

Image Inpainting Approaches – A Review

Rohit T. Pushpalwar, Smriti H. Bhandari
 Department of Computer Science and Engineering
 Walchand College of Engineering,
 Sangli, India
 rohit.pushpalwar@walchandsangli.ac.in
 smriti.bhandari@walchandsangli.ac.in

Abstract—Image inpainting is an ancient art. There may be different factors which cause deterioration of images including environmental factors, chemical processing, improper storage and more. Inpainting is the process which restores the deteriorated parts of the image. This paper proposes two major types of inpainting approaches i.e. 2D image inpainting and depth map inpainting. Image inpainting refers to recover corrupted parts of images by applying structural, textural or both the methods simultaneously. Whereas, depth map inpainting mainly used to improve the 3D visualization effect by improving the depth map associated with scene. Some of the image inpainting approaches can also be applied to depth maps. This paper gives overview of such approaches as well as those which are specifically meant for depth maps.

Keywords—*inpainting, depth map.*

I. INTRODUCTION

Image inpainting refers to reconstruction of deteriorated parts in images and videos. This has to be done in an undetectable way. Image inpainting can be used to reconstruct damaged paintings or photographs and removal or replacement of selected objects. 3D imaging has been the most emerging technology in recent era. Mechatronics and robotics areas are influenced by Microsoft's kinect sensor in last few years due to recent advancement in sensors. These sensors have many applications in video surveillance, healthcare and communication.

3D images or videos can be generated with the help of depth maps. Due to some limitations of the sensors we may not get the exact depth map of scene. So, inpainting of depth maps is also an important issue. Disocclusion is the process of recovering scene information obstructed by visible parts. Disocclusion is also considered a major part in image and depth map inpainting.

Damaged and occluded parts which we want to restore in image are generally called as holes. Bertalmio et al. [1] have given a solution to inpaint such holes in an image by formulating partial differential equation that propagates information in the direction of isophotes. This process is called as digital image inpainting. Research in the area of image inpainting mainly focuses on connecting boundary pixels in the area of holes.

In 3D imaging, depth map plays an important role, since it contains information related to depth between objects in the scene. Due to some limitation of sensors or deflection of infrared light from certain objects in the scene like glass

windows gives missing values in depth maps. These are also termed as holes. These holes should also be filled with proper depth values to get a better view of the scene in 3D. Following figure 1 [1] and 2 [2] show 2D image inpainting and depth map inpainting respectively.

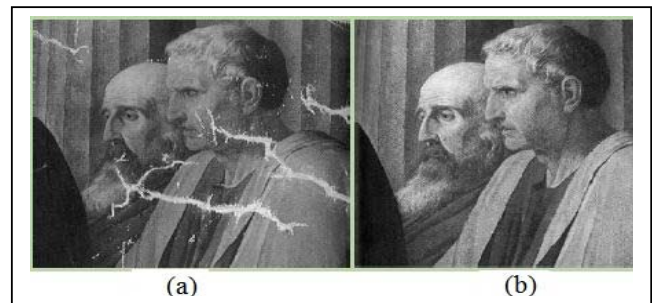


Fig. 1. 2D image inpainting. (a) Distorted 2D image. (b) 2D image after image inpainting.

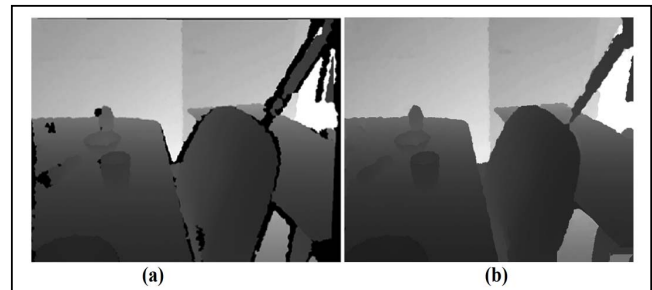


Fig. 2. Depth map inpainting. (a) Depth image with missing depth values denoted by black color. (b) Depth image after depth map inpainting.

This paper basically reviews various techniques for 2D image inpainting and depth map inpainting. 2D image inpainting is referred in this paper as “image inpainting”. Section II gives an overall overview of inpainting techniques. Section III provides discussion related to various categories of image inpainting whereas section IV focuses on depth map inpainting with respect to 3D imaging. Section V is provided as summary of the work discussed in this paper. Conclusion is given in section VI.

II. OVERVIEW

Image inpainting tasks include object disocclusion, image restoration and filling holes etc. The notion of digital inpainting was first introduced by Bertalmio et al.[1]. They used higher order partial differential equation (the Laplacian of the image) for restoration purpose. The algorithm proposes that the gradient direction gives the direction of the region to be filled. The two main categories of inpainting include structural and textural inpainting. Image compression can be effectively done by skipping some of the regions at the encoder side and inpainting those regions by same approach at decoder side [3]. Morphological operation like erosion can also be used to inpaint small regions of missing values [4]. Cheng-Shian Lin and Jin-Jang Leou suggested a four step approach for inpainting, in which salient structures and its surrounding regions are inpainted first and then remaining missing regions are inpainted [5].

To inpaint large missing regions Sung Ha Kang et al. proposed an approach with the help of global information from multiple images [6]. Anat Levin et al. used training images to construct exponential family distribution over images which will be used to inpaint the missing regions [7]. Another approach suggested by Tijana Ruzic and Aleksandra Piurica is context aware patch based inpainting in which textural descriptors are used to match the patches. They applied this general approach to global image inpainting by using Markov Random Field Modeling [8].

Marcelo Bertalmio et al. suggested an approach which inpaint missing regions with simultaneous structure as well as texture [9]. The idea of using both structural and textural inpainting together is extended by Shantanu D. Rane, Guillermo Sapiro, and Marcelo Bertalmio for filling-in of holes in wireless compression applications and transmission. The main task in this approach is to decompose the image into two functions based on certain characteristics and then work on these two functions with structure and texture based filling-in algorithms [10].

Low cost devices such as Microsoft Kinect sensor have influenced many researchers to deal with depth data. The main issue with this device is its accuracy and resolution [11]. Shaoguo Liu et al. proposed a depth inpainting technique with energy minimization function [12]. Disocclusion is the main problem which causes holes in the depth map. Another approach to fill-in missing values in depth video is to use the knowledge of similarity of neighboring pixels [13]. Fei Qi et al. have combined the strategy used for color image inpainting with that of non-filtering schemes to generate inpainted depth maps [14]. Kinect sensor uses infrared sensor to determine depth of the scene. Some objects like glass windows distract infrared light which causes holes in depth map. One approach has been proposed to inpaint these holes based on background estimates of unoccluded scene [15]. Weihai Chen et al. have used edge information from both color as well as depth images to guide the process of depth map inpainting with the help of neighboring pixels [2]. Fig. 3 shows the inpainting classification discussed in this paper.

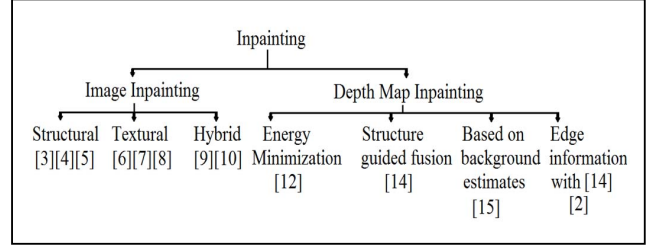


Fig. 3. Inpainting Classification

Inpainting is an ancient art. It is being carried out traditionally by professional artists. Inpainting deals with set of techniques to make unrecognizable modification in the images. With these techniques we can remove an object or rebuild distorted portion of the scene.

III. IMAGE INPAINTING

A. Structure based image inpainting

Chen wang et al. proposed a method of structure aware inpainting which is used for image compression at the decoder side [3]. In this approach, structure is propagated in the missing region and then texture synthesis is performed at decoder side. Distance-based inpainting method is used for structure propagation. The unknown and known region stands for ω and $\bar{\omega}$ respectively. Two structures \hat{S}_1 and \hat{S}_2 adjacent to structure S in ω are found. Two windows consisting of set of pixels from \hat{S}_1 and \hat{S}_2 are derived and denoted as \hat{W}_1 and \hat{W}_2 . Widths of \hat{W}_1 and \hat{W}_2 are given as $(2\hat{T}_1 + 1)$ and $(2\hat{T}_2 + 1)$ respectively. A set W is formed consisting of pixels having Euclidean distance $D(p) \leq T$ where $T \equiv \max \{\hat{T}_1, \hat{T}_2\}$. All pixels in $W \equiv \hat{W}_1 \cup \hat{W}_2 \cup W$ are classified into three structure info as $S \equiv \hat{S}_1 \cup \hat{S}_2 \cup S : L, R$ and S itself. If $p_k \in L$ and $D(p_k) = n$ then p_k can be filled as follows:

$$f(v_1, v_2) = a \cdot v_1 + (1 - a) \cdot v_2 \quad (1)$$

Where $v_1 \in V_{L_1}^n$, $v_2 \in V_{L_2}^n$ and a is a linear combination factor formulized as $d = d_2 / (d_1 + d_2)$, here d_1 and d_2 are distance of $p_k \in L$ from \hat{W}_1 and \hat{W}_2 . Due to this, texture is also preserved while inpainting.

Morphological erosion operation can also be used for inpainting of unknown regions in an image with structure guided approach [4]. Erosion is one of the crucial morphological operations, which is used for removing pixels around the boundaries of an object. The erosion of the binary image A by the structuring element B is defined by:

$$A \ominus B = \{x \in A \mid (B_x \in A)\} \quad (2)$$

Where (B_x) denotes the structuring element B centered at pixel x . The structuring element is chosen from the missing region of the area to be inpainted. So after successive erosions the missing region is eroded.

This erosion process can be treated as an inpainting process. Since a thin line is maintained between missing

region and known region, the erosion process is suited for inpainting. The missing region is denoted as Ω and boundary of the missing region at time t is denoted as $\partial\Omega^t$. At each successive erosion, the boundary of Ω reduces. This process of successive erosions is carried out until missing region is completely restored.

Another approach given by Cheng-Shian Lin and Jin-Jang Leou which uses salient structure of image to inpaint comparatively large missing region [5]. The overall method is divided into four steps as follows:

- 1) Salient structure(s) detection .
- 2) Salient structure(s) inpainting.
- 3) Inpainting of neighboring area of salient structure(s).
- 4) Inpainting of remaining unknown parts.

B. Texture based image inpainting

Structure based inpainting are suitable mainly for missing regions of small areas which propagate neighboring information into the missing region. Hence to inpaint large missing areas, texture information is used. Sung Ha Kang et al [6] presented an approach to inpaint large missing regions in an image with the help of global information from multiple images. This inpainting approach has following three phases:

- 1) Extraction of landmarks and matching.
- 2) Interpolating directional information.
- 3) Copy information from one image to another.

If the inpainting shows some irregularities in the intensity levels in the modified image then local inpainting techniques can also be used along with the interpolation to improve the final result.

Most techniques which uses global information for inpainting are related to textural inpainting. Anat Levin et al. proposed one such technique which is based on histograms of local features and uses exponential family distribution over images [7]. Probability of image is given by small number of sufficient features which are taken from arbitrary location from image. The probability of image is given as:

$$P(I; \{\Phi_i\}) = \frac{1}{Z} e^{\sum_i \Sigma_{x,y} \Phi_i(f_i(x,y))} \quad (3)$$

where, $f_i(x, y)$ is the value of feature i at location (x, y) in the image I and Z is a normalization factor. Based on the feature values to $\Phi_i(f)$, the probability of the image changes i.e. average value of $\Phi_i(f)$ for a particular image is directly proportional to probability of image. So inpainting requirement is that predicted marginal of all features matches with empirical marginal in the data.

$$P(F_i = s; \{\Phi\}) = \hat{P}(F_i = s) \quad (4)$$

Optimization is also done on the probability, subject to boundary of missing area. Hence, to inpaint missing values in the image histograms with the features selected are measured over the training images and an integrable gradient field is searched that matches with the image gradients on the boundary of the missing values and escalate the probability.

Patch based approaches are very popular in the field of inpainting. One of such approach is context aware patch based image inpainting [8]. This technique uses textural descriptors to guide and speedup the process of matching patches. This method is used with global inpainting technique with Markov random field (MRF). The main concept of context aware method is to restrict search of patch selection to the region of concern based on contextual features. The method can be applied to fixed size blocks when missing pixels in a block are not dominant. When missing pixels are more than half pixels in a block, adaptive blocks sizes should be used. Adaptive sizes of blocks can be formed with the help of simple top-down splitting procedure based on the ‘‘homogeneity’’ of texture.

C. Hybrid: Combination of Structure and Texture based

The algorithms proposed in subsection *A* and *B* works for pure structure or pure texture. But for images which contains structural and textural part in significant amount these techniques fail to produce acceptable inpainting result. Marcelo Bertalmio et al proposed an algorithm to use both structural and textural inpainting on images of missing regions [9]. The main idea of using this technique is to decompose the main image into two images, one containing the image structure and other containing image texture. Both of these images are inpainted using their respective methods and the result is the combination of these two processed images. The algorithm proposed in [9] uses three main phases.

- 1) Image decomposition
- 2) Texture synthesis
- 3) Structure inpainting

The two main parts of decomposition are the space of oscillating functions to model texture and noise and total variation minimization for image restoration and denoising. Texture template is used to fill the missing texture and for structure inpainting. The idea is to propagate information coming from boundary of missing region in the direction of isophotes.

Most images are compressed in the form of JPEG, so that it can be used for transmission throughout the internet. JPEG divides image into blocks of 8×8 and calculates a two dimensional (2-D) discrete cosine transform (DCT). When these images are transmitted over wireless channel, some of these blocks may get corrupted. Forward error correction (FEC) and automatic retransmission query protocols (ARQ) are available for resolving this issue but requires some extra overhead. Shantanu D. Rane et al. [10] proposed an approach of applying structure and texture inpainting for lost data. This technique is also used for compression along with JPEG to improve the compression ratio.

The first step in algorithm is to classify lost blocks into structure or texture. This decision can be taken by querying the information surrounding lost blocks. Coarseness measure given by number of local extrema in the neighborhood of lost block. The number of local extrema in window of size $s \times s$ is given as:

$$n = s^2 \frac{(UB + LB)}{2} \quad (5)$$

Where UB and LB are respectively the upper and lower bounds for texture coarseness. Thus classification of a block depends on n . Thus if a block has fewer than n extrema, it is classified to have structure, else is considered to contain texture. The filling of structure and texture information is inspired from the algorithm proposed in [9].

IV. DEPTH MAP INPAINTING

The visual appearance of data objects and their relationships with other objects have always been an important issue. To generate 3D environment, multiple cameras were used at earlier stages. Advancement in technology give rise to new devices which can generate 3D visualization of the captured scene. Still there are some problems associated with it. Recent improvements in computers and sensors have encouraged the operation of computer systems in increasingly complex functions and situations, which deals with human computer interaction. The suitable option for scenarios like this is the optical sensors since they grant for a contactless, non-interfering sensing of the environment.

Video surveillance, mechatronics, robotics etc. are some of the areas on which Microsoft kinect sensor has maximum impact. Depth map provided by kinect often contain some missing values (holes) in it. These holes should be filled with appropriate values. This technique of filling in missing values in depth map is called as depth map inpainting. There are many techniques proposed for depth map inpainting. This paper covers some of these techniques.

Shaoguo Liu et al.[12] proposed a technique of energy minimization for depth map recovery with TV_{21} regularization. The method proposed in this technique is divided into three major components which are as follows.

- 1) Energy minimization for depth inpainting
- 2) Adding TV21 prior
- 3) Alternating energy minimization

(1) *Energy minimization for depth inpainting*: Kinect device takes both depth map as well as color image of the scene at the same time. Since these two descriptors are of the same environment, there exist a firm interaction between them. Hence within a local window N_k centered at pixel k , the depth values r_i and pixel color colors c_i exhibits following linear regression model:

$$r_i = w_k^T c_i + b_k, \forall i \in N_k \quad (6)$$

where $w_k = (w_{k1}, w_{k2}, w_{k3})^T \in \mathbb{R}$ are the linear regression coefficients. Based on this model w ; b ; r should be found out that minimize the following energy function:

$$J(w, b, r) = \int_{k \in C} \left(\int_{i \in N_k} G_\Phi(k, i) |w_k^T c_i + b_k - r_i|^2 di + \lambda_1 |w_k|^2 \right) dk \quad (7)$$

Where Gaussian kernel is G with width Φ , which gives the contribution in the neighborhood for each pixel. To avoid over-fitting $\lambda_1 |w_k|^2$ is a used. Hole filling in depth map can be accomplished by adding a condition in above equation (7) as:

$$\min J(w, b, r) \text{ s.t. } r_i = S_i, \forall i \in S \quad (8)$$

Where, known depth values are given by set S . Minimizing above energy function gives w_k and b_k for every unknown pixel k , this can be used to compute its depth value from equation (6).

(2) *Adding TV_{21} prior*: To produce sharp edges in inpainted regions and remove noise, TV_{21} prior is used. TV_{21} prior is given on depth maps in its continuous form as:

$$TV_{21}(r) = \int \left(\int r_x^2(x, y) dy \right)^{\frac{1}{2}} dx + \int \left(\int r_y^2(x, y) dx \right)^{\frac{1}{2}} dy \quad (9)$$

Gradient is sparse both horizontally and vertically to resolve the two issues stated above. For depth restoration the entire energy function is given as:

$$\min J(w, b, r) + \lambda_3 TV_{21}(r) \text{ s.t. } r_i = S_i, \forall i \in S \quad (10)$$

Where λ_3 is the weight of the TV_{21} prior.

(3) *Alternating energy minimization*: Constrained energy function mentioned in Eqn. 10 can be minimized by using Lagrange Multiplier as:

$$\min \int_{k \in C} \left(\int_{i \in N_k} G(k, i) |w_k^T c_i + b_k - r_i|^2 di + \lambda_1 |w_k|^2 \right) dk + \int \lambda_i |r_i - s_i|^2 di + \lambda_3 TV_{21}(r) \quad (11)$$

The three unknown variables are w , b , r are calculated by two step minimization procedure. In first step w and b is calculated by differentiating Eqn. 11 with respect to w and b and fixing r . In second step r is calculated by differentiating Eqn. 11 with respect to r and fixing w and b . These two steps are performed iteratively till the difference between two successive inpainted results is less than the predefined threshold.

Another approach suggested by Fei Qi et al. [14] integrates inpainting strategy used for color images with that of recently developed non-filtering scheme. The proposed method for inpainting the pixels in missing region is given by following Eqn. 12.

$$d(p) = \sum_{q \in N_p} w(p, q) [d(p) + \langle \Delta d(p), p - q \rangle] \quad (12)$$

Where, N_p is a set of valid depth pixels near p . To inpaint color images, diffusion gives better result but since depth images are less textural and hence propagating depth information from outside to inside of region to be inpainted gives better result. The problem to determine where to stop the propagation is given by weighting function which depends on

geometrical distance, depth similarity, and the structure information provided by the color image.

To achieve a real-time inpainting of holes in depth map, Martin Stommel et al. [15] proposed an approach which incorporate low complex and efficient inpainting method with a depth based background estimator. The proposed method uses pyramid based inpainting method which gives a fast inference of a primary, dense disparity map. Then non-linear mapping of the depth map to a disparity map is carried out to normalize noise. With the help of background estimates the disparity map is updated. To inpaint small regions of missing values selective 3×3 median filter is used, whereas to inpaint regions of larger missing values pyramid based inpainting is used. Background estimates are used in the regions where foreground objects cover region of missing values.

A similar approach to that of given in [14] is proposed by Weihai Chen et al. using improved edge detection algorithm [2]. The idea behind using edge information while inpainting is that the regions where missing values in depth maps are observed are around object boundaries. Hence, in proposed method the approach given in [14] is modified according to edge information. First, edge detection is carried out in both color and depth image. Then the edges extracted from color image are optimized using depth data, while the edges extracted from depth map are optimized using color data. Then the optimized color-edges and depth edges are fused together to get a final edge extraction. Finally, depth map inpainting is done with the help of extracted edges from depth fusion. The difference between [14] and [2] is that not only depth values of pixels are considered, but the edges extracted from both the color image and the depth map are considered as well. Only the pixels that are on the same side of the curve represented by extracted edges with valid depth values are added to set N_p . This is due to the fact that the pixels which are taken into consideration to calculate depth of missing pixel are nearer and belongs to same object to that of missing pixel. In this way, the pixels are removed from N_p if they are not belonging to the same object with valid depth value. Therefore the prediction of the depth value of missing pixel p can be more accurate.

V. SUMMARY

Following tables I, II, III and IV give brief analysis of reviewed work.

TABLE I. IMAGE INPAINTING – APPROACH: STRUCTURE BASED

Ref.	Methodology Used	Advantages	Disadvantages
[3]	Distance based inpainting	Can be used for Compression	Region classification unknown
[4]	Erosion based	Preserves Isophotes continuity	Not suitable for complex structures
[5]	Multiscale salient structure propagation	Both structure and texture are preserved	Complex

TABLE II. IMAGE INPAINTING – APPROACH: TEXTURE BASED

Ref.	Methodology Used	Advantages	Disadvantages
[6]	Landmark matching, interpolation, and copying.	Work well with large inpainting domain	Identifying feature points is difficult
[7]	Exponential Model of image statistics	Works well on sharp corners as well as curves	Poor job preserving texture
[8]	Patch based and MRF	Speed	Limited to images, not for videos

TABLE III. IMAGE INPAINTING – APPROACH: HYBRID

Ref.	Methodology Used	Advantages	Disadvantages
[9]	Image decomposition with inpainting and texture synthesis	Work even in the areas where variations in texture is more	Limited to images, not for videos
[10]	Synthesis of texture and inpainting of structure	Used in wireless transmission to save retransmissions of lost blocks of image	Used only with monochrome images

TABLE IV. DEPTH MAP INPAINTING

Ref.	Methodology Used	Dataset Used	Advantages	Disadvantages
[12]	Energy minimization with regularization	Middlebury	Achieve smooth depth maps with sharp boundaries, robust to noise	Not applied for videos
[14]	Fusion based inpainting	-	Mean Absolute Error very less (20mm)	Does not perform extrapolation
[15]	Selective median filter and background estimator	Kinect recorded video frames	Avoids blurring effect	Not tested for time of flight (TOF) sensors
[2]	Improved edge detection algorithm and fusion based inpainting	RGB-D object dataset, online dataset	Performs better around object boundaries	Comparatively slower

VI. CONCLUSION

This paper provides a brief overview of different categories of inpainting approaches and their analysis.

Digital image inpainting has been widely used in various applications such as compression, wireless transmission and improving quality of image by restoring damaged parts of the image. Current structure based inpainting methods are complex but give accurate results by preserving structure of missing regions after inpainting. Texture based inpainting approaches mainly use patch based technique to fill the missing texture in image, but fail to preserve the structure. To overcome this limitation combined structure and texture inpainting techniques can be used with the cost of complexity.

There are no current methods that can reliably inpaint large holes in depth map. But some of the promising techniques are explained in this paper which inpaint missing depth values in depth map captured from Microsoft kinect like devices. Depth inpainting still has long way to go before it can be applied to critical systems like healthcare.

REFERENCES

- [1] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image Inpainting," in *Comput. Graph. (SIGGRAPH 2000)*, July 2000, pp.417-424
- [2] W. Chen, H. Yue, J. Wang, X. Wu, "An improved edge detection algorithm for depth map inpainting," *Optics and Lasers in Engineering*, vol. 55, 2014, pp. 69-77
- [3] Chen Wang, Xiaoyan Sun, Feng Wu and Hongkai Xiong, "Image Compression with Structure Aware Inpainting," *IEEE International Synopsium, ISCAS*, 2006, pp. 4
- [4] Hao Guo, N. Ono, S. Sagayama, "A structure-synthesis image inpainting algorithm based on morphological erosion operation," *Congress on Image and Signal Processing*, , vol. 3, 2008, pp. 530-535
- [5] Cheng-Shian Lin, Jin-Jang Leou, "Image Inpainting Using Multiscale Salient Structure Propagation," *International Conference on Multimedia and Signal Processing*, vol. 1, pp. 201-204
- [6] Sung Ha Kang, Chan, T. Chan, S. Soatto, "Inpainting from Multiple Views," *First International Symposium on 3D Data Processing Visualization and Transmission*, 2002, pp. 622-625.
- [7] A. Levin, A. Zomet, Y. Weiss, "Learning How to Inpaint from Global Image Statistics," *Ninth IEEE International Conference on Computer Vision*, vol. 1, 2003, pp. 305-312.
- [8] T. Ruzic, A. Pizurica, "Context-Aware Patch-Based Image Inpainting Using Markov Random Field Modeling," *IEEE Transactions on Image Processing*, vol. 24, 2015, pp. 444-456
- [9] M. Bertalmio, L. Vese, G. Sapiro and S. Osher, "Simultaneous Structure and Texture Image Inpainting," *IEEE Transaction on Image Processing*, vol. 12, no. 8, August 2003, pp. 882-889.
- [10] S. Rane, G. Sapiro, and M. Bertalmio, "Structure and Texture Filling-In of Missing Image Blocks in Wireless Transmission and Compression Applications," *IEEE Transaction on Image Processing*, vol. 12, no. 3, March 2003, pp. 296-303.
- [11] K. Khoshelham and S. Elberink, "Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications," *Sensors*, vol. 12, 2012, pp. 1437-1454
- [12] S. Liu, Y. Wang, J. Wang, H. Wang, J. Zhang, C. Pan "Kinect Depth Restoration via Energy Minimization with TV_{21} Regularization," *20th IEEE International Conference on Image processing*, 2013, pp.724-724.
- [13] D. Zhang, Y. Yao, D. Zang, Y. Chen, "A Spatio Temporal Inpainting Method for Kinect Depth Video, " *IEEE International Conference on Signal and Image Processing Applications*, 2013, pp. 67-70.
- [14] Fei Qi, J. Han, P. Wang, G. Shi, Fu Li, "Structure guided fusion for depth map inpainting," *Pattern Recognition Letters*, vol. 34, January 2013, pp. 70-76
- [15] M. Stommel, M. Beetz and W. Xu, "Inpainting of Missing Values in the Kinect Sensor's Depth Maps Based on Background Estimates," *IEEE Sensors Journal*, vol. 14, no. 4, April 2014, pp. 1107-1116.