Research Journal of Applied Sciences, Engineering and Technology 7(20): 4313-4318, 2014
DOI:10.19026/rjaset.7.803
ISSN: 2040-7459; e-ISSN: 2040-7467
© 2014 Maxwell Scientific Publication Corp.
Submitted: January 01, 2014 Accepted: January 05, 2014 Published: May 20, 2014

Research Article

An Efficient Multi-resolution Video Compression Approach with Inpainting Algorithm

[1]R. Gomathi and [2]A. Vincent Antony Kumar
[1]Department of Electronics and Communication Engineering, Anna University-Dindigul Campus,
[2]Department of Information Technology, PSNA College of Engineering and Technology,
Dindigul 624 622, Tamilnadu, India

Abstract: This study carves out a way to video coding that is motivated by the recent advancement in video inpainting. In the proposed video coding approach, the regions are mainly divided into two types namely, Local Motion Region and Global Motion Region. The regions are removed by using block based motion estimation in Local Motion Regions and the method of texture synthesis is used for obtaining the removed regions. Shearlet *p*-Laplacian based Partial Differential Equation (PDE) inpainting is carried out to recover the removed regions in Global Motion Regions. In both regions, Exemplar Selection (Region Removal) Process is based on the edges extracted from the input region. Wavelets like curvelets and contourlets are not very efficient when dealing with edges in multidimensional signals. In this study, a Discrete Shearlet Transform (DST) is used for edge detection. Using DST helps to improve the overall compression performance together with the Shearlet *p*-Laplacian based inpainting and texture synthesis Schemes. The Scheme has been integrated into H.264/AVC and achieves better bit rate saving.

Keywords: Discrete shearlet transform, inpainting, texture synthesis, video coding

INTRODUCTION

Video compression Luo *et al*. (2009) is the process of reducing the amount of data used to represent video images. The two important parts of video compression are spatial image compression and temporal motion compensation. Video compression belongs to the concept of source coding in Information theory. Video is basically a three-dimensional array of color pixels. Spatial directions of the moving picture are represented by two dimensions and the time domain is represented by one dimension.

Video data contains spatial and temporal redundancy. By eliminating these redundancies, we can achieve video compression. In video coding, motion compensation and transform are used to exploit spatio-temporal redundancy based on the Mean Squared Error (MSE) criterion. But the recent video coding schemes concentrate only on pixel wise redundancy. It is illustrated by the inefficiency of coding texture regions such as fabric and wood. But, texture synthesis (Efros and Leung, 1999) and video inpainting (Bertalmio *et al*., 2000; Rane *et al*., 2003) algorithms proves their efficacy on these textural regions. The above mentioned schemes are also used for other applications such as object removal.

A Ridgelet transform based hybrid video coder (Lorenzo *et al*., 2003; Raymond *et al*., 2009) scheme demonstrates the ability of Ridgelets and results show the considerable improvements when compared to coding methods which uses wavelets only. Several recently proposed methods use the lifting scheme (Gerek and Ctin, 2006; Chang and Gimod, 2006; Velisavljevic *et al*., 2006) in image and video compression algorithms. Nowadays commonly used video coding standards use (e.g., those in standards approved by the ITU-T or ISO) a Discrete Cosine Transform (DCT) for reducing spatial redundancy.

Initially the image is partitioned into number of blocks and each block is subjected with Discrete Cosine Transform (DCT) to develop the correlation between the blocks. This method reduces the difficulty of processing the whole image. At the same time, it induces some of the blocking artifacts. Also, when the DCT is applied to each block, the inter block correlation is not perfectly developed. These drawbacks limit the coding performance.

Wavelets have been used to eliminate all the above mentioned problems. But the directional wavelets such as curvelets, ridgelets and contourlets are not very effective in dealing multidimensional signals containing distributed discontinuities called edges. This drawback is avoided by using the basis elements with higher directional sensitivity and of various shapes. This study uses Discrete Shearlet Transform (DST) (Kutyniok and Labate, 2009; Glenn *et al*., 2008; Gao *et al*., 2009)

Corresponding Author: R. Gomathi, Department of Electronics and Communication Engineering, Anna University-Dindigul Campus, Dindigul 624 622, Tamilnadu, India
This work is licensed under a Creative Commons Attribution 4.0 International License (URL: http://creativecommons.org/licenses/by/4.0/).

which is able to capture the geometry of multidimensional data and gives efficient representation of images with edges.

In this study, we propose a video coding scheme which utilizes the advantages of textural synthesis and the *p*-Laplacian Operator inpainting based on Discrete Shearlet Transform Domain. The scheme is carefully designed towards the video coding. The concept employed is, highly correlated regions present in video frames are removed in the encoder and then restored at the decoder. Some assistant information for example the edge information of removable blocks and motion estimation parameters are extracted efficiently by using DST and used for restoration at the decoder.
The important features of the scheme are:

- For maintaining temporal consistency, the local motion and global motion regions are used in both encoding and decoding processes.
- To perform the decoding process very efficiently the edges are extracted from the correlated regions and they are used as assistant information.
- For extracting edge information, the Discrete Shearlet Transform (DST) is employed to improve the performance.
- In both the encoding and decoding processes, the texture and structural regions are considered.

METHODOLOGY

Design of an encoder:
Segmentation: As shown in Fig. 1, by using a simple segmentation model, both the Local Motion Regions and Global Motion Regions are easily extracted from the given input video frame. The employed segmentation algorithmic steps are described here:

- Obtain the differential image by subtracting the sprite from the input frame.
- Detect the motion-occlusion zones by performing morphological filter.
- Detect the edge pixels using Canny Operator from the raw frame.
- Detected edge pixels are located in the motion zones and it is assumed as an initial segmentation model.
- Segment the foreground objects from the background by filling in scheme to the initial model.

Edge detection using Discrete Shearlet Transform (DST): Obtaining the edge map of Local Motion Region is very essential step in this video coding process. The point singularities are efficiently represented by wavelets. When the images of higher dimensions are considered, the singularities are not effectively represented by wavelets. Therefore, representation of these singularities require more terms in the wavelet representation. This disadvantage is eliminated by using a Discrete Shearlet Transform (DST).

Discrete Shearlet Transform (DST): The Discrete Shearlet Transform (DST) is obtained by sampling the Continuous Shearlet Transform based on scaling, shear and translation parameters *a*, *s*, *t*. Figure 2 shows the construction of DST. The algorithmic procedure is described here. The Laplacian pyramid method (Peter and Edward, 1983) is used for decomposing the input image into low pass image and a high pass image. The high pass filtered image is computed on a pseudo polar grid. The obtained matrix is subjected with band pass filtering. Then the Cartesian sampled values are re-assembled and inverse 2-D FFT is applied to obtain the shearlet transformed image.

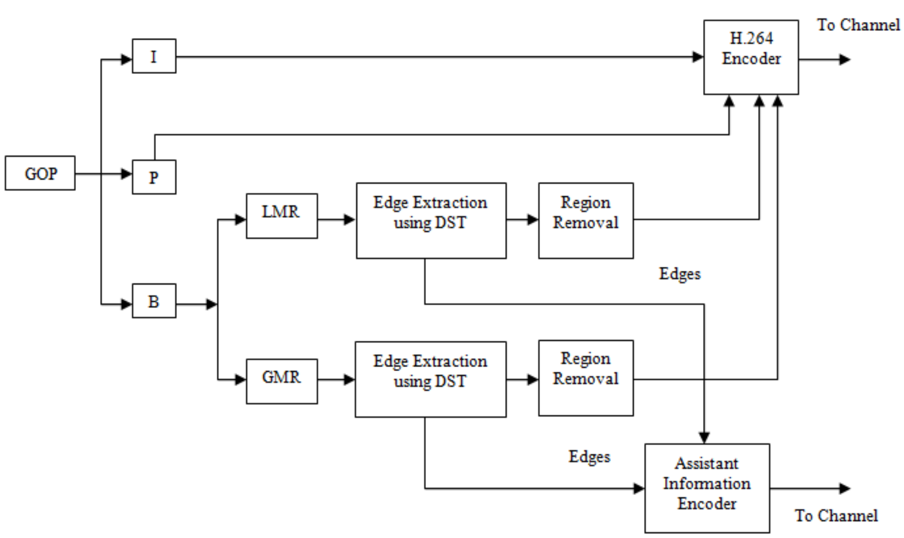

Fig. 1: Block diagram of an encoder

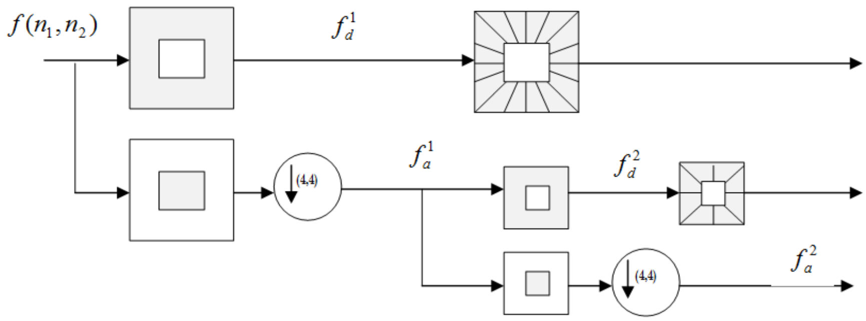

Fig. 2: Succession of laplacian pyramid and directional filtering

Detection of edge orientation: Let *u* be an image containing an edge. The information about the edge points of *u* can be analyzed from the properties of the corresponding wavelet transform. Using the continuous wavelet transform $W_\psi u\ (a,t)$ the orientation of the edges of an image 'u' can be obtained by looking at the horizontal and vertical components of $W_\psi u\ (a,t)$ Letting, $\psi_a = \nabla G_a$ and $\psi_a^y = \frac{\partial G_a}{\partial x}, \psi_a^y = \frac{\partial G_a}{\partial y}$ the edge orientation of *u* at τ is given:

$$\angle(u * \nabla G_a(\tau)) = \arctan\left(\frac{u * \psi_a^y(\tau)}{u * \psi_a^x(\tau)}\right)$$

The above expression measures the direction of the gradient ∇u_a at τ The advantage of the shearlet transform is that, by decomposing an image as function of scale, location and orientation, it allows one to extract directly the information about the orientation of edges.

Region removal in LMR: In a video frame, the pure textures can be easily recovered at decoder but structural regions are very difficult to restore. Therefore the LMR must be divided into two regions namely structural and textural blocks based on the edge information extracted using Discrete Shearlet Transform (DST). The textural blocks are constructed by using the process called block based motion estimation. The exemplars are selected based on spatio temporal variation.

Region removal in GMR: For each Group of Pictures (GOP), the image called sprite is created (Yan *et al*., 2003; Lu *et al*., 2001; Krutz *et al*., 2008). The region identification is carried out by dividing the sprite image into removable and non-removable regions. Then the original frames are formed by mapping. The blocks to be removed in sprite image are indicated by R_s. The R_f blocks to be removed in original frame are indicated by Blocks in R_f are matched with the grid of 8×8 blocks during the exemplar selection module.

In an original frame if a block have more than half of the highly correlated pixels then it is called as a removable block and they are removed during encoding. The remaining regions are considered as exemplars and encoded by using H.264 encoder.

Design of a decoder: The block diagram of decoder is shown in Fig. 3. In the decoder design, two restoration methods are proposed to recover the removed blocks in foreground and back ground regions.

Spatio temporal texture synthesis for LMR: The removed blocks in LMR are restored by texture synthesis. To maintain the temporal consistency, the unidentified textural regions are covered by the available neighborhood pixels.

PDE inpainting algorithm based on *p*-Laplacian operator for GMR: In the decoder side, the removed blocks of GMR are restored by using PDE based *p*-Laplacian operator inpainting algorithm (Dong *et al*., 2007; Guillermo *et al*., 2010; Tony *et al*., 2006). The shearlet domain *p*-Laplacian operator inpainting Model is described as follows:

- Take sprite image as an input image and denote it as $\alpha_{j,l,k}$
- Start with initial guess $\beta_{j,l,k}^{new} = \alpha_{j,l,k} x_{j,l,k}$

where,

$$\chi_{j,l,k} = \begin{Bmatrix} 1; & (j,l,k) \in I \\ 0; & (j,l,k) \in I \end{Bmatrix}$$

I = Inpainting region
$\alpha_{j,l,k}$ = Wavelet coefficients

- Set $\beta_{j,l,k}^{old} = 0$ and the initial error E = $\|\beta^{new} - \beta^{old}\|_2$
- While $E \le \delta$, do
  - Set $\beta^{old} = \beta^{new}$
  - Calculate SH_{curv} by using

$SH_{curv} = FSHT\ (curv)$

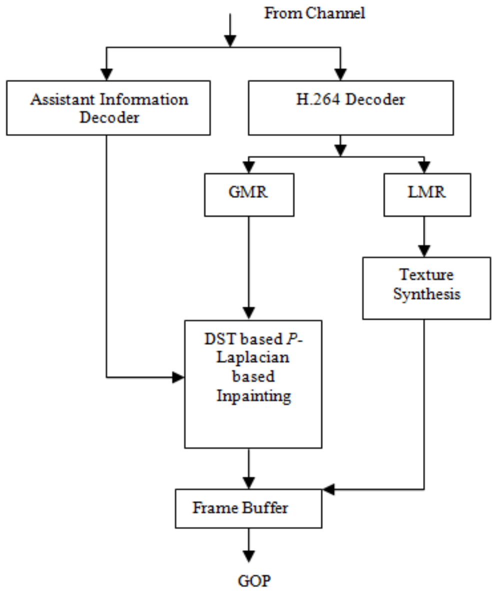

Fig. 3: Block diagram of decoder

where,

$$curv_{i,j} = D_1^- \frac{D_1^+ u_{i,j}}{\left(\left|D_1^+ u_{i,j}\right|^2 + \left|D_2^+ u_{i,j}\right|^2 + \varepsilon\right)^{\frac{2-p}{2}}} + D_2^- \frac{D_2^+ u_{i,j}}{\left(\left|D_1^+ u_{i,j}\right|^2 + \left|D_2^+ u_{i,j}\right|^2 + \varepsilon\right)^{\frac{2-p}{2}}}$$

where, ε is a small positive number which is used to prevent the numerical blow up when denominator becomes zero.

- For all (i, j), update:

$$\beta_{j,l,k}^{new} = \beta_{j,l,k}^{old} + \frac{\Delta_t}{\Delta_x}\left(SH_{curv} - \lambda_{j,l,k}\left(\beta_{j,l,k} - \alpha_{j,l,k}\right)\right)$$

where,
Δ_t = The time step size
Δ_x = The space grid size

- Compute error E = $\|\beta^{new} - \beta^{old}\|_2$ and set $i = i + 1$
- End the while loop
- Obtain a Shearlet based *p*-Laplacian inpainted sprite image

RESULTS AND DISCUSSION

In this section, the performance of the proposed algorithm for video coding with inpainting in Shearlet Domain using *p*-Laplacian is illustrated. The codes are written in MATLAB 2008a.

Figure 4a shows test sprite image 'cricket' and corresponding results of proposed system. In this test, Fig. 4b shows the edges extracted from the input image and Fig. 4c shows the sprite image with missed regions at decoder. Based on the preserved blocks, the Shearlet *p*-Laplacian inpainting gives results in Fig. 4d.

Compared with the restored sprite image shown in Fig. 4e by H.264/AVC, proposed scheme saves 35.79% of bits (Table 1). Furthermore, the system is also tested using the color video sequences.

The visual quality comparisons are shown in Fig. 5 and 6 in which the first row shows proposed scheme results and second row presents H.264 results. From the figures we can easily observed that the visual quality

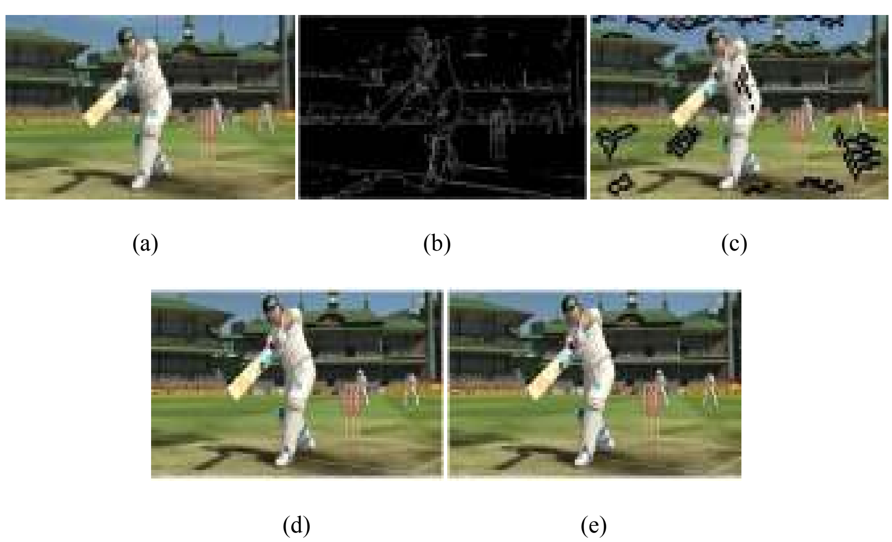

Fig. 4: Comparison with H.264/AVC on test video frame 'cricket', (a) original sprite, (b) edge map, (c) generated sprite with holes (40% removal) at decoder, (d) restored sprite after in painting and texture synthesis, (e) restored sprite by H.264/AVC

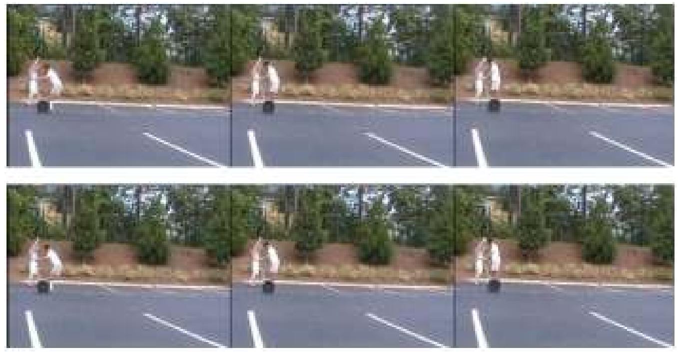

Fig. 5: Comparison with H.264 on ‘stop test’ video sequences. The first row shows the reconstructed video sequences by proposed scheme and the second row shows the restored video sequences by H.264 with QP = 26

Fig. 6: Comparison with H.264 on ‘car’ video sequences. The first row shows the restored video sequences by proposed scheme and the second row shows the restored video sequences by H.264 with QP = 26

Table 1: Bit rate savings of proposed scheme compared to H.264

Original image	H.264	Proposed scheme	Bit-rate savings compared with H.264 (%)
Forman	0.985	0.6850	30.00
Stefan	0.993	0.6180	37.50
City	1.311	1.0860	22.50
Cricket	0.948	0.5901	35.79
Bird	0.710	0.5888	12.12
Stop test	0.876	0.6380	23.80
Car	1.354	0.6399	71.41
Worker	1.246	0.6082	63.78
Highway	0.823	0.4526	37.04
Class	0.622	0.5584	6.36

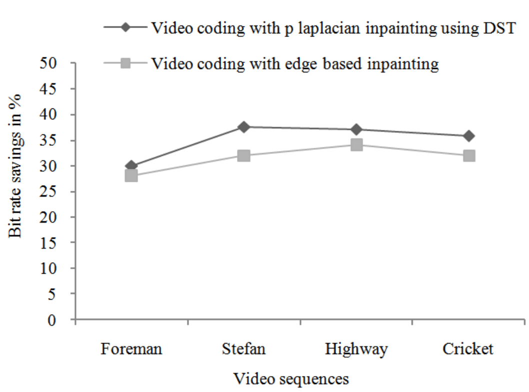

Fig. 7: Plot of bit-rate saving of proposed scheme (40% removal) vs. video sequences with QP = 26

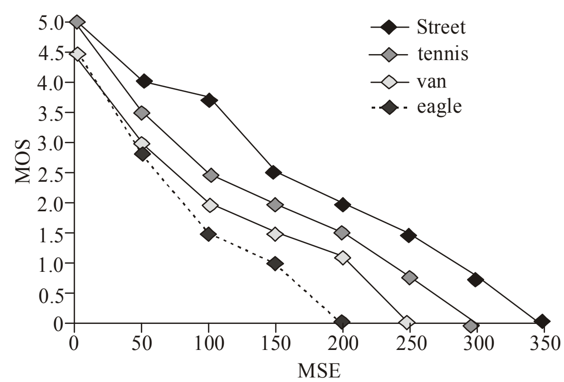

Fig. 8: Plot of MSE vs. MOS

of obtained video sequences is same as that of H.264. The bits-saving of proposed system is indicated in Table 1.

Proposed method using Shearlet based p-Laplacian inpainting at decoder averagely saves 34.03% bits for the video sequences at the similar visual quality levels. Figure 7 shows Bit rate saving vs. video sequences curves of four testing sequences. Up to 37.50% bit rate saving can be obtained for sequence “Stefan”.

The plot of Mean Square Error (MSE) versus Mean Opinion Score (MOS) is shown in Fig. 8 for four video sequences namely Street, Tennis, Van, Eagle. From the

Fig. 8, we can easily observe that the curves are linear in shape from the MOS value of 5.0 to the MOS value of 1.0. Here, the value 1 represents ‘bad’ and the value 5 represents ‘excellent’. Therefore the quality of picture is very poor at low bit rate i.e., below the MOS value of 1.0.

CONCLUSION

A new method for compressing video frames, which selects Partial Differential Equation (PDE) based Shearlet inpainting to remove spatial redundancy is introduced in this study. The assistant information called edges is extracted from video frames and highly correlated regions are intentionally skipped during encoding. The remaining areas along with information associated with the edges are encoded to form the compressed output data. Removed information is to be recovered with the assistance of information sent to the decoder side. At the decoder, using a PDE-based inpainting algorithm, the removed areas are recovered after the remaining regions and edges are decoded.

The proposed scheme is tested with number of color video sequences. From the Bit-rate saving results, the ability of the proposed scheme is justified. When comparing the proposed scheme with H.264, up to 71% bits-saving can be acquired. Numerical results show state-of-the-art performance and speed.

Further improvements of the current scheme are still promising. Edge extraction can be more flexible and adaptable for compression. Finding the regions that can be eliminated is considered to be an open problem and it seems that solving this problem will lead to increase in Bit-rate saving values and output quality.

REFERENCES

Bertalmio, M., G. Sapiro, V. Caselles and C. Ballester 2000. Image inpainting. Proceedings of ACM SIGGRAPH, pp: 417-424.

Chang, C.L. and B. Gimod, 2006. Direction adaptive discrete wavelet transform via directional lifting and bandletization. Proceedings of IEEE International Conference on Image Processing. Atlanta, GA.

Dong, L., S. Xiaoyan, W. Feng, L. Shipeng and Z. Ya Qin, 2007. Image compression with edge based in painting. IEEE T. Circ. Syst. Vid., 17(10): 1273-1286.

Efros, A. and T. Leung, 1999. Texture synthesis by non-parametric sampling. Proceedings of International Conference on Computer Vision, pp: 1033-1038.

Gao, K., D. Labate and W. Lim 2009. Edge analysis and identification using the continuous shear let transform. Appl. Comput. Harmon. A., 27(1): 24-46.

Gerek, O.N. and A.E. Ctin, 2006. A 2-D Orientation adaptive prediction filter in lifting structures for image coding. IEEE T. Image Process., 15(1): 106-111.

Glenn, E., L. Demetrio and Q.L. Wang, 2008. Sparse directional image representation using discrete shear let transform. Appl. Comput. Harmon. A., 25(1): 25-46.

Guillermo, S., B. Aurelie, B. Marcelo and C. Vicent, 2010. A comprehensive framework for image in painting. IEEE T. Image Process., 19(10): 2634-2645.

Krutz, A., A. Glants, M. Haller and M. Droese, 2008. Multiple Background sprite generation using camera motion characterization for object-based video coding. Proceeding of 3DTV Conference: The True Vision-capture, Transmission and Display of 3D Video, pp: 313-316.

Kutyniok, G. and D. Labate, 2009. Resolution of the wavefront set using continuous shear lets. T. Am. Math. Soc., 361: 2719-2754.

Lorenzo, G., M. Fulvio and V. Piere, 2003. Ridgelet transform applied to motion compensated images. Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing. Hong Kong, pp: 6-10.

Lu, Y., G. Wen and W. Feng, 2001. Fast and robust sprite generation for MPEG-4 video coding, Proceedings of the 2nd IEEE Pacific Rim Conference on Multimedia: Advances in Multimedia Information Processing, pp: 118-125.

Luo, L., F. Wu, S. Li, Z. Xiong and Z.Q. Zhuang, 2009. Advanced motion threading for 3D-wavelet video coding, signal processing. Image Commun., 19(7): 601-616.

Peter, J.B. and H.A. Edward, 1983. The lapalcian pyramid as a compact image code. IEEE T. Commun., 31(4): 532-540.

Rane, S., G. Sapiro and M. Bertalmio, 2003. Structure and Texture filling-in of missing image blocks in wireless transmission and compression applications. IEEE T. Image Process., 12(3): 296-303.

Raymond, H.C., Y.W. Wen and A.M. Yip, 2009. A Fast optimization transfer algorithm for image in painting in wavelet domain. IEEE T. Image Process., 18(7): 1467-1476.

Tony, F.C., S. Jianhong and M.Z. Hao, 2006. Total variation wavelet in painting. J. Math. Imaging Vis., 25(1): 107-125.

Velisavljevic, V., B. Beferull-Lozano, M. Vitterli and P.L. Dragotti, 2006. Direction lets: Anisotropic multidirectional representation with separable filtering. IEEE T. Image Process., 15: 1916-1933.

Yan, L., G. Wen and W. Feng 2003. Efficient background video coding with static sprite generation and arbitrary shape spatial prediction techniques. IEEE T. Circ. Syst. Vid., 13(5): 394-405.