NOVEL AUTOREGRESSIVE MODEL BASED ON ADAPTIVE WINDOW-EXTENSION AND PATCH-GEODESIC DISTANCE FOR IMAGE INTERPOLATION

*Wenhan Yang, Jiaying Liu, Shuai Yang, Zongming Guo**

Institute of Computer Science and Technology, Peking University, Beijing, P.R. China, 100871

ABSTRACT

In this paper, we propose a novel autoregressive (AR) model based on the adaptive window and the patch-geodesic distance for the image interpolation. The model combines the information of inner/inter-patch correlation. To model the inner-patch correlation, we introduce a patch-geodesic distance similarity metric. The proposed metric shows the desirable capacity to depict the piecewise-stationarity of natural images. For the inter-patch correlation, we introduce the inter-patch structure variation and propose an adaptive window-extension AR model. The model extends the interpolation window according to the local structural variation, increasing the adaptation without violating the consistency. Comprehensive experiments demonstrate that the proposed method is better than or competitive with state-of-the-art interpolation methods in both objective and subjective quality evaluations.

Index Terms— Structural variation, interpolation, autoregressive model, pixel similarity

1. INTRODUCTION

Interpolation is a general and economical technique for the image enlargement. It generates the high resolution (HR) images by utilizing the information of the low resolution (LR) images with some prior knowledge. The conventional interpolation method interpolates an image by convolving the pixels with a fixed kernel, such as Bilinear and Bicubic. These methods do not consider the local structural information in the images. Thus the interpolation results do not adapt to the local structure. Therefore, artifacts, such as blurring and ringing, occur in the edge or texture regions.

Then, the explicit adaptive interpolation method, which utilizes the structural information in an explicit way, turns up. This kind of method identifies the edge or isophote directions and interpolate along the direction [1][2]. The isophote is a line consisting of the connected pixels of the same intensity. The continuous isophote meets the visual rules in the natural image. In Wang's method [1], the isophote direction is

estimated with an angle. Then it interpolates along a parallelogram. One edge of the parallelogram is parallel to the estimated isophote. In [2], a segment adaptive gradient angle interpolation (SAGA) builds an isophote lattice and interpolates along the local isophote rather than the image lattice. By interpolating along the isophote lattice, these methods achieve desirable performance. And fractional locations can be represented in the isophote lattice, so the general scale enlargement is supported.

Another type of methods, the implicit adaptive methods, incorporates the local structural information into the objective function and interpolates by optimizing the function. A representative branch is the autoregressive (AR) based interpolation, such as new edge-directed interpolation (NEDI) and soft-adaptive interpolation (SAI). Based on the geometric duality, NEDI [3] employs the LR image to estimate the high resolution (HR) covariances by a least square problem and estimates the HR pixels with their neighbouring LR pixels. In [4], SAI extends the framework of NEDI with a cross-direction AR model and estimates the interpolation result by a joint estimation. These implicit methods are valid when the statistical stationarity assumption is tenable in the local area. However, due to the diversified content in natural images, the stationarity assumption may be violated even in a small area. Some works [5-7] that extend the AR-based interpolation with the tolerance to the stationarity assumption violation are proposed. In [5], the iterative curve based interpolation (ICBI) forces the AR parameters in the opposite directions equal and uses the second order information to refine the interpolation result to avoid over-constraint. In [6], in order to model the piecewise stationarity, an AR-based interpolation with the geodesic distance weighting is put forward. The weighting scheme tries to simultaneously measure both the spatial distance and the colour difference. In [7], an adaptive general scale interpolation employs a weighting scheme that supports the general scale situations on account of the pixel similarity to increase the accuracy of the estimation. By introducing the inner-patch correlation into the AR model, these methods achieve desirable results. However, another kind of correlation, the inter-patch correlation, is ignored.

In this paper, we follow the AR-based method and propose a novel interpolation method incorporating the information of the inner/inter-patch correlations. In our method,
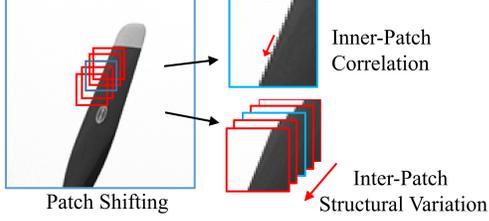
Fig. 1. Illustration for the patches shifting.

Fig. 2. Notations for the AR model in the diagonal (red arrows) and cross (black arrows) directions.

the interpolation window is extended in the isophote direction. And a two-layer similarity metric based on the patch-geodesic distance and the patch distance is performed to model the inner-patch correlation and the inter-patch structural variation. Comprehensive experiments demonstrate that our method achieves desirable performance compared with other interpolation methods in both the objective and subjective quality evaluations.

The rest of the paper is organized as follows: Section 2 presents our method based on the adaptive window AR model. Section 3 suggests a two-layer similarity metric founded on the patch-geodesic distance and the patch distance. Experimental results and analysis are presented in Section 4. Finally, Section 5 concludes this paper.

2. INTERPOLATION BASED ON ADAPTIVE WINDOW AR MODEL

In natural images, "patch shifting" is represented as a widespread phenomenon. As shown in Fig. 1, the successive adjacent image patches in the local area are similar in the spatial intensity distribution. Along the isophote, a line consisting of successive pixels of the same intensity, the similar spatial intensity distributions of the adjacent image patches keep stable. This phenomenon contains two kinds of constraints: the *inner-patch correlation* and the *inter-patch structural variation*. The **inner-patch correlation** indicates the correlation between pixels within a local window. Such correlation has been fully exploited by most of the AR-based methods, which present rather impressive results. However, they neglect the fact that there is another kind of correlation between patches: the *inter-patch structural variation*. The **inter-patch structural variation** indicates the variation trends of the similar successive adjacent patches along the isophote. Both the consistency and the differences between the spatial intensity distributions of the adjacent patches reflect a kind of local structural pattern. In this section, we model the image patch shifting, capturing both the inner-patch similarity and the inter-patch structural variation and interpolate on account of the two types of information.

We first review traditional AR interpolations. Let X and Y be the LR and HR images, x and y be LR and HR pixels, respectively. Given x and X, our goal is to estimate y and Y. $\otimes$ and $\oplus$ denote the neighbours in diagonal and cross directions respectively. As in Fig. 2, $x_{i\otimes t}$ and $y_{i\otimes t}$ represent the t-th neighbours of x_i and y_i in the diagonal direction. $x_{i\oplus t}$ and
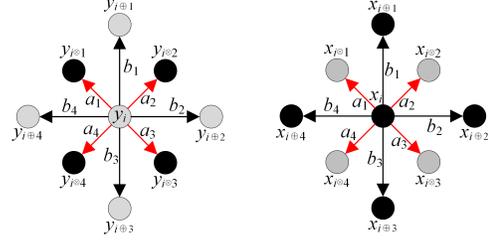
$y_{i\oplus t}$ represent the t-th neighbours of x_i and y_i in the cross direction. Two sets of parameters $\boldsymbol{a} = [a_1, a_2, a_3, a_4]$ and $\boldsymbol{b} = [b_1, b_2, b_3, b_4]$ depict the model parameters in two directions. $\sigma_i^{\oplus}$ and $\sigma_i^{\otimes}$ refer to random perturbations independent of spatial locations and image signal levels. The AR squations for a given pixel in diagonal and cross directions can be represented as:

$$x_i = \sum_{t=1}^{4} a_t x_{i\otimes t} + \sigma_i^{\otimes}, \ y_i = \sum_{t=1}^{4} a_t y_{i\otimes t} + \sigma_i^{\otimes}, \tag{1}$$

$$x_i = \sum_{t=1}^{4} b_t x_{i\oplus t} + \sigma_i^{\oplus}, \ y_i = \sum_{t=1}^{4} b_t y_{i\oplus t} + \sigma_i^{\oplus}. \tag{2}$$

Based on the assumption that images maintain the stationarity in a local window W, we minimize the matching error of the pixels in W by solving the linear least squares problem in Equ. (3):

$$\min_{\{y_i\}} \left\{ \sum_{i\in W} \left(y_i - \sum_{t=1}^{4} a_t y_{i\otimes t} \right)^2 + \sum_{i\in W} \left(x_i - \sum_{t=1}^{4} a_t x_{i\otimes t} \right)^2 \right.$$
$$\left. + \sum_{i\in W} \left(y_i - \sum_{t=1}^{4} b_t y_{i\oplus t} \right)^2 \right\}. \tag{3}$$

Two issues are ignored in the traditional AR models. First, the fixed size of the interpolation window cannot adapt to the local structure at different scales and the stationarity assumption may be violated. Second, the structural information across patches is neglected.

Based on these considerations, we propose a new AR model based on the adaptive window-extension. The model estimates the isophote by the local neighbouring similar patches and then extends the interpolation window in the isophote direction. Then, an irregular window that contains the similar adjacent patches is built. Modulated with the similarity metric mentioned in Section III, the objective function is obtained. Finally, we estimate the HR pixels' value by a closed-form solution. The whole process is shown in Fig. 3.

For the adaptive-window extension, we employ a simple strategy. We define four kinds of windows: basic window, extension window, basic compared window and compared window. The pixel set in a given window is defined as the corresponding patch. The size of the basic window is set as 6×6

in the LR scale. The size of the basic compared window and the compared window is set as 4×4 in the LR scale. The initial interpolation window is set as the basic window and then begins to extend. We search the similar patches in eight directions. Taking left as an example, the six HR pixels in the left direction are considered to be the center of the compared patch. If the mean square error (MSE) between one of the six compared patches and the basic compared patch is less than a threshold, then the window extends in the "left" direction. The grown regions are bounded by the dash line box in left direction. Situations are alike in other directions. In the cross directions, 6 patches are compared in each direction. And in the diagonal directions, 4 patches are compared in each direction. Then, the matching error is represented as:

$$
\min_{\{y_i\}} \left\{ \sum_{i \in \rho(W_b)} \left[sim_i^H w_i \left(y_i - \sum_{t=1}^{4} a_t y_{i \otimes t} \right) \right]^2 + \right.
$$
$$
\sum_{i \in \rho(W_b)} \left[sim_i^L w_i \left(x_i - \sum_{t=1}^{4} a_t x_{i \otimes t} \right) \right]^2 + \quad (4)
$$
$$
\left. \lambda \sum_{i \in \rho(W_b)} \left[sim_i^H w_i \left(y_i - \sum_{t=1}^{4} b_t y_{i \oplus t} \right) \right]^2 \right\},
$$

where W_b is the basic window and ρ is the window-extension operator. λ is the Lagrange multiplier. $\omega(1, k)$ measures the similarity between the current interpolated patch and the k-th compared patch W_k. sim_i^H and sim_i^L are the probabilities between the center interpolated pixel and the HR pixel and the LR pixel, respectively. The details of the similarity metric and the weighting scheme will be elaborated in Section 3.

Let x and y be the vectors consisting of the LR and HR pixels in W, respectively. Let C and D be the vectors consisting of the covariance (a_t or b_t) between pixels. Let S be the diagonal matrix composed of the similarity probability ($sim_i^H w_i$ or $sim_i^L w_i$). We can deduce the objective function in Equ. (4) to a vector form:

$$
\hat{y} = \arg \min_y ||S(Cy - Dx)||_2^2. \quad (5)
$$

Then a close-formed resolution can be obtained:

$$
\hat{y} = (C^T S^2 C)^{-1} C^T S^2 D x.
$$

3. TWO-LAYER SIMILARITY METRIC

In this section, we will present a two-layer similarity metric based on the inter-patch structural variation and the inner-patch correlation.

3.1. Similarity Based on Inter-Patch Structural Variation

Improving the capacity to depict the inter-patch structural variation benefits the AR model and the interpolation. In the proposed method, the interpolation is performed based on an extended window, which is decided by the inter-patch structural variation. Besides, we use a patch similarity to measure the consistency and the difference between the basic compared patch and the compared patch. We use the MSE to
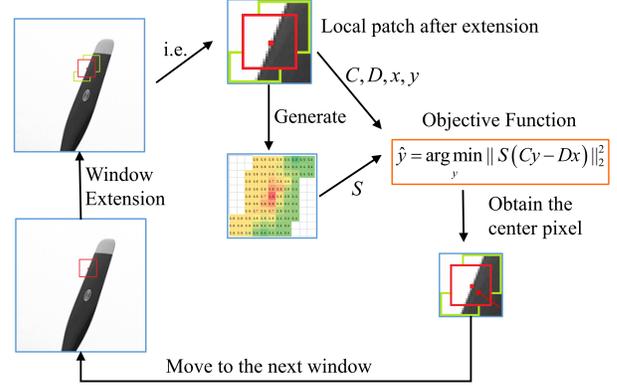

Fig. 3. The flowchart of the proposed method.

depict the patch similarity and modulate it into the AR model in Equ. (4). The weight w_i is defined as:

$$
w_i = \begin{cases} 1, \ p_i \in P_B, \\ \exp \left\{ -\dfrac{max \left\{ MSE(P_{com}(k), P_B) \right\}}{\alpha} \right\}, \\ \qquad \forall k, \ s.t. \ p_i \in P_{com}(k), \end{cases} \quad (6)
$$

where p_i is the i-th pixel (an HR pixel or LR pixel) in the interpolation window. P_B is the basic patch and $P_{com}(k)$ is the k-th compared patch. If a pixel is in the basic compared patch, its probability is set as 1. Otherwise, its probability is set as the maximum value measured by the MSE between the basic patch and all the compared patches where the pixel exists in.

3.2. Similarity Based on Inner-Patch Correlation

The stationarity in natural images provides the way to estimate the parameters in the AR models. On the basis of the stationary assumption, AR models in a local window share the same parameters. Unfortunately, the assumption does not hold even in a very small region due to the diversification of the natural images. Then, the piecewise stationarity, a more general and universal assumption is introduced to model the image. In order to better characterize the piecewise stationarity in a local region, we introduce a novel similarity metric. The metric measures whether two pixels are in the same region and share similar AR model parameters.

A method presented in [6] introduces a geodesic distance as a metric to measure the probability whether two pixels are similar. It incorporates both the spatial distance and the pixel intensity distance, thus very robust to outlier. Due to not considering the pattern similarity, the geodesic distance of two pixels in the same component locating close to different frontiers is small. However, in such a scenario, the pixels own different AR model parameters, which are not represented by the geodesic distance. On account of the consideration, we employ a patch-geodesic distance. Incorporating the pattern similarity, we sum up the differences between patches instead of pixels, with other formulations like the geodesic distance calculation in [6].

4. EXPERIMENTAL RESULTS

All evaluations are performed on the basis of MATLAB 8.20. Bicubic is performed with the MATLAB built-in functions, the source codes of other compared methods are kindly provided by their authors. For thoroughness and fairness of our comparison study, we test ten widely used test images selected from the Kodak database and the USC-SIPI image database in the experiments.

State-of-the-art interpolation methods, new edge directed interpolation (NEDI) [3], soft-decision adaptive interpolation (SAI) [4], iterative curvature-based interpolation (ICBI) [5], segment adaptive gradient interpolation (SAGA) [2] and sparse mixing estimators (SME) [8], are used for comparison.

Table 1. PSNR(dB) results of the five interpolation methods.

Images	Bicubic	SAI	SAGA	SME	Proposed
Child	35.49	35.63	35.48	35.53	**35.67**
Lena	34.01	34.76	34.45	34.61	**34.77**
Pepper	32.06	31.84	32.43	**34.61**	32.71
Monarch	31.93	33.08	32.58	32.69	**33.34**
Airplane	29.40	29.62	29.84	30.00	**30.09**
Caps	31.25	31.64	31.58	31.60	**31.71**
Statue	31.36	31.78	31.81	31.55	**31.96**
House	22.20	22.28	**22.49**	22.34	22.40
Woman	31.17	31.27	31.33	31.15	**31.35**
Bike	25.41	26.28	25.92	26.08	**26.33**
Lighthouse	26.97	26.70	**27.25**	27.23	27.04
Average	30.11	30.44	30.47	30.51	**30.67**

To compare the objective quality of the different interpolation methods, the original HR images are first directly downsampled by a factor of two to generate the input LR images. Then the different interpolation methods are applied to interpolate the input LR images to their original resolutions. Table 1 tabulates the PSNR results of the five interpolation methods on several images in our experiments. From Table 1, we can see that the proposed method produces the comparable or often better PSNR results than other methods. The average PSNR gain is 0.16dB higher than the second-best SME algorithm. It is worth noticing that for *Monarch* and *Statue* image, the proposed method gains 0.65 dB and 0.40 dB respectively higher than the second-best SME algorithm.

We also compare the visual quality of the different interpolation methods In Fig. 4. For *Barbara*, because the downsampling may cause the direction of the stripes changing across different scales, the interpolation results based on the LR image will cause the prediction errors. In the sub-regions of *Barbara* in Fig. 4, most methods make the wrong prediction and cause the interpolation errors. Due to the robust modelling capacity of the patch-geodesic similarity, our method can suppress the artifacts and reduce errors. For *Cameraman* and *Sailboat*, the test images exhibit the strong and sharp edges. Our method presents the desirable local results and obtains the darkest difference image, which means the results of our method contains fewer errors and more similar to the corresponding HR image. It can be obviously observed
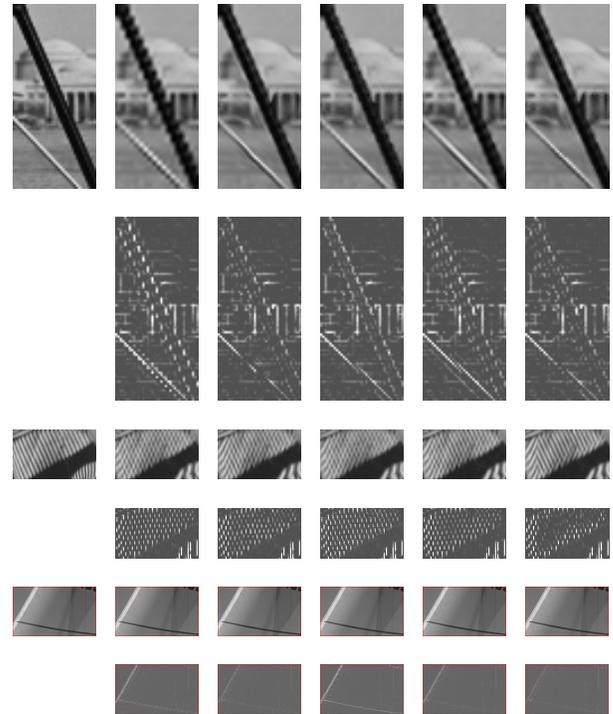

Fig. 4. Visual comparison between different algorithm results for the local images and the difference images. From up to down: *Cameraman*, *Barbara*, *Sailboat*. From left to right: Original, Bicubic, SAI, SAGA, SME, Proposed method.

that the images interpolated by the Bicubic interpolator suffer from blurred edges, jaggies and annoying ringing artifacts. SAGA and SME show improvements over SAI and NEDI in the regions of edges and textures, reducing the visual defects these methods bring. Thanks to the combination of the inner-patch correlation and the inter-patch structural variation, our method achieves better visual quality compared with all other methods. Our algorithm produces fewer interpolation errors than other methods. Such results clearly demonstrate the superiority of our method in reconstructing the high frequency part, such as edges and textures, of the images.

5. CONCLUSION

In this paper, we propose a novel interpolation method based on the adaptive window-extension AR model. The model combines the information of the inner-patch correlation and the inter-patch structural variation. The interpolation window is extended in the direction of the isophote or to the location similar patches exist in. A two-layer metric based on the patch-geodesic similarity and the patch distance is performed and the metric is modulated into the AR model. Comprehensive experiments demonstrate that our method achieves desirable performance compared with other interpolation methods no matter in objective or subjective quality evaluations.

6. REFERENCES

[1] Qing Wang and Rabab Kreidieh Ward, "A new orientation-adaptive interpolation method," *IEEE Transactions on Image Processing*, vol. 16, no. 4, pp. 889–900, 2007.

[2] C. M. Zwart and D. H. Frakes, "Segment adaptive gradient angle interpolation," *IEEE Trans Image Process*, vol. 22, no. 8, pp. 2960–9, 2013.

[3] X. Li and M. T. Orchard, "New edge-directed interpolation," *IEEE Trans Image Process*, vol. 10, no. 10, pp. 1521–7, 2001.

[4] L. Zhang and X. Wu, "An edge-guided image interpolation algorithm via directional filtering and data fusion," *IEEE Trans Image Process*, vol. 15, no. 8, pp. 2226–38, 2006.

[5] A. Giachetti and N. Asuni, "Real-time artifact-free image upscaling," *IEEE Trans Image Process*, vol. 20, no. 10, pp. 2760–8, 2011.

[6] Tang Ketan, O. C. Au, Guo Yuanfang, Pang Jiahao, and Li Jiali, "Arbitrary factor image interpolation using geodesic distance weighted 2d autoregressive modeling," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pp. 2217–2221.

[7] Li Mading, Liu Jiaying, Ren Jie, and Guo Zongming, "Adaptive general scale interpolation based on similar pixels weighting," pp. 2143–2146, 2013.

[8] S. Mallat and G. Yu, "Super-resolution with sparse mixing estimators," *IEEE Trans Image Process*, vol. 19, no. 11, pp. 2889–900, 2010.