wide the right eyebrow will be raised and the forehead will be wrinkled. With aPCA you need to have two different spaces; one for encoding and one for decoding. In some applications the spaces can model different image areas; for example the side and front view [9] while in some applications the two spaces can be a subspace as well as the entire space, for example the mouth region and the entire face [10]. In this article we will only use the second approach; a subspace consisting of a few frequencies and the entire space with all frequencies.

aPCA is applied to a sequence of images or a video sequence. First we create a space with regular PCA which models only a small area of the images; called subset *s*. Each image or video frame is transformed into a row vector and is called $\mathbf{I}$ so the model for the image subset $\mathbf{I}^s$ is modeled in an eigenspace Φ^s=$\{\phi_1^s\ \phi_2^s\ ...\ \phi_N^s\}$ according to the following formula (*N* is the number of frames in the video):

$$\phi_j^s = \sum_i b_{ij}^s(\mathbf{I}_i^s - \mathbf{I}_{\underline{0}}^s) \tag{1}$$

where b_{ij}^s are values from the eigenvectors of the covariance matrix $\mathbf{C}^s = \sum_{i,j}(\mathbf{I}_i^s - \mathbf{I}_{\underline{0}}^s)^T(\mathbf{I}_j^s - \mathbf{I}_{\underline{0}}^s)\}$ and $\mathbf{I}_{\underline{0}}^s$ is the mean of the subset pixel area. The subscripts *i* and *j* are horizontal and vertical pixel positions in the images.

A pseudo eigenspace is constructed with the same spatial size as the original images $\mathbf{I}$. By using the eigenvectors from the subset area this part becomes orthogonal while the rest of the pixels are modeled according to the correspondence with the subset pixels. This space Φ^p=$\{\phi_1^p\ \phi_2^p\ ...\ \phi_N^p\}$ is constructed as:

$$\phi_j^p = \sum_i b_{ij}^s(\mathbf{I}_i - \mathbf{I}_{\underline{0}}) \tag{2}$$

Coefficients $\{\alpha_j^s\}$ can be extracted from the subset eigenspace Φ^s through projection:

$$\alpha_j^s = \phi_j^s(\mathbf{I}^s - \mathbf{I}_{\underline{0}}^s)^T \tag{3}$$

These coefficients can then be used with the pseudo eigenspace for all frequencies:

$$\hat{\mathbf{I}} = \mathbf{I}_{\underline{0}} + \sum_{j=1}^{M} \alpha_j^s \phi_j^p \tag{4}$$

where M is the selected number of pseudo components used for reconstruction.

A full image can be reconstructed (Eq. 4) using the projection coefficients from only the subset pixels (Eq. 3). This means that all pixels are estimated from the subset pixels. That only a few pixels, or frequencies, are needed for encoding while another subset, or all, frequencies can be decoded is the strength of aPCA.

B. Choice of wavelet function

In our experiments, four types of wavelet families are examined; Haar, Daubechies, Symlets, and Coiflets. These orthogonal wavelets are compactly supported, meaning that they have finite-impulse response filters leading to efficient implementations. In this implementation, the four different mother wavelets are examined extensively. The results demonstrate that for any sequence of the images, the percentage of zeros is higher for Haar wavelets compared to the other wavelets and eventually more energy is retained. Haar wavelets are chosen in our implementation since the loss of information is smaller than for the others, leading to better image quality.

C. Discrete wavelet transform (DWT) in different decomposition levels

A discrete wavelet transform (DWT) divides the frequency content of an image into four parts called subbands where information from different orientations is assigned to different subbands. The four parts are frequency content from horizontal, vertical and diagonal direction plus one subband with the most important low frequency content. These bands are obtained by low-pass and high-pass filtering of the image in the horizontal and vertical directions. The subband labeled LH_1 (horizontal frequencies) is low-pass filtered in the vertical direction and high-pass filtered in the horizontal direction. LH_1 together with HL_1 (vertical frequencies) and HH_1 (diagonal frequencies) represent the wavelet coefficients for the different directions and subband LL_1 contains the most important information of the image, called approximation image. To obtain the next coarse level of wavelet coefficients subband LL_1 is further decomposed into four new subbands, LL_2, LH_2, HL_2 and LH_2. The decomposition is shown in figure 1. As the decomposition continues the resolution is lowered so LH_2 has a higher resolution than LH_3.

D. Correlation models for the wavelet domain

aPCA works by estimating missing or new information based on its correlation with existing information. If there is no correlation between the information parts it will not work. A wavelet transform alters the original data into uncorrelated frequency components and it should not be possible to find any correlation between the different subbands. Although two coefficients might be close to each other in the spatial domain they can be located on distantly separated subbands of the wavelet transform. Consequently a standard wavelet quad-tree, modeling only parent-child relationships, poorly represents the existing correlations between the coefficients in different subbands. By looking at prior models on Gaussian Markov random field (GMRF) covariance structure it is possible to see the actual correlation between the frequencies. This issue is studied by Azimifar et.al. [11]. Their results clearly show that the wavelet coefficients in different subbands are correlated

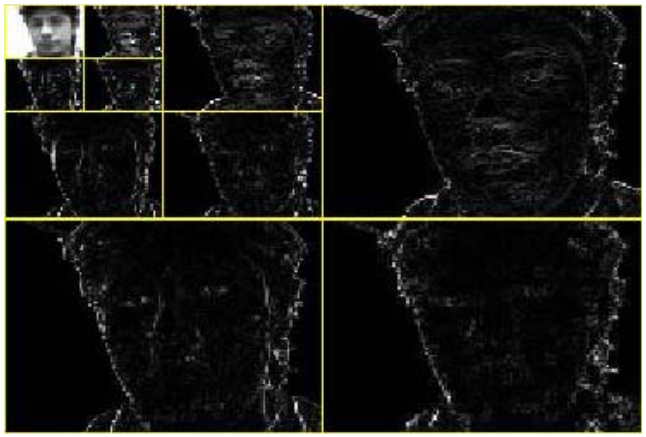

Fig. 1. Wavelet decomposition of a facial image (third level decomposition). Top left corner: LL3, bottom right corner: HH1.

with each other. A coefficient will absorb correlation from other coefficients in the neighborhood, both within and across scales, thus allowing the use of aPCA estimation

E. aPCA-DWT

The combination of aPCA and DWT is accomplished by regarding the DWT subbands as images. Even if the subbands represent frequencies, they can be regarded as image parts with pixel values (Figure 1). The process can be divided into two separate steps; first, a video frame is transformed with DWT and secondly the DWT version is encoded with aPCA. The inverse procedure means that a DWT image with more frequencies than the original image is reconstructed with aPCA and this image is inversely transformed (IWT) to become a reconstructed image.

After DWT a certain subband is regarded as an area in an image and it is simple to apply aPCA to this image. Using figure 1 as example, the upper left corner is the LL_3 subband. This subband is regarded as subset s and used to create the eigenspace Φ^s. The pseudo eigenspace Φ^p is constructed from all subbands in figure 1. Φ^s is used to extract coefficients $\overline{\alpha}^s$ and these coefficients are used with Φ^p to estimate all the subbands based on the information from the LL_3 subband.

For this technique to work we need knowledge of the objects in the scene. In this paper we use facial images and the model will not work well for other types of objects. The pseudo eigenspace Φ^p has known relationships with the subset eigenspace Φ^s. If the content in the subset or pseudo eigenspace is altered so is the model.

III. Experiments

Several different experiments have been performed with aPCA-DWT where all of them use a certain subband as subset s for encoding and all subbands for decoding. All experiments show how it is possible to estimate frequencies in a higher resolution than the original image. There is also one experiment directly showing how the technique can be used to zoom and one experiment directly showing how the technique can be used for deblurring. Each experiment is described in

TABLE I. ENCODING WITH LL_3 SUBBAND FROM COLOR IMAGES. AVERAGE FOR ALL VIDEO SEQUENCES.

M	Reconst. qual. [PSNR]
5	36,5
10	37,7
15	38,3
20	38,8
25	39,3

TABLE II. ENCODING WITH LL_3 SUBBAND FROM COLOR IMAGES. AVERAGE FOR ALL VIDEO SEQUENCES.

Method	Reconst. qual. [PSNR]
Bicubic interpolation	32,9
Sparse representation	35,1
aPCA-DWT (M=25)	40,2

detail along with the results in separate subsections. First we provide a description of the data used for the experiments.

A. Video sequences

We have used six video sequences; each one showing a person where he/she is displaying the six basic emotions proposed by Ekman [12]. The video sequences are approximately 30 seconds long and a new expression is displayed for three seconds every five seconds. Between expressing emotions the person in the video returns to a neutral expression. The framerate is 15 fps and the resolution is 240x176 for each video sequence. When we have used LL_3 subband as I^s it has a size of 22x30 pixels and for LL_4 the size is 11x15.

B. Encoding with LL_3 with color images

The three color channels; Red, Green and Blue are encoded individually so there are actually three different subset eigenspaces Φ^s, pseudo eigenspaces Φ^p and coefficient sets $\overline{\alpha}^s$. The eigenspaces are created individually for each video sequence. We have used the same number of coefficients M for all color channels. This means that when the number of M is 10, R, G and B use 10 dimensions each and the actual amount of dimensions used is 30. The reconstruction quality is presented as an average for the three color channels over all video sequences in table I. In this experiment we have used LL_3 subband for encoding and all available subbands for decoding.

We have compared our results with the results from Yang et.al [2]. The input images are subsampled to a quarter of the size in both horizontal and vertical direction, meaning they are 16 times smaller than the original images. The subsampled images are used to estimate a high frequency version of them. Their method uses image patches for training and we have trained their model with general image patches. The results are shown for 2 frames in figure 2 and the average objective quality (PSNR) for all video sequences are shown in table II. When we use our images for training we have used 250 images from a video sequence for training and one for testing.

Fig. 2. Example frames reconstructed with different methods. (left)=Bicubic interpolation reconstruction (middle)=Images reconstructed with sparse representation (right)=Images reconstructed with aPCA-DWT.

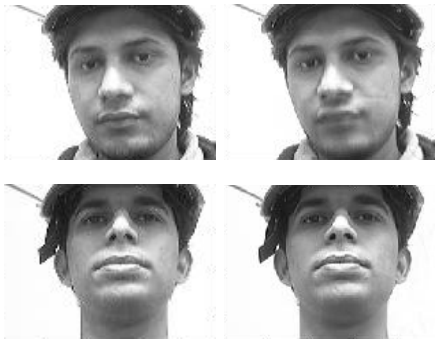

Fig. 3. Encoding with LL_3 from color images. (left)=Original images (right)=Reconstructed images.

TABLE III. ENCODING WITH LL_3 SUBBAND OF SEVERAL COLOR IMAGE VIDEOS. AVERAGE FOR ALL VIDEO SEQUENCES.

M	Reconst. qual. [PSNR]
5	34,5
10	35,0
15	35,8
20	36,7
25	37,4

C. Encoding with LL_3 on frames from different video sequences with color images

In the previous experiment we used frames from one video sequence where the difference between frames is the facial mimics. To test if the technique works well with more diverse video content we use frames from five different video sequences. The frames from the five video sequences are treated as frames in one long video sequence and the encoding procedure is exactly the same as in the previous experiments. Original and reconstructed frames from two video sequences are shown in figure 3 and the results measured in PSNR are shown in table III. We have used the same setup as in the previous experiment (LL_3 subband for encoding, all subbands for decoding) so they are directly comparable.

(a) (b)

Fig. 4. Blurred and reconstructed version of a grayscale video frame. (a)=Image blurred with a blur disk size of 8 (b)=Reconstructed image from blurry encoding with disk size 8.

TABLE IV. ENCODING WITH BLURRY GRAYSCALE IMAGES. AVERAGE FOR ALL VIDEO SEQUENCES.

	Reconst. qual. [PSNR]	
M	**Disk size 4**	**Disk size 8**
5	38,0	38,0
10	39,2	39,1
15	39,9	39,8
20	40,5	40,4
25	41,0	40,8

TABLE V. DEBLURRING QUALITY. AVERAGE FOR ALL VIDEO SEQUENCES.

Method	**Reconst. qual. [PSNR]**
Blind image deconcolution	34,5
aPCA-DWT (M=25)	41,1

D. Encoding with frequencies from blurred images with grayscale images

One major difference between sharp and blurry images is that sharp images contain much more frequencies than blurry ones. A low-pass filtered image will look blurry and the only difference from the sharp version of it is that some frequencies have been removed. By using the same eigenspace (from LL_3 and pseudo eigenspace from all subbands) as in the first experiment we have correspondence between blurry and sharp images. In addition, we create blurry images by low-pass filtering them with a Gaussian blur kernel. This blurry image is used for encoding on the LL_3 eigenspace and a sharp image can be decoded with the pseudo eigenspace. We use blurring with a Gaussian disk with size 4 and 8, where the image with a larger blur disk will be blurrier. A blurred image using a disk size of 8 and the decoded (deblurred) image are shown in figure 4 and the result is noted in table IV. In this experiment we use grayscale images and the result is not directly comparable to the previous experiments. We have compared our result with blind image deconvolution deblurring, presented by Kundur and Hatzinkos [13]. Example frames are shown in figure 5 and the overall average quality is shown in table V.

E. Encoding with LL_3 of subsampled video with grayscale images

We have performed one experiment where we use small size images for encoding compared to the reconstructed images. When a subband of a higher order than all available subbands is used for encoding it actually corresponds to using a smaller image so all experiments in this article in fact show the capability of creating SR images with aPCA-DWT. This experiment clearly shows how this method can be used to zoom beyond the original resolution of an image. Again, we use the video sequences and scale down the resolution four times through Photoshop, yielding a video sequence with a resolution four times lower than before. The use of DWT in this down-scaled image is similar to using the LL_2 subband which is four times smaller than the original image. We use the LL_3 subband of the smaller image for encoding, equivalent to using LL_4 for the original image. The frequencies in all other subbands are estimated based on their relationship to the frequencies in LL_3 of the sub-sampled image and a DWT image with higher frequency content is reconstructed. Example frames from two video sequences are shown in figure 6. The quality measured in PSNR when M=25 for the video sequences and the quality yielded by images scaled with interpolation are shown in table VI. We use grayscale images in this experiment so it can be compared to the previous experiment but not the first two experiments.

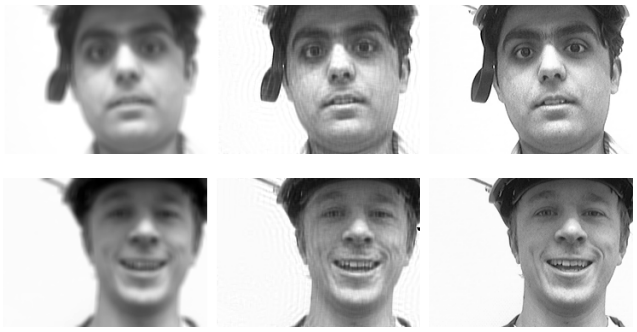

Fig. 5. Image deblurring comparison. (left)=Blurred images (middle)=deblurred with blind image deconvolution (right)=deblurred with aPCA-DWT.

TABLE VI. ZOOMING (ENLARGING) (LEFT)=INTERPOLATION RECONSTRUCTION (RIGHT)=APC-DWT RECONSTRUCTION (M=25). AVERAGE FOR ALL VIDEO SEQUENCES.

Interpol. quality [dB]	**Reconst. qual. [PSNR]**
34,7	41,0

IV. DISCUSSION AND CONCLUSIONS

It can be seen from the experiments in section III-B and III-C that the quality is reduced when the video content is diverse and based on the facial expressions by several different persons but as can be seen from the resulting images, (Figure 3), the image quality is still very high.

Zooming beyond the existing resolution essentially means creating a version of the image with a higher resolution than the original image. So the zooming experiment and the creation of SR image experiments basically show the same capability. The zooming experiment is included to highlight the techniques capacity for such a task.

A possible improvement is to divide the database into

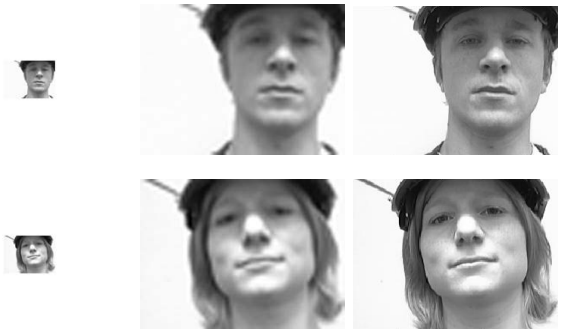

Fig. 6. Zooming 4 times in both direction (16 times zoom). (left)=Original images (middle)=zoomed with interpolation (right)=zoomed with aPCA-DWT.

gender and ethnicity. At present, we have limited our material to six video sequences and we use them all in a single database to create a reasonably large data set. The results will be much improved if the database is divided based on gender and/or ethnicity; i.e. different databases for males, females and different ethnic backgrounds. Our videos consist of images of males and females, Caucasians and Asians. Previous research by Kirby and Sirovich has shown how modeling of faces through PCA works well if the database is divided into such categories [14].

The technique can easily be used to remove noise from an image. A simple way for noise removal is to low-pass filter the image but this also means that the sharp edges in the image are blurred. A low frequency version of an image, without noise, can be used to estimate a high frequency version of the image meaning that it is possible to remove noise with low-pass filtering and estimate a high frequency image without noise.

The PSNR values in experiment III-D and III-E are calculated based on grayscale images while experiment III-B and III-C use color images. We have found that the result is poor for the blue color channel but the result is still shown as an average for the three color channels. Should they be presented individually, it could be seen that the red and green color channels produce higher quality results than the blue channel.

It is not really possible to see the improvement of aPCA-DWT from figure 2 but the objective quality in table II clearly states that our method outperforms the results that Yang et.al have. Their method is however more general, works with several different kinds of images and is not data dependent. Our method only works for facial images and only for images which are similar in size and shape. This is something that needs to be addressed in future work.

The most important future work for this research is to create a more general model than the one used in this article. The present model can be used to estimate higher frequencies based on specific frequency content. This means that the model can only estimate higher frequencies for certain objects; in this work we have used head-and-shoulders images. With a general model, higher frequencies can be created based on shapes, edges or any kind of objects in the image. This will probably require that the image is divided into subparts, a solution we are currently working on. Another possible solution is to have several models for known objects. If the video shows a tree, one model is used, if it shows a car, a second model is used, and so on.

The last experiment clearly shows how accurately the technique works when a correct model exists. With accurate models for specific kinds of images it will be possible to create SR versions of such images. With a carefully designed general model the method could be used for any kind of image, opening up a vast array of possibilities.

REFERENCES

[1] J. Yang, J. Wright, T.S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *Image Processing, IEEE Transactions on*, vol. 19, no. 11, pp. 2861–2873, 2010.

[2] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, 2008, pp. 1–8.

[3] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *Computer Vision, 2009 IEEE 12th International Conference on*, 2009, pp. 349–356.

[4] K. I. Kim and Y. Kwon, "Single-image super-resolution using sparse regression and natural image prior," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 6, pp. 1127–1133, 2010.

[5] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, 2004, vol. 1, pp. I–I.

[6] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *Computer Graphics and Applications, IEEE*, vol. 22, no. 2, pp. 56–65, 2002.

[7] I. Jolliffe, *Principal Component Analysis*, Springer-Verlag, New York, 1986.

[8] U. Söderström and H. Li, "Asymmetrical principal component analysis for video coding," *Electronics letters*, vol. 44, no. 4, pp. 276–277, February 2008.

[9] U. Söderström and H. Li, "Side view driven facial video coding," in *Advanced Video Coding for Next-Generation Multimedia Services*, Prof. Yo-Sung Ho, Ed. InTech, 2013, DOI: 10.5772/53101.

[10] U. Söderström and H. Li, "Asymmetrical principal component analysis theory and its applications to facial video coding," in *Effective Video Coding for Multimedia Applications*, Sudhakar Radhakrishnan, Ed., pp. 95–110. InTech, 2011.

[11] Z. Azimifar, P. Fieguth, and E. Jernigan, "Towards random field modeling of wavelet statistics," in *Image Processing. 2002. Proceedings. 2002 International Conference on*, 2002, vol. 1, pp. I–361 – I–364 vol.1.

[12] P. Ekman, *Emotion in the Human Face*, Cambridge University Press, New York, 1982.

[13] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *Signal Processing Magazine, IEEE*, vol. 13, no. 3, pp. 43–64, May 1996.

[14] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *Journal of the Optical Society of America*, vol. 4, pp. 519–524, 1987.