Color Image Super Resolution: A Two-Step Approach Based on Geometric Grouplets

Aldo Maalouf and Mohamed-Chaker Larabi; University of Poitiers, XLIM Laboratory, Department of Signal, Image and Communication, UMR CNRS 6172; SP2MI-2 Bd Marie et Pierre Curie, PO Box 30179, 86962 Futuroscope Chasseneuil, France

Abstract

In this work, a two-step technique for constructing a super-resolution (SR) image from a single multi-valued low-resolution (LR) input image is proposed. The problem of SR is treated from the perspective of image geometry-oriented interpolation. The first step consists of computing the image geometry of the LR image by using the grouplet transform. Having well represented the geometry of each color channel in the LR image, we propose a grouplet-based structure tensor whose role is to couple the geometrical information of the different image color components. In a second step, a functional is defined on the multispectral geometry defined by this structure tensor. The minimization of this functional insures the synthesize of the SR image. The proposed super-resolution algorithm outperforms the state-of-the-art methods in terms of visual quality of the interpolated image.

Introduction

Image super-resolution is the process of increasing the resolution of a given image [2, 3, 4]. This process has also been referred to in the literature as resolution enhancement. One such application to image super-resolution can be found in streaming video websites, which often store video at low resolutions for various reasons. The problem is that users often wish to expand the size of the video to watch at full screen with resolutions of 1024×768 or higher, and this process requires that the images be interpolated to a higher resolution. Another application comes from the emergence of digital cinema, where filmmakers are increasingly turning towards an all-digital solution, from image capture to postproduction and projection. Due to its fairly recent appearance, the digital cinema chain still suffers from limitations which can hamper the productivity and creativity of cinematographers and production companies. One of these limitations is that the cameras used for high resolutions are expensive and the data files they produce are large. Because of this, studios may chose to capture some sequences at lower resolution (2K for example). These sequences can later be interpolated to 4K sequences by using a super-resolution technique and projected in higher resolution display devices.

Increasing the resolution of the imaging sensor is clearly one way to increase the resolution of the acquired images. This solution, however, may not be feasible due to the increased associated cost and the fact that the shot noise increases during acquisition as the pixel size becomes smaller. Furthermore, increasing the chip size to accommodate the larger number of pixels increases the capacitance, which in tern reduces the data transfer rate. Therefore, image processing techniques, like the ones described in this paper, provide a clear alternative for increasing the resolution of the acquired images.

There are various possible models for performing resolution enhancement. These models can be grouped in three categories: interpolation-based methods, reconstruction-based methods, and learning-based methods.

The most common methods used in practice are the bicubic and bilinear-based interpolation methods[5] [6], which require only a small amount of computation. However, these simple methods often produce images with various artefacts along object boundaries, including aliasing, blurring, and zigzagging edges. The reconstruction based methods [8] [7] enforce a reconstruction constraint which requires that the smoothed and down-sampled version of the high resolution (HR) image should be close to the LR image. The learning based methods [9] [10] "hallucinate" high frequency details from a training set of HR/LR image pairs. The learning based approach highly relies on the similarity between the training set and the test set. It is still unclear how many training examples are sufficient for the generic images.

Meanwhile, various algorithms have been proposed to improve the interpolation-based approaches and reduce edge artifacts, aiming at obtaining images with regularity (i.e. smoothness) along edges. Jensen and Anastassiou [11] propose to estimate the orientation of each edge in the image by using projections onto an orthonormal basis and the interpolation process is modified to avoid interpolating across the edge. Allebach and Wong [12] propose to use an estimate of the high-resolution edge map to iteratively correct the interpolated pixels. Li and Orchard [5] propose a statistical estimation method that tunes interpolation coefficients according to local edge structures. While this method produces good results, its computational complexity is prohibitive due to the large window size used to estimate local covariances. Improvements for the interpolation methods have also been proposed in [13], where partial differential equations (PDE) are used to satisfy smoothness and orientation constraints. Other proposed approaches perform interpolation in a transform (e.g. wavelet) domain [14] [15]. These algorithms assume the low-resolution image to be the lowpass output of the wavelet transform and make use of dependence across wavelet scales to predict the "missing" coefficients in the more detailed scales.

The above listed SR methods have been designed to increase the resolution of a single channel (monochromatic) image. To date, there is very little work addressing the problem of color SR. The typical solution involves applying monochromatic SR algorithms to each of the color channels independently [16] [17], while using the color information to improve the accuracy. Another approach is transforming the problem to a different color space, where chrominance layers are separated from luminance, and SR is applied only to the luminance channel [18]. Both of these methods are sub-optimal as they do not fully exploit the correlation across the color bands. Other methods using learning based techniques have been proposed for color image interpolation [19], yet results still depend on the training phase and the used dataset.

In this paper, we propose a novel variational color image interpolation algorithm based on the new grouplet transform [20] which provides an efficient multiscale geometric representation for natural images. The grouplet transform was proposed by Mallat as a

directional multiresolution transform that can efficiently capture and represent boundaries and textures in natural images. Furthermore, it allows to define a geometrical flow that can be used to direct our interpolation technique. Having well represented the geometry of each color channel by using geometrical grouplets, we propose a grouplet-based structure tensor whose role is to couple the geometrical information of the different image color components. Then, a functional is defined on the multispectral geometry defined by this structure tensor. The minimization of this functional insures the synthesize of the SR image.
The rest of the paper is organized as follows. Section 2 presents the grouplet transform and provides motivation for its use in our algorithm. In Section 3, we present our variational super-resolution algorithm. We report the results of our experiments in Section 4 and conclude the paper in Section 5.

Geometrical Grouplets

Geometrical grouplets have been recently introduced by Mallat in [20]. They are constructed with association fields that group points to take advantage of the geometrical image regularities. We only present here a brief review of the Grouplet transform. The reader can refer to [20] for a full detailed description of the Grouplet transform.
Because they are constructed with the spirit of the "Gestalt" psychophysic school, the grouplets construct the geometry of the image with grouping processes which define a multiscale association field. This field define the different orientations even for the fine structures in the images.
Then, an orthogonal multiscale grouping is implemented with a weighted Haar lifting applied successively to points that are grouped by the association field. The associated field is computed by first performing group matching on the 2D wavelet transform coefficients of the image. The role of this field is to group together points that have similar neighborhoods in order to exploit the geometry of the image. The computation of the association field is performed as follows: First the image grid Ω is divided into two subgrids Ω_{even} (even columns) and Ω_{odd} (odd columns). Then, each point in the odd subgrid is associated to a point in the even subgrid according to a block matching criterion. This can be achieved by searching for each pixel m in Ω_{odd} for a best matching pixel $\tilde{m}$ in Ω_{even}. The search is performed on the neighborhood of m denoted by $N(m)$ using the following equation:

$$\tilde{m} = \underset{p \in N(m)}{\arg\min} |a[m] - a[m+\tilde{m}]|^k, \qquad k = 1 \text{ or } k = 2 \tag{1}$$

where a is the corresponding wavelet coefficient.
We denote by $A_j[m]$ the association field that stores a displacement from $m \in \Omega_{odd}$ to an image point $\tilde{m} \in \Omega_{even}$ at a wavelet scale j.
Figure 1 shows an example of the association field computed for 'Lenna' image. As we can see, the block matching (1) computes regular multiscale association fields with directions along which the image is smoothly varying. Once the association field is obtained, the grouplet coefficients are obtained by performing a weighted multiscale Haar transform on the associated wavelet coefficients.
For scales 2^j that increase from 1 to 2^J, the successive groupings are performed as follows:
each point m is associated to another point $m = \tilde{m} + A_j[\tilde{m}]$. The grouplet transform computes a normalized difference between associated averages:

$$d_j[\tilde{m}] = (h[\tilde{m}] - h[m]) \frac{\sqrt{s[m]\, s[\tilde{m}]}}{\sqrt{s[m] + s[\tilde{m}]}} \tag{2}$$

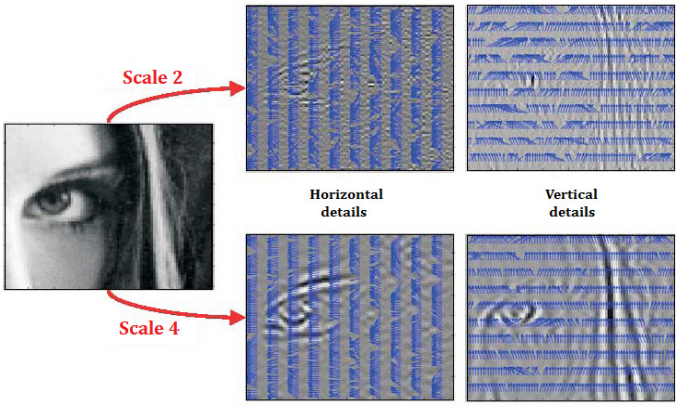


Figure 1. *Association field of 'lenna' image*

A weighted average is then defined by:

$$\tilde{h} = \frac{s[m]\, h[m] + s[\tilde{m}]\, h[\tilde{m}]}{s[m] + s[\tilde{m}]} \tag{3}$$

where

$$h[m] = \frac{h[2m] + h[2m+1]}{2}$$

The averaging size is updated by adding the averaging size of the two averaged points

$$\tilde{s} = s[m] + s[\tilde{m}] \tag{4}$$

Compared to other geometrical representations, such as bandelet or curvelet transforms, the grouplet transform is more flexible since the association fields can deviate from the integral lines in order to converge to singularity points such as junctions or crossings. Fine image structures are consequently well represented. Therefore, the interpolation of the represented information in the "missing" (or to be synthesized) pixels of the SR image following the directions of the association field yields to a precise synthesis of the SR image. We present in the following section this interpolation technique.

Grouplet-Based Super-Resolution Technique

Given that the image geometry is efficiently represented and characterized by the multiscale association field, we present in this section an interpolation method oriented by the captured geometry. First, we present our grouplet-based structure tensor and then, we describe our interpolation technique.

Grouplet-based structure tensor

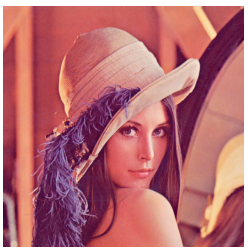
Extending differential-based operations to color or multi-valued images is hindered by the multi-channel nature of color images. The derivatives in different channels can point in opposite directions, hence cancelation might occur by simple addition. The solution to this problem is given by the structure tensor for which opposing vectors reinforce each other.
In [1] Di Zenzo pointed out that the correct method to combine the first order derivative structure is by using a local tensor. Analysis of the shape of the tensor leads to an orientation and a gradient norm estimate. For a multichannel image $I = \left(I^1, I^2, ..., I^n\right)^T$ the structure tensor is given by

$$M = \begin{pmatrix} I_x^T I_x & I_x^T I_y \\ I_y^T I_x & I_y^T I_y \end{pmatrix} \tag{5}$$

The multichannel structure tensor describes the 2D first order differential structure at a certain point in the image.
The motivation of this work is to make the interpolation oriented by the optimal geometry direction captured by the grouplet transform in order to synthesize fine structures for the SR image. For that purpose, a multiscale multistructure grouplet-oriented tensor for an m-valued ($m=3$ for color images and $m=1$ for gray images) image is defined by:

$$G^j = \begin{bmatrix} \sum_{i=1}^{m} \left(\frac{\partial}{\partial x}\tilde{h}_i^j \cos\theta_i\right)^2 & \sum_{i=1}^{m} \frac{\partial}{\partial x}\tilde{h}_i^j \cos\theta_i \frac{\partial}{\partial y}\tilde{h}_i^j \sin\theta_i \\ \sum_{i=1}^{m} \frac{\partial}{\partial x}\tilde{h}_i^j \cos\theta_i \frac{\partial}{\partial y}\tilde{h}_i^j \sin\theta_i & \sum_{i=1}^{m} \left(\frac{\partial}{\partial y}\tilde{h}_i^j \sin\theta_i\right)^2 \end{bmatrix} \quad for \quad i=1,2,.....m \tag{6}$$

The norm of G^j is defined in terms of its eigenvalues λ_+ and λ_-, $\left|\left|G^j\right|\right| = \sqrt{\lambda_+ + \lambda_-}$. The angle θ_i represents the angle of the direction of the grouplet association field. j is the scale of the grouplet transform. g_i^j is the corresponding grouplet coefficient. i designates the image channel ($i=1,2,\ldots,m$). Figure 2 shows the norm of the grouplet-based structure tensor defined in (6) of the 'Lenna' image.



(a) Lenna Image

(b) Norm of the structure tensor defined in (6)

Figure 2. *Norm of the grouplet-based structure tensor*

Until now, we have characterized edges and the geometrical flow (the association field) of the image. We present in the following subsection our super-resolution variational approach that is oriented by these two geometric features.

Super-Resolution

We formulate our interpolation approach as the following variational problem,

$$\tilde{I}_i = \min_{I_i}\left(\int_\Omega \left(\|\nabla I_i(x,y)\| + \left\|\nabla G^j(x,y)\right\| + \lambda \left\|G^j(x,y)\right\|\right) d\Omega\right) \tag{7}$$

subjected to the following constraints,

$$I(xs\Delta, ys\Delta) = I'(x,y) \qquad \begin{matrix} 0 \le x \le \left\lfloor \frac{w}{s\Delta} \right\rfloor \\ 0 \le y \le \left\lfloor \frac{h}{s\Delta} \right\rfloor \end{matrix} \tag{8}$$

where $I'(x,y)$ is the original image before interpolation, Δ is the grid size of the upsampled image, w and h are the width and the height of the image respectively and s is the scaling factor.
$\tilde{\nabla}$ is the directional gradient with respect to the grouplet geometric direction θ and λ is a constant.
The first term in (7) is a marginal regularization term oriented by the directions of the geometrical flow defined by the association fields of the grouplet transform. The second is a multispectral regularization term while the third is edge-driven, which aims at the orientation of the interpolation process to the orientation of the color edges. In fact, the norm $\left\|G^j(x,y)\right\|$ is a weighting factor such that more influence is given to points where the color gradient is high in the interpolation process.
The Euler equation of (7) is

$$\tilde{\nabla}\cdot\left(\frac{\nabla I_i(x,y)}{\|\nabla I_i(x,y)\|}\right) + \tilde{\nabla}\cdot\left(\frac{\nabla\left\|G^j(x,y)\right\|}{\|\nabla I_i(x,y)\|}\right) + \lambda\frac{\tilde{\nabla}\left\|G^j(x,y)\right\|}{\|\nabla I_i(x,y)\|} = 0 \tag{9}$$

By expanding (9) we obtain after simplification,

$$\begin{aligned} &I_{i_{xx}}\cos\theta + I_{i_{yy}}\sin\theta - \left(I_{i_{xy}} + I_{i_{yx}}\right)\cos\theta\sin\theta + \\ &\left\|G^j(x,y)\right\|_{xx}\cos\theta + \left\|G^j(x,y)\right\|_{yy}\sin\theta + \\ &\lambda\left\|G^j(x,y)\right\|_x\cos\theta + \lambda\left\|G^j(x,y)\right\|_y\sin\theta = 0 \end{aligned} \tag{10}$$

where $\left\|G^j(x,y)\right\|_x$ and $\left\|G^j(x,y)\right\|_y$ are, respectively, the horizontal and vertical derivatives of the norm matrix $\left\|G^j(x,y)\right\|$ computed at a scale j to extract the horizontal and vertical details of the color image.
Equation (10), which yields to a factor-of-two interpolation scheme, is applied to each color band i and it can be easily discretized by using the following two equations:

$$\begin{aligned} I_{i_{xx}}(x\Delta, y\Delta) &\approx \left[I_i(x-1,y) + I_i(x+1,y) - 2I_i(x,y)\right]/\Delta^2 \\ I_{i_{yy}}(x\Delta, y\Delta) &\approx \left[I_i(x,y-1) + I_i(x,y+1) - 2I_i(x,y)\right]/\Delta^2 \\ I_{i_{xy}}(x\Delta, y\Delta) &\approx \left[I_i(x+1,y+1) + I_i(x-1,y-1) - \right. \\ &\qquad \left. I_i(x-1,y+1) \quad - I_i(x+1,y-1)\right]/4\Delta^2 \end{aligned} \tag{11}$$

$$\begin{aligned} I_i(x,y-1) &\approx \left[I_i(x+1,y-1) + I_i(x-1,y-1)\right]/2 \\ I_i(x,y+1) &\approx \left[I_i(x+1,y+1) + I_i(x-1,y+1)\right]/2 \\ I_i(x-1,y) &\approx \left[I_i(x-1,y-1) + I_i(x-1,y+1)\right]/2 \\ I_i(x+1,y) &\approx \left[I_i(x+1,y-1) + I_i(x+1,y+1)\right]/2 \end{aligned} \tag{12}$$

By using equation (10) and solving for $I(x\Delta, y\Delta)$ (λ is set equal to -2 to eliminate the term $G^j(x,y)$ which is unknown), we obtain the final interpolating scheme on the upsampled grid (for simplicity $I(x,y)$ is used to denote $I(x\Delta, y\Delta)$):

$$\begin{aligned} I_i(x,y) = &\left(\tfrac{\cos\theta_i+\sin\theta_i-4\cos\theta_i\sin\theta_i}{4(\cos\theta_i+\sin\theta_i)}\right)\left(I_i(x-1,y-1) + \right. \\ &\left. I_i(x+1,y+1)\right) + \left(\tfrac{\cos\theta_i+\sin\theta_i+4\cos\theta_i\sin\theta_i}{4(\cos\theta_i+\sin\theta_i)}\right)\left(I_i(x-1,y+1) + \right. \\ &\left. I_i(x+1,y-1)\right) + \tfrac{1}{4}\left\|G^j(x-1,y-1)\right\| + \\ &\left(\tfrac{\cos\theta_i+3\sin\theta_i}{4(\cos\theta_i+\sin\theta_i)}\right)\left\|G^j(x-1,y+1)\right\| + \\ &\left(\tfrac{\sin\theta_i-\cos\theta_i}{4(\cos\theta_i+\sin\theta_i)}\right)\left\|G^j(x+1,y-1)\right\| + \\ &\left(\tfrac{3\sin\theta_i-\cos\theta_i}{4(\cos\theta_i+\sin\theta_i)}\right)\left\|G^j(x+1,y+1)\right\| \end{aligned} \tag{13}$$

where

$$\begin{aligned} \theta_i = &\tfrac{1}{4}\left(\theta_i(x-1,y-1) + \theta_i(x+1,y-1)\right) \\ &+ \left(\theta_i(x-1,y+1) + \theta_i(x+1,y+1)\right) \end{aligned}$$

The process can be iterated with the resolution increased by a factor of two on each iteration. We represent in the following section some experimental results.

Experimental Results

We propose to evaluate subjectively (panel of observers) and objectively (metrics) the results obtained by the proposed algorithm and compare them with the results obtained by the state-of-the-art approaches.

Subjective evaluation

Subjective experiments consist in asking a panel of subjects to watch a set of images or video sequences and to score their quality. The main output of these tests is the Mean Opinion Score (MOS) computed using the values assigned by the observer. In order to obtain meaningful and useful values of MOS, the experiment need to be constructed carefully by selecting rigorously the test material and defining scrupulously the subjective evaluation procedure. The most important recommendations have been published by ITU [21, 22] or described in VQEG test-plans [23]. Figure 3 gives an overview of the used test material.

Figure 3. *Overview of the test material. a-Lena, b- Lena (grayscale), c-Lighthouse, d- Iris, e-Caster and f-Haifa.*

All images described above have been downsampled by a ratio of 2 in width and height. Then it has been provided as an input of the super-resolution algorithms. This process allows to compare algorithms with regards to the original image.

Environment setup

The subjective experiments took place in a normalized test-room built with respect to ITU standards [21] (cf. figure 4). It is very important to control accurately the environment setup in order to ensure the repeatability of the experiments and to be able to compare results between different test locations.

Figure 4. *A Synthetized view of the used test room*

Only one observer per display has been admitted during the test session. This latter is seated at a distance between 2H and 4H; H being the height of the displayed image. His vision is checked for acuity and color blindness. Table 1 provides the most important features of the used display. The ambient lighting of the test-room has been chosen with a color temperature of 6500 K.

Display characteristics for the subjective evaluation

Type	Dell 3008WFP
Diagonal size	30 inches
Resolution	2560 $\times$ 1600 (native)
Calibration tool	EyeOne Display 2
Gamut	sRGB
White point	D65
Brightness	370 cd/m^2
Black level	lowest

Subjective evaluation procedure

In order to compare our SR approach with other super-resolution algorithms from the point of view of subjective evaluation, we used a single stimulus approach. This means that processed images are scored without any comparison with the original image (reference image). This latter is used as a result image and is scored in order to study the reliability and accuracy of the observer results.

The test session starts by a training allowing to show to the observer the types of degradation and the way to score impaired images. The results for these items are not registered by the evaluation software but the subject is not told about this. Then, each image is displayed for 10 seconds, three times, to the observer to stabilize his judgment. At the end of each presentation, a neutral gray is displayed with a GUI containing a discrete scale as shown in figure 5. This scale corresponding to a quality range from bad to excellent ([0 - 5]) allows to affect a score to each image. Of course, numbers shown on figure 5 are given here for illustration and do not exist on the GUI.

Figure 5. *Discrete quality scale used to score images during the test*

Each subjective experiment is composed of 198 stimulus: 6 images $\times$ 11 (10 algorithms + reference image) $\times$ 3 repetitions in addition to 5 stabilizing images (training). A panel of fifteen observers has participated to the test. Most of them were naive subjects. The presentation order for each observer is randomized. To better explain the aim of the experiment and the scoring scale, we give the following description to the observers: Imagine you receive an image as an attachment of an email. The resolution of this latter does not fit with the display and you want to see it in full screen. A given algorithm performs the interpolation and you have to score the result as: Excellent: the image content does not present any noticeable artifact; Good: The global quality of the image is good even if a given artifact is noticeable; Fair: several noticeable artifacts are noticeable all over the image; Poor: many noticeable artifacts and strong artifacts corrupt the visual quality of the image; Bad: strong artifacts are detected and the image is

unusable.

Scores processing

The raw subjective scores have been processed in order to obtain the final Mean Opinion Scores (MOS) presented in the results section.

The MOS $\bar{u}_{jkr}$ is computed for each presentation:

$$\bar{u}_{jkr} = \frac{1}{N}\sum_{i=1}^{N} u_{ijkr} \tag{14}$$

where u_{ijkr} is the score of the observer i for the impairment j of the image k and the r^{th} iteration. N represents the number of observers. In a similar way, we can calculate the global average scores, $\bar{u}_j$ and $\bar{u}_k$, respectively for each test condition (algorithm) and each test image.

Objective evaluation

Objective quality measurement is an alternative of a tedious and time consuming subjective assessment. In literature there is a plenty of metrics (Full reference, Reduced reference and no reference) that models or not the Human Visual System (HVS). Most of them are not very popular due to their complexity, difficult calibration or lack of freely available implementation. This is why metrics like PSNR and SSIM [25] are widely used to compare algorithms.

PSNR is the most commonly used metrics and its calculation is based on the mean squared error (MSE).

$$PSNR(x,y) = 20log_{10}\frac{255}{MSE(x,y)} \tag{15}$$

Structural SIMilarity (SSIM) works under the assumption that human visual perception is highly adapted for extracting structural information from a scene. So, it directly evaluates the structural changes between two complex-structured signals.

$$SSIM(x,y) = l(\mu_x,\mu_y)^{\alpha}c(\sigma_x,\sigma_y)^{\beta}s(\sigma_x,\sigma_y)^{\gamma} \tag{16}$$

Evaluation results

Nine super-resolution algorithms coming for the state of the art have been evaluated objectively and subjectively and compared with our SR algorithm: *A* for Allebach [12] , *C* for Chang [14], *E* for Elad [26], *H* for Hardie [27], *I* for Irani [18], *J* for Jensen [11], *L* for Li [5], *Mu* for Muresan [15], *P* for Patti [28]. *Ma* denotes our approach in the results shown below.

Graphics a,b,c d and e of figure 6 show, for each image from the test material, the MOS values obtained after the processing applied to the subjective scores. It shows also the confidence interval associated with each algorithm.

From the subjective scores, one can notice that our approach outperforms the other existing approaches and that the evaluated algorithms can be grouped in three classes: Low quality algorithms (*E*, *H*, *I* and *J*), Medium quality algorithms (*A* and *C*) and High quality algorithms (our approach, *Mu*, *P*). Only one algorithm, *L*, seems to be content dependent and provide results that can be put in medium and high quality groups.

For the subjective experiments, we inserted the original images with the test material without giving this information to the observers. The scores obtained for these images are high, approximatively ranging within the 20% highest scores. However, The difference between images is relatively high. This is due to the acquisition condition of the image itself. Figure 7 gives the MOS values and the associated confidence interval for the original images. Obviously, Haifa and Lighthouse are around 5 and

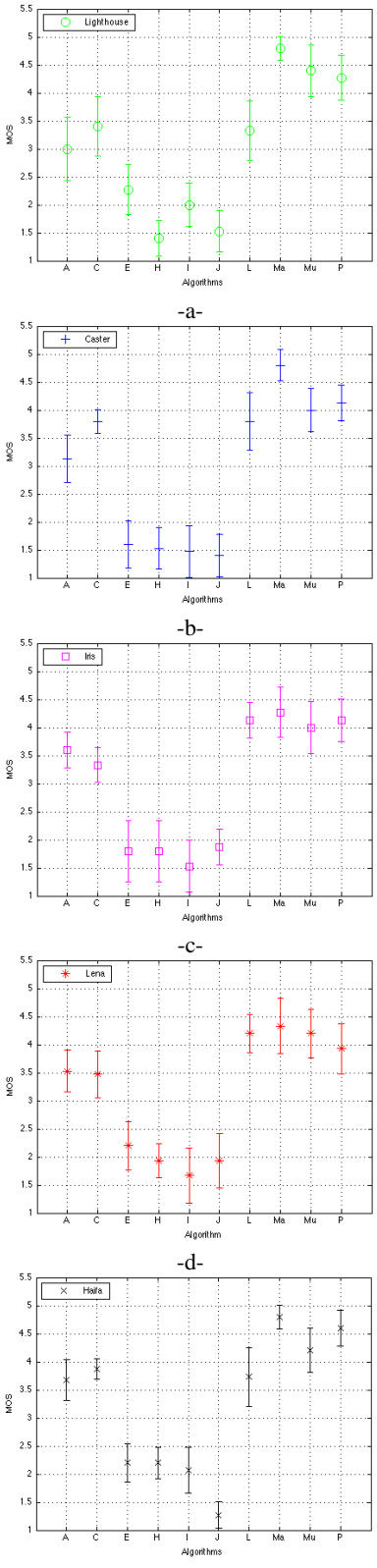


Figure 6. *Mean Opinion Scores (MOS) values and 95% confidence interval obtained for the different algorithms for: a- Lighthouse, b- Caster, c- Iris, d- Lena and e- Haifa.*

have a very small confidence interval because these images are relatively sharp and colourful. The worst was Lena because its background contains some acquisition artifacts that can be considered by the observers as generated by the super-resolution algorithm.
One important thing that we can exploit from these results is to use them as an offset to calibrate the MOS of the test images.

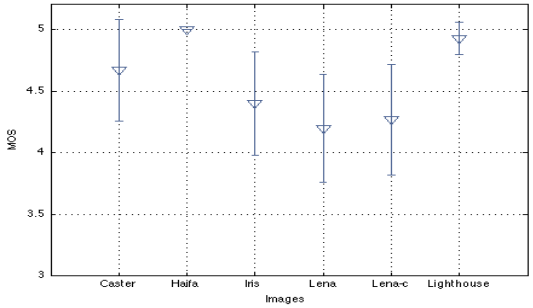


Figure 7. *Mean Opinion Scores (MOS) values and 95% confidence interval obtained for the original images.*

The test material contains two versions of Lenna i.e. color and grayscale. This has been used to study the effect of the super-resolution algorithms on colors and on human judgment. Figure 8 shows MOS values and their confidence intervals for color and grayscale versions. First of all, the scores are relatively close and it is impossible to draw a conclusion about which is the best. Then, the confidence intervals are of the same size approximatively. Finally, these results leads to the conclusion that the used algorithms either allow to conserve the color information or do not deal with the color information in their conception. Hence, for the evaluation of super-resolution algorithms (those used here at least) one can use the Luminance information rather than the three components.

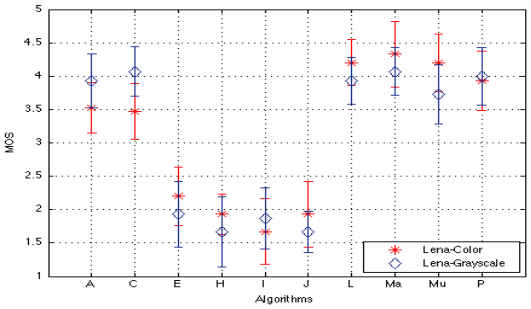


Figure 8. *Mean Opinion Scores (MOS) values and 95% confidence interval obtained for Lena in color and grayscale*

The PSNR and the SSIM have been used to evaluate the quality of the test material used for the subjective evaluation. Figure 9 and figure 10 show respectively the results for PSNR and SSIM. It is difficult to draw the same conclusion than the subjective assessment because the categories are not clearly present especially for PSNR. This confirms its lack of correlation with the human perception. However, from figure 9 the low quality category (PSNR lower than 32dB) is confirmed for algorithms *E*, *H*, *I* and *J*.

One can notice that the other algorithms perform better especially our algorithm (*Ma*). Results of figure 10 are more correlated to human perception than the PSNR because, on the one hand, we can retrieve the same group of high quality algorithms

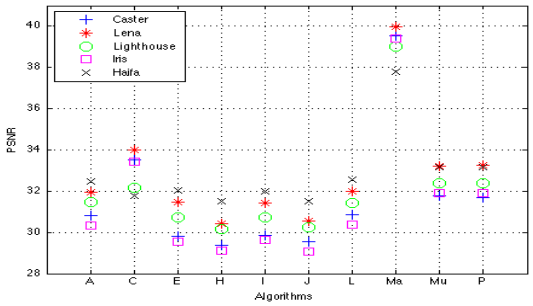


Figure 9. *PSNR results for the five images and the ten algorithms.*

with values very close to one. On the other hand, the medium quality group can be considered at values between 0.96 and 0.98. For low quality algorithms, it is really difficult to have a clear range of scores.

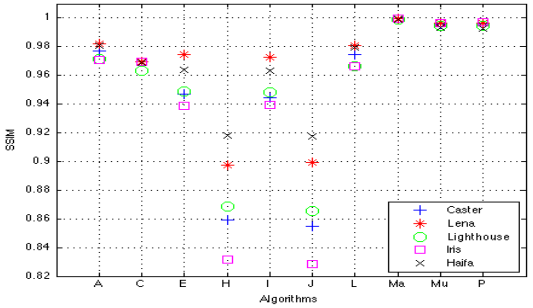


Figure 10. *SSIM results for the five images and the ten algorithms.*

In order to confirm the declaration about the more or less correlation of the PSNR and the SSIM with the human perception, we computed the Pearson correlation coefficient (PCC). Table 2 gives the PCC values first for each image and then for the global data. The PCC values show clearly that SSIM is more correlated than PSNR but the correlation is not very high.

Pearson correlation coefficient between objective metrics (PSNR, SSIM) and subjective scores.

Image	SSIM	PSNR
Caster	0,8224	0,7486
Haifa	0,8758	0,6542
Iris	0,7749	0,6232
Lena	0,7223	0,6465
Lighthouse	0,8963	0,7510
Global	0,7866	0,6745

Figures 11-a and 11-b give scatter plots for the correlation of the PSNR and the SSIM. It easy to notice that the correlation of the first is lower than the second and both are low with regards to human perception. This means that the use of these metrics to replace the human judgment for super-resolution algorithms evaluation is to a certain extent incorrect.

These subjective and objectives evaluations showed that our grouplet-based algorithm outperform the other state-of-the-art approaches and confirmed our hypothesis that taking into account the geometry of the color image can enhance the quality of the SR image.

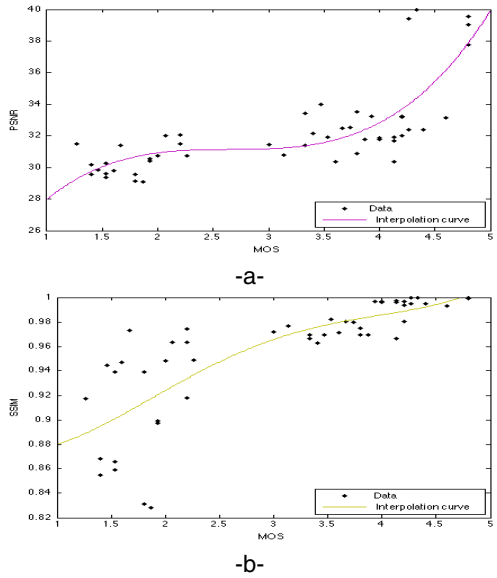


Figure 11. *Scatter plots between the MOS values collected for all images and the PSNR (a) and the SSIM (b)*

Conclusion

In this work, a grouplet-oriented color image interpolation technique is presented. First, the geometrical flow of the image is computed by using the grouplet transform. Then, a grouplet-based structure tensor is defined to combine geometric information of the different color channels of the image. A variational approach is finally defined on the captured geometry. The minimization of the proposed functional ensures the synthesize of the SR image. The proposed algorithm has been evaluated subjectively by psychophysical experiments allowing to quantify the human judgment and objectively by using two common metrics: PSNR and SSIM. The performed evaluations showed that our approach has an improvement over existing SR techniques on a subjective and objective scale.

References

[1] DiZenzo S., "A note on the gradient of multi images," *Computer Vision Graphics and Image processing*, vol. 33 (1), pp. 116–125, 1986.

[2] S. Borman and R. L. Stevenson, Super-Resolution from Image Sequences-A Review, Midwest Symposium on Circuits and Systems, 374-378, 1998.

[3] S. C. Park, M. K. Park and M. G. Kang, Super-resolution image reconstruction: a technical overview, IEEE Signal Processing Magazine, 20(3):21-36, May 2003.

[4] S. Farsiu, D. Robinson, M. Elad and P. Milanfar, Advances and Challenges in Super-Resolution, International Journal of Imaging Systems and Technology, 14(2), pp. 47-57, August 2004

[5] X. Li and M. T. Orchard, New edge-directed interpolation, IEEE transactions on Image Processing, 10(10):1521-1527, 2001.

[6] T. Blu P. Thevenaz and M. Unser, Image interpolation and resampling, academic press, san diego, USA, Handbook of Medical Imaging, Processing and Analysis, 2000.

[7] Z. C. Lin M. Ben-Ezra and B. Wilburn, Penrose pixels: Superresolution in the detector layout domain, Proc. ICCV 2007, 2007.

[8] S. Baker and T. Kanade, Limits on super-resolution and how to break them, IEEE Trans. on PAMI, 24(9):1167-1183, 2002.

[9] H. Y. Shum C. Liu and W. T. Freeman, Face hallucination: Theory and practice, International Journal of Computer Vision, 75(1):115-134, 2007.

[10] X. Tang Q. Wang and H. Y. Shum, Patch based blind image super resolution, Proc. ICCV 2005, 1:709-716, 2005.

[11] K. Jensen and D. Anastassiou, Subpixel edge localization and the interpolation of still images, IEEE transactions on Image Processing, 4:285-295, 1995.

[12] J. Allebach and P. W. Wong, Edge-directed interpolation, Proc. IEEE Int. Conf. on Image Proc., 1996.

[13] H. Jiang and C. Moloney, A new direction adaptive scheme for image interpolation, Proc. IEEE Int. Conf. on Image Proc., pages 369-372, 2002.

[14] D. B. Chang W. K. Carey and S. S. Hermami, Regularity-preserving image interpolation, Proc. IEEE Int. Conf. on Image Proc., pages 1293-1297, 1999.

[15] D. D. Muresan and T. W. Parks, Prediction of image detail, Proc. IEEE Int. Conf. on Image Proc., pages 323-326, 2000.

[16] N. R. Shah and A. Zakhor, Resolution enhancement of color video sequence, IEEE transactions on Image Processing, 6(8):879-885, 1999.

[17] B. C. Tom and A. Katsaggelos, Resolution enhancement of monochrome and color video using motion compensation, IEEE transactions on Image Processing, 2(10):278-287, 2001.

[18] M. Irani and S. Peleg, Improving resolution by image registration, CVGIP:Graph. Models Image Process., 53:231-239, 1991.

[19] Truong Q. Nguyen Karl S. Ni, Color image superesolution based on a stochastic combinational classification-regression algorithm, Proc. IEEE Int. Conf. on Image Proc., pages 89-92, 2007.

[20] Stephane Mallat, Geometrical grouplets, Applied and Computational Harmonic Analysis, 26(2):161-180, 2009.

[21] ITU-T, Recommendation ITU-R BT500-10, Methodology for the subjective assessment of the quality of the television pictures, March 2000.

[22] ITU-T, Recommendation ITU-R P910, Subjective video quality assessment methods for multimedia applications, September1999.

[23] VQEG, VQEG testplans, ftp://vqeg.its.bldrdoc.gov

[24] Z. Wang and A. C. Bovik, A Universal Image Quality Index, IEEE Signal Processing Letters, vol. 9, no. 3, pp. 81-84, March 2002.

[25] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, Apr. 2004.

[26] M. Elad and A. Feuer, Restoration of single super-resolution image from several blurred, noisy and down-sampled measured images, IEEE Transactions on Image Processing, 6(12):1646-1658, 1997.

[27] R. Hardie, K. Barnard, and E. Amstrong, Joint map registration and high-resolution image estimation using a sequence of undersampled images, IEEE Transactions on Image Processing, 6 (12):1621-1633, 1997.

[28] A. J. Patti, M. I. Sezan, and A. M. Tekalp, Superresolution video reconstruction with arbitrary sampling lattices and nonzero aperture time, IEEE Transactions on Image Processing, 6(8):1064-1076, 1997.

Author Biography

Aldo MAALOUF received the B.Sc. degree in Computer Engineering with distinction and the M.Sc. degree in Electrical Engineering with first class honor from the University of Balamand, Tripoli, Lebanon in 2003 and 2005 respectively. He received his PHD in signal and image processing with high distinction from the University of Poitiers in 2008. Currently, he is appointed as post-doc in the XLIM-SIC laboratory (UMR CNRS 6172) at the University of Poitiers where he is working on the optimization of stereoscopic image coding, image super resolution and video surveillance.