INTERPOLATION SPECIFIC RESOLUTION SYNTHESIS

Ramez Yoakeim and David Taubman

The University of New South Wales, Sydney, Australia

ABSTRACT

We introduce a new approach to Resolution Synthesis which is specifically matched to the interpolation problem. This is achieved by explicitly aligning classification and subsequent interpolation activities with image models conducive to interpolation. Our image models are pre-determined rather than discovered and represent a range of edges of arbitrary orientation, profile and relative position. We demonstrate superior interpolation outcomes compared to statistical classification based resolution synthesis.

Index Terms— Image processing, Interpolation, Resolution Synthesis

1. INTRODUCTION

Resolution Synthesis (RS) describes a class of algorithms that aim to estimate a higher resolution image, given a single low resolution instance of the same image. Resolution Synthesis fits within the broader class of problems known as "Inverse Problems in Imaging," which includes related ill-posed inverse problems such as super-resolution, deblurring, denoising and other restoration objectives. While various approaches to resolution synthesis have been proposed [1, 2], in this paper we focus on the approach introduced by Atkins et al. [3, 4]. Previous work [5] dealt with examining various aspects of the fitness of the RS model in [3, 4] for the purpose of image interpolation. In this work we consider the underlying statistical model, and examine an alternative better aligned with the interpolation problem resulting in improved interpolation performance.

We start with a brief overview of the RS algorithm as proposed by Atkins et al, then we examine its suitability for the image interpolation problem. In earlier work [5] we noted a number of considerations in the algorithm design. At a high level, the original RS approach consists of two key elements: a classification strategy for the low resolution image pixels; and a high resolution synthesis (interpolation) strategy, which is based on the classification. In this approach the neighborhood surrounding each low resolution image pixel is viewed as a realization of one of M Gaussian generators, each corresponding to a hidden class model which represents some image feature. There exists no deterministic relationship between any particular class, j, and the specific low resolution neighbourhood. Each class is characterised by its probability distribution, a prior likelihood and a Linear Minimum Mean Squared Error estimator (LMMSE), all determined during training. It is then possible to generate the high resolution pixels corresponding to the observed pixel values by deducing a posterior class membership probability distribution across all classes and taking the expectation of the output of the LMMSE estimators over the posterior class distribution. To determine the parameters describing each of the class models, a training set of low resolution images are used to together with the EM algorithm. Given a particular training set, the EM algorithm is used to determine the parameters of the class models that best explain the features of the images. At the conclusion of the classification process, classical estimation theory is used to derive the LMMSE interpolators.

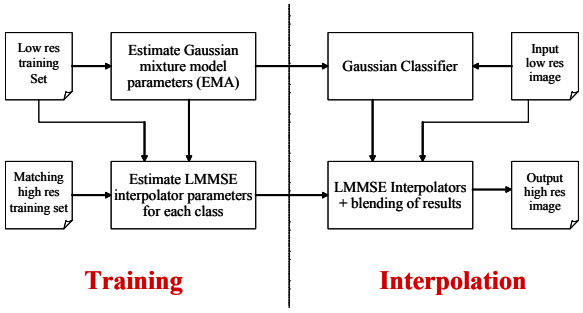

Fig. 1. Structure of classification based RS

Firstly, we note that the Gaussian class models as proposed are not particularly effective in discriminating between semantically recognizable image features. The use of a common scalar variance in the classification process effectively reduces the characterisation of the image source models to sole dependence on the mean vector of each class. In view of this, we propose the use of multivariate Gaussian class models which can fully exploit the covariance structure found within each low resolution pixel neighborhood.

Furthermore, we note the lack of direct link between the classification and interpolation aspects of the statistical modelling approach to RS. Our proposed alternative approach more closely integrates the interpolation objective into the classification activity so that the parameters of each class, indeed the classes themselves are chosen a priori to represent various aspects common to natural images which should aid in the interpolation process. Specifically, we design these classes to correspond to edges of varying orientation, profile and relative postion within the region of interest.

The paper is organised as follows. Section 2 provides a brief overview of the salient aspects of classification based RS algorithm relevant to this work. We follow with an analysis of the low resolution image modelling aspects of RS and present our alternative statistical modelling approach in Section 4. Section 5 presents the details of our novel interpolation specific approach to RS, including its performance results.

2. SCALAR RESOLUTION SYNTHESIS

Classification based resolution synthesis breaks the familiar image interpolation exercise into a two stage process illustrated in Figure (1). The first stage starts with the image being broken into individual non-overlapping regions centred around pixels of interest. This is followed by an arbitrary classification of those image regions in-

tended to cluster regions of similar characteristics together and derive the parameters of each such cluster. Those parameters are then used in the second stage where LMMSE interpolators optimised for each set of cluster parameters are used to operate on the same low resolution image regions employed by the classification process, producing a corresponding higher resolution approximation.

We will restrict ourselves to a brief overview of the salient aspects; for a more comprehensive review please refer to our prior work [5]. Formally, we write $x[\mathbf{n}] \equiv x[n_1, n_2]$ for the high resolution image which gives rise to a corresponding low resolution image $u[\mathbf{n}]$. The low resolution version of the image is obtained using a conventional filtering and subsampling process, with some reductive kernel $h[\mathbf{k}]$

$$u[\mathbf{n}] = \sum_{\mathbf{k}} h[\mathbf{k}]\, x[2\mathbf{n} - \mathbf{k}]$$

Associated with each low resolution image pixel $u[\mathbf{n}]$, we identify a neighbourhood $\mathcal{N}_{\mathbf{n}}$ and the vector of samples $\mathbf{z}[\mathbf{n}]$, which belong to $\mathcal{N}_{\mathbf{n}}$. From $\mathbf{z}[\mathbf{n}]$, [3, 4] derives a feature vector

$$\mathbf{y}[\mathbf{n}] = \begin{cases} \mathbf{v}[\mathbf{n}] \cdot \|\mathbf{v}[\mathbf{n}]\|^{-3/4} & \mathbf{v}[\mathbf{n}] \neq \mathbf{0} \\ \mathbf{0} & \text{otherwise} \end{cases} \quad (1)$$

where $\mathbf{v}[\mathbf{n}]$ is obtained by subtracting $u[\mathbf{n}]$ from its 8 immediate neighbours in $\mathcal{N}_{\mathbf{n}}$.

The mean-removed, non-linearly transformed neighbourhood, $\mathbf{Y}[\mathbf{n}]$, is modeled as a multivariate Gaussian mixture, with PDF

$$p_{\mathbf{Y}}(\mathbf{y}) = \sum_{j=1}^{M} \pi_j p_{\mathbf{Y}|J}(\mathbf{y}, j) \quad (2)$$

Here, j denotes one of M underlying classes, π_j is the prior probability that $J = j$, and

$$p_{\mathbf{Y}|J}(\mathbf{y}, j) = \frac{1}{\left(\sqrt{2\pi}\right)^8} e^{-\frac{1}{2\sigma^2}\|\mathbf{y}-\boldsymbol{\mu}_j\|^2} \quad (3)$$

Note that each element of $\mathbf{Y}[\mathbf{n}]$ is being modeled here as an independent Gaussian random variable, with a unique class-dependent mean, and a common class-independent variance, σ^2.

Given an observed low resolution image $u[\mathbf{n}]$, we compute $\mathbf{y}[\mathbf{n}]$ for each $\mathbf{n}$, and evaluate the posterior likelihood that the low resolution neighbourhood vector $\mathbf{z}[\mathbf{n}]$ was generated by class j, for each $j = 1, 2, \ldots M$. Applying Bayes rule, we find that

$$p_{J|\mathbf{Z}}(j, \mathbf{z}) = p_{J|\mathbf{Y}}(j, \mathbf{y}) = \frac{\pi_j \exp\left(\frac{-1}{2\sigma^2}\left\|\mathbf{y} - \boldsymbol{\mu}_j\right\|^2\right)}{\sum_{l=1}^{M} \pi_l \exp\left(\frac{-1}{2\sigma^2}\left\|\mathbf{y} - \boldsymbol{\mu}_l\right\|^2\right)} \quad (4)$$

where the first equality holds by assumption. The parameters $\boldsymbol{\mu}_j$, π_j and σ are derived by a training process, which aims to maximize

$$\sum_i \log p_{\mathbf{Y}}\left(\mathbf{y}_i | \boldsymbol{\mu}_j, \pi_j, \sigma\right), \quad (5)$$

where the $\mathbf{y}_i$ are drawn from the mean-removed transformed neighbourhood vectors $\mathbf{z}[\mathbf{n}]$ of a collection of low resolution training images, $u[\mathbf{n}]$. The "expectation maximization" (EM) algorithm is used for this purpose. Focusing only on the update equation for the global variance parameter σ then

$$\sigma^{2(k+1)} = \frac{1}{8} \sum_{l=1}^{M} \left[\frac{\pi_j^{(k+1)}}{N_j^{(k+1)}} \sum_{i=1}^{n} \left\| \mathbf{y}_i - \boldsymbol{\mu}_j^{(k+1)} \right\|^2 \cdot p_{J|\mathbf{Y}}\left(j, \mathbf{y}_i; \theta^{(k)}\right) \right] \quad (6)$$

where $N_j^{(k+1)}$, $\boldsymbol{\mu}_j^{(k+1)}$ and $\pi_j^{(k+1)}$ are the update equations for the other parameters.

For synthesis, we first write $\mathbf{x}[\mathbf{n}]$ for the high resolution image block we seek to synthesise given a unique low resolution neighbourhood vector $\mathbf{z}[\mathbf{n}]$. We model the conditional distribution of $\mathbf{X}$ given $\mathbf{Z} = \mathbf{z}$, for each class j, using a multivariate Gaussian, with mean $\mathbf{A}_j\mathbf{z} + \boldsymbol{\beta}_j$. This leads to the following formula for the expected value, $\hat{\mathbf{x}}[\mathbf{n}]$, of $\mathbf{X}[\mathbf{n}]$ given $\mathbf{z}[\mathbf{n}]$

$$\hat{\mathbf{x}}[\mathbf{n}] = \sum_{j=1}^{M} p_{J|\mathbf{Y}}(j, \mathbf{y}[\mathbf{n}]) \cdot \left(\mathbf{A}_j\mathbf{z} + \boldsymbol{\beta}_j\right) \quad (7)$$

Noting that the mode and mean of a Gaussian distribution are equal; we point out that the MAP estimate, $\hat{\mathbf{x}}'[\mathbf{n}]$, is

$$\hat{\mathbf{x}}'[\mathbf{n}] = \left.\mathbf{A}_{j_{\mathbf{n}}}\mathbf{z} + \boldsymbol{\beta}_{j_{\mathbf{n}}}\right|_{j_{\mathbf{n}} = \operatorname{argmax}_j p_{J|\mathbf{Y}}(j, \mathbf{y}[\mathbf{n}])} \quad (8)$$

The parameters, $\mathbf{A}_j$ and $\boldsymbol{\beta}_j$ are derived from a collection of low and high resolution training images, using classical estimation theory.

3. FEATURE CHARACTERISATION AND CLASSIFICATION MODEL

The purpose of the non-linear transformation in (1) is to distort the space of the feature vector used for clustering, thereby placing variable emphasis on different image features to obtain a representative mixture of classes favourable to the interpolation process. During the synthesis stage, the feature vector is used to characterise the low resolution neighbourhood and proportionally identify the mixture model to use for interpolating it. As such, the effectiveness of this feature characterisation process is critical to both the training and subsquent interpolation activities.

We note that the non-linear transformation is not insensitive to variations in luminance. While the transformation promotes closer clustering of image samples representing similar edge orientation, the normalisation is insufficient to collapse samples representing the same edge with various luminance into a single cluster. As demonstrated in [5], during training this results in the creation of redundant classes, which model essentially the same feature with different luminances. As a result, individual locations can have a high probably of membership in several classes, even when the classes represent important features such as oriented image edges. Futhermore, this poor semantic association is a direct result of the classification scheme's reliance on class mean patterns $\boldsymbol{\mu}_j$, as the primary classification differentiator. This builds an intensity dependence into the classes which works against semantic association and reveals the inherent weakness in the interpolation-independent approach to classification when used for interpolation purposes. We note the following:

1. The classification objective of equation (5) has no dependence whatsoever on the high resolution image, and hence the synthesis problem itself thus providing no guarantee of optimal utility in the interpolation problem.
2. The choice of a single scalar variance parameter, σ, with independent Gaussians for each element in the mean-removed transformed neighbourhood vector $\mathbf{Y}$, reduces the classification procedure to a pattern matching exercise. The patterns are represented by the class centroids $\boldsymbol{\mu}_j$, while σ controls the degree to which we prefer the best matching pattern over other, more distant matches.

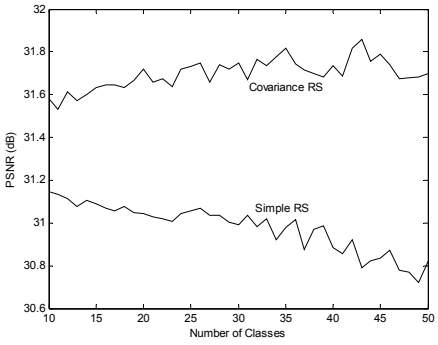

Fig. 2. Interpolation PSNR vs number of classes for scalar and multivariate RS

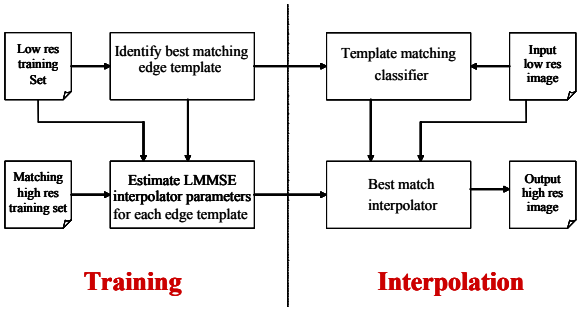

Fig. 3. Interpolation specific RS overview

For the purpose of focusing on the dynamics of the classifcation approach we opt to train on a randomly selected subset of neighbourhoods from the same image used for testing. This also allows us to highlight the need for better modeling of the image features. We note that despite this restriction, the ability of scalar RS to adequately generalise within the same image is limited suggesting that the classification strategy is having a hard time discovering semantically meaningful image statistics. Indeed, considering the illumination dependence of the class models, any small deviation between the statistics of the image and those represented by the training vectors can result in the selection of inappropriate LMMSE interpolators.

4. MULTIVARIATE RESOLUTION SYNTHESIS

Consider the case when the mean-removed, non-linearly transformed neighbourhood vector, $\mathbf{Y}\,[\mathbf{n}]$, obtained from the transformation of (1) is modeled as a jointly Gaussian random vector, with a unique class-dependent mean and a full covariance matrix. These parameters must be estimated by the EM algorithm.

More specifically, we can now rewrite equation (3) for class j with mean μ_j as before but here with a symmetric, positive semi-definite covariance matrix Σ_j

$$p_{\mathbf{Y}|J}\left(\mathbf{y},j\right)=\frac{1}{\left(\sqrt{2\pi}\right)^{8}\left|\Sigma_j\right|^{1/2}}\exp\left(-\frac{1}{2}\left(y-\mu_j\right)^{T}\Sigma_j^{-1}\left(y-\mu_j\right)\right) \tag{9}$$

This leads to the following class probability membership in the multivariate case

$$p_{J|\mathbf{Y}}\left(j,\mathbf{y}_i\right)=\frac{\pi_j\left|\Sigma_j\right|^{-1/2}\exp\left(D\left(y;\boldsymbol{\mu}_j;\Sigma_j\right)\right)}{\sum_{l=1}^{M}\pi_l\left|\Sigma_l\right|^{-1/2}\exp\left(D\left(y;\boldsymbol{\mu}_l;\Sigma_l\right)\right)} \tag{10}$$

where Σ_j is the covariance of the class dependent distribution and D is the weighted Mahalanobis distance function

$$D\left(y;\boldsymbol{\mu}_j;\Sigma_j\right)=-\frac{1}{2}\left(y-\mu_j\right)^{T}\Sigma_j^{-1}\left(y-\mu_j\right) \tag{11}$$

Proceeding with the derivation we can show that the iterative update equations of the other classification parameters remain unchanged. The 8 dimensional covariance matrix, Σ_j , is updated with each iteration of the EM algorithm as was the case with the scalar variance σ in equation (6).

$$\Sigma_j^{(k+1)}=\frac{1}{N_j^{(k+1)}}\sum_{i=1}^{n}\left(y-\mu_j\right)\left(y-\mu_j\right)^{T}\cdot p_{J|\mathbf{Y}}\left(j,\mathbf{y}_i;\theta^{(k)}\right) \tag{12}$$

The classification modelling objective used in the EMA (5) is now given by

$$\sum_i \log p_{\mathbf{Y}}\left(\mathbf{y}_i|\boldsymbol{\mu}_j,\pi_j,\Sigma_j\right) \tag{13}$$

Figure (2) provides some experimental results comparing RS with a common scalar variance against RS with class specific covariance matrices. It is worth noting that the full covariance RS algorithm significantly increases the cost of training. However, the computational impact on the actual interpolation process is relatively small, since this is dominated by the cost of applying the LMMSE interpolation operators. Furthermore, as we noted in section (3), we can observe a progressive reduction of interpolation quality as the number of classes increases for scalar RS. This arises out of the poor semantic association with specific image features; namely edges; often associated with better quality interpolated images. Multivariate RS delivers improved classification and consequently an increase in the number of classes generally yields a corresponding improvement in interpolated image quality.

5. INTERPOLATION SPECIFIC RESOLUTION SYNTHESIS

Regions of natural images may be broadly classified into three categories: smooth regions, textured regions and edges. We focus on edge regions because they provide both high frequency content to be synthesized and sufficient structure to have some hope of unwrapping the aliasing effects from the low resolution imagery alone. The interpolation-specific RS approach introduced in this section differs from the statistical classification based RS briefly outlined earlier in a few notable ways. The most significant is that in classification based RS we statistically estimate a set of classes broadly representing the statistical properties of images typical of the training set through an iterative process optimising a statistical likelihood function. With interpolation specific RS we create the classes deterministically so as to represent various edge profiles. Since the classes are determined a priori, the training process is now limited to the derivation of the LMMSE interpolators. The overall system is represented in Figure 3.

As an alternative to the feature characterisation approach outlined earlier, we use reference templates or masks which are then

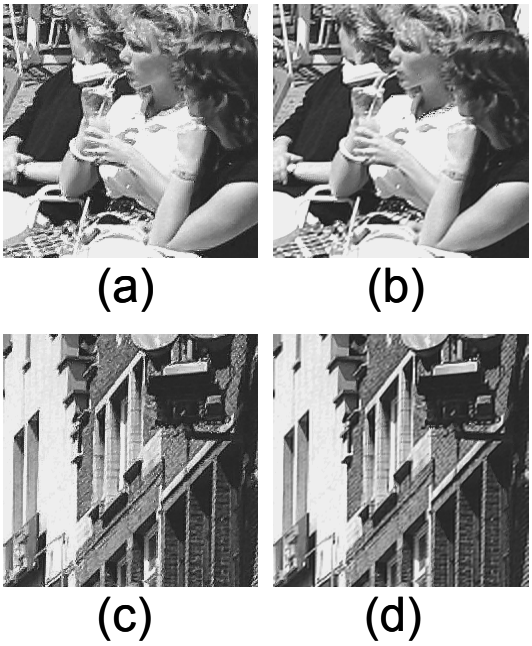

Fig. 4. Relative performance of Scalar classification based RS (a,c) and Interpolation specific RS (b,d).

compared statistically to each image sample to determine an extent of similarity. More specifically, to characterise an edge we choose to identify its orientation, its relative position within $\mathcal{N}_{\mathbf{n}}$ and its profile. We could additionally assign a binary parameter for the directionality of the edge, denoting its transition from light to dark or vice versa. As we shall see, however, this turns out not to be necessary.

Within $\mathcal{N}_{\mathbf{n}}$, we identify possible edges through their orientation θ, distance τ from $\mathbf{n}$, which is the centre of the neighbourhood, and the profile of the edge, which we model by a single Gaussian blur parameter σ. In this way, our representative edges are described completely by the triplet (θ, τ, σ).

To identify an image feature, the image samples are compared with a set of normalised edge templates representing a range of edge orientations, positions and profiles. Each candidate edge template, $\mathbf{t}_{\theta,\tau,\sigma}$, is a vector of edge samples defined over the same neighbourhood $\mathcal{N}_{\mathbf{0}}$ used for the low resolution image vectors $\mathbf{z}[\mathbf{n}]$. Specifically, for each $\mathbf{p} \in \mathcal{N}_{\mathbf{0}}$, the value of $\mathbf{t}_{\theta,\tau,\sigma}$ at that location is given by

$$t_{\theta,\tau,\sigma}[\mathbf{p}] = \frac{1}{2} + \frac{1}{\pi}\int_0^{\frac{\tau + p_1\sin(\theta) - p_2\cos(\theta)}{\sqrt{2}\sigma}} e^{-t^2} dt \qquad (14)$$

As can be seen from equation (14) the normalisation of the edge template renders it invariant to luminance variations, so that an edge of a specific characteristic orientation, position and profile can be identified without reference to the luminance transition it undertakes – light to dark or vice-versa. The extent of match between an image feature and a template is measured using the Normalised Cross Correlation γ of the template $\mathbf{t}_{\theta,\tau,\sigma}$ and $\mathbf{z}[\mathbf{n}]$ given by:

$$\gamma = \frac{\langle (\mathbf{z}[\mathbf{n}] - \bar{\mathbf{z}}[\mathbf{n}]), (\mathbf{t}_{\theta,\tau,\sigma} - \bar{\mathbf{t}}_{\theta,\tau,\sigma}) \rangle}{\sqrt{\|\mathbf{z}[\mathbf{n}] - \bar{\mathbf{z}}[\mathbf{n}]\|^2 \cdot \|\mathbf{t}_{\theta,\tau,\sigma} - \bar{\mathbf{t}}_{\theta,\tau,\sigma}\|^2}} \qquad (15)$$

Table 1. Interpolation PSNR results for classical methods, scalar and multivariate classification RS and interpolation specific RS.

Method	Bilinear	Bicubic	SRS	MRS	IRS
Cafe	22.4 dB	23.0 dB	19.4 dB	19.6 dB	25.0 dB
Facade	21.4 dB	21.8 dB	21.3 dB	21.6 dB	23.5 dB

where $\bar{\mathbf{z}}[\mathbf{n}]$ is the mean of the image sample and $\bar{\mathbf{t}}_{\theta,\tau,\sigma}$ is the mean of the template.

For interpolation, we adopt the same definitions of $\mathbf{x}[\mathbf{n}]$ and $\mathbf{z}[\mathbf{n}]$ as noted above. But here we can compute the estimated value $\mathbf{x}^e[\mathbf{n}]$, of $\mathbf{X}[\mathbf{n}]$ given $\mathbf{z}[\mathbf{n}]$ directly using

$$\mathbf{x}^e[\mathbf{n}] = \mathbf{A}_{j_\mathbf{e}}\mathbf{z}|_{j_\mathbf{e}=\text{argmax}_j\,\gamma(t,\mathbf{z}[\mathbf{n}])} \qquad (16)$$

The parameter, $\mathbf{A}_j$ is derived from a collection of low and high resolution training images, using classical estimation theory. Note that unlike equation (8) there is no mean term in equation (16).

To test the relative performance of interpolation specific resolution synthesis compared to classification based resolution synthesis we selected two different type of images. The first is very rich in edges, while the second one includes more smooth regions and human subjects. Both resolution synthesis interpolators were trained using the same set of images which did not include the two test images.

As Figure (4) illustrates the improved performance of interpolation specific RS compared to scalar classification based RS. Numerical results are provided in Table 1 including a comparison with classical interpolation methods. As anticipated the focus on the modelling of edge regions in the construction of classes results in improved overall interpolation performance.

6. CONCLUSIONS

In this paper we presented an analysis of the underlying model of resolution synthesis used for image interpolation. We proposed a novel approach to RS which is better aligned with the interpolation objective. By constructing an edge-centric model for the classification and interpolation of images we were able to demonstrate qualitative and quantitative improvements in interpolation outcomes.

7. REFERENCES

[1] S.-H. G. Chang, Z. Cvetković, and M. Vetterli, "Resolution enhancement of images using wavelet transform extrema extrapolation," *IEEE Trans. Acoust. Speech and Sig. Proc.*, vol. 4, pp. 2379–2382, May 1995.

[2] Z. Ye, Q. Guohui, and L. Shaobin, "Wavelet transform and multi-resolution synthesis of texture images," *Proc. IEEE Region 10 Conf. Computer, Communication, Control and Power Engineering*, pp. 442–445, 1993.

[3] C. Atkins, C. Bouman, and J. Allebach, "Optimal image scaling using pixel classification," *Proc. IEEE Int. Conf. Image Proc.*, vol. 3, pp. 864–867, Sep 2001.

[4] C. Atkins, "Classification-based methods in optimal image interpolation," *Ph.D. thesis, Purdue*, 1998.

[5] R. Yoakeim and D. Taubman, "Quantitative analysis of resolution synthesis," *Proc. IEEE Int. Conf. Image Proc.*, vol. 3, pp. 864–867, Oct 2004.