IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 8, August 2017

Asymmetric Stereoscopic Video Super-Resolution and Depth Estimation

Ashwini Mahadik[1], Prof. R. P. Patil[2], Prof. Dr. S. K. Shah[3]

M.E (E&TC), SKNCOE, Vadgaon, Pune, Maharashtra, India [1]

Asst. Professor (E&TC Department), SKNCOE, University of Pune, Pune, Maharashtra, India[2]

Head of Dept, Dept. of ENTC Engineering, SKNCOE, University of Pune, Pune, Maharashtra, India[3]

Abstract: Recreation of full-resolution stereoscopic video from a asymmetric stereoscopic video is a challenging assignment. The existing techniques accept that the depth information is available, which forces an additional challenge in data acquisition. In this paper, we propose a novel plan that is fit for acquiring super-resolution and depth estimation at the same time from an asymmetric stereoscopic video. The proposed plot models the video super-resolution and stereo matching with a unified energy function. At that point, we apply alternating optimization technique to minimize this energy function, which can be implemented with a two-step algorithm. In the initial step, initial depth map is calculated using region based co-operative optimization technique to minimize this energy function, which can be implemented with two step function. In the second step, the super-resolution issue under the direction of the depth information is resolved. It is powerful on the grounds that each step provides extra benefit over the past step. These two steps are iteratively updated until stable depth and super-resolution results are acquired. Series of experiments are conducted on open stereoscopic video arrangements to assess the performance of the proposed technique. Based on comparison of objective indices and subjective visual results, it confirms that proposed scheme can achieve satisfactory super-resolution results and high-quality depth map simultaneously. Specifically, the subjective assessment experiment on 3D screen shows that this technique outperforms others and accomplishes the best visual sharpness.

Keywords: Asymmetric Stereoscopic Video, Super Resolution, Depth Estimation, Stereo Matching.

I. INTRODUCTION

Multi-frame super resolution, namely estimating the high-resolution frames from a low-resolution sequence, is one of the fundamental problems in computer vision and has been extensively studied for decades. The problem becomes particularly interesting as high-definition devices such as HDTV's dominate the market. There is a great need for converting low-resolution, low-quality videos into high-resolution, noise-free videos that can be pleasantly viewed on HDTV's. Although a lot of progress has been made in the past 30 years, super resolving real-world video sequences still remains an open problem. Most of the previous work assumes that the underlying motion has a simple parametric form, and/or that the blur kernel and noise levels are known. But in reality, the motion of objects and cameras can be arbitrary, the video may be contaminated with noise of unknown level, and motion blur and point spread functions can lead to an unknown blur kernel.

Therefore, a practical super resolution system should simultaneously estimate optical flow, noise level and blur kernel in addition to reconstructing the high-res frames. As each of these problems has been well studied in computer vision, it is natural to combine all these components in a single framework without making oversimplified assumptions.

B. Problem Definition

The method is used to reconstruct the low-resolution view while simultaneously performing stereo matching in a unified energy minimization framework. It provide unified energy function model and minimize it using optimization technique. The high-frequency information of the left full-resolution view for enhancing the resolution of the right view. After the unified energy function is obtained it perform objective experiments to recover the full resolution stereoscopic video with high PSNR and SSIM.

C. Motivation

Previously used methods deals with video super resolution and depth estimation separately which requires more energy. Algorithm proposed in this work reconstructs the low-resolution view while simultaneously performing stereo matching in a unified energy minimization framework. The proposed algorithm can recover the full resolution stereoscopic video with high PSNR and SSIM. Also it provides a high-quality depth map, which can be applied in many video applications, such as multiview video coding and video semantic segmentation.

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 8, August 2017

D. Objectives

- To reconstruct the low-resolution view while simultaneously performing stereo matching in a unified energy minimization framework.
- To provide unified energy function model and minimize it using optimization technique.
- To perform objective experiments to recover the full resolution stereoscopic video with high PSNR and SSIM.

E. Benefits of Proposed Methodology:

- Different form the existing methods, which deal with video super-resolution and depth estimation separately, the proposed algorithm reconstructs the low-resolution view while simultaneously performing stereo matching in a unified energy minimization framework. On the one hand, the stereo correspondence can guide on how to borrow the high-frequency information from the full-resolution view to enhance the quality of the low-resolution view.
- On the other hand, the enhanced stereo pair can generate more matching points and consequently improve the result of stereo correspondence. In the proposed algorithm, the energy minimization is performed by using an alternating optimization method.
- The proposed algorithm can recover the full resolution stereoscopic video with high PSNR and SSIM.
- Furthermore, it can also provide a high-quality depth map, which can be applied in many video applications, such as multi-view video coding and video semantic segmentation.
- Both objective and subjective experimental results verify that the proposed algorithm can achieve high-quality depth and super-resolution results while preserving good 3D visual experience.

F. Methodology

Given an asymmetric stereoscopic video, without loss of generality, we assume that the left view is of full-resolution and the right view is of low-resolution. The goal of the proposed method is to obtain a full-resolution video of the right view, by exploiting the plentiful details information of the left full-resolution video. The exploiting step builds up on the calculation of view correspondence using depth information. This method models the video super-resolution and stereo matching in a unified energy function. Then, we use an alternating optimization method to minimize the energy function

Inputs : Full-resolution video of left view IL and low-resolution video of right view IR (low).

Outputs : Full resolution video of right view IR N and the depth map DN.

The proposed method consists of two parts: Depth estimation and Video super-resolution.

G. Depth Estimation

In the first part, the depth map is estimated by using cooperative optimization for an energy function, which is composed of a data term, an occlusion term, a smoothness term, and a temporal consistency term for modeling the temporal correlations between adjacent depth frames. Then, the estimated depth map is used for guiding the super-resolution process in the second part.

H. Video Super-Resolution

Then, the estimated depth map is used for guiding the super-resolution process in the second part. Specifically, the super-resolved right view is estimated by using conjugate gradient method to optimize a quadratic energy function, which is composed of a data term, a mapping term for modelling the pixel correspondence between left and right views, and a nonlocal term for exploiting the nonlocal self-similarity.

II. BACKGROUND AND LITERATURE SURVEY

A. Literature Survey

H. Brust, A. Smolic, K. Mueller, G. Tech, and T. Wiegand, "Mixed resolution coding of stereoscopic video for mobile devices," in Proc. of 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2009

The results for full and mixed resolution stereo video coding including subjective and objective evaluation for 3DTV on Mobile devices are presented in this paper. For this, objective as well as subjective tests have been carried out with different sequences at different image sizes and evaluated on two different stereoscopic displays. The subjective tests showed that for uncoded sequences full resolution was more highly rated, however for coded sequences at low bit rates mixed resolution was rated better. A masking algorithm which is not sharp enough for up-sampling the lower resolution view only improved to a certain extent the overall quality because coding artifacts were improved and made more robust as well. Objective tests showed that the best obtainable bit rate distribution for mixed resolution stereo video coding is 30% to 35% of the total bit rate for the lower resolution view (half horizontal and half vertical). Finally, tests showed that mixed resolution stereo video decoding with up-sampling is less complex and time consuming than decoding full resolution stereo video which is important for implementation on mobile devices.

P. Aflaki, W. Su, M. Joachimiak, D. Rusanovskyy, M. M. Hannuksela, H. Li, and M. Gabbouj, "Coding of mixed-resolution multiview video in 3d video application," in Proc. of IEEE Int. Conf. Image Processing (ICIP), 2013
The emerging MVC+D standard specifies the coding of Multiview Video plus Depth (MVD) data for enabling advanced 3D video applications. MVC+D specifications define the coding of all views of MVD at equal spatial resolution and apply a conventional MVC technique for coding the multiview texture and the depth independently. This paper presents a modified MVC+D coding scheme, where only the base view is coded at the original resolution whereas dependent views are coded at reduced resolution. To enable inter-view prediction, the base view is down sampled within the MVC coding loop to provide a relevant reference for dependent views. At the decoder side, the proposed scheme consists of a post-processing scheme which up samples of the decoded views to their original resolution. The proposed scheme is compared against the original MVC+D scheme and an average of 4% delta bitrate reduction (dBR) in the coded views and 14.5% of dBR in the synthesized views are reported.

D. Garcia, C. Dorea, and R. de Queiroz, "Super resolution for multiview images using depth information," IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 9, pp. 1249–1256, sept. 2012
In stereoscopic and multiview video, binocular suppression theory states that the visual subjective quality of 3-D experience is not much affected by asymmetrical blurring of the individual views. Based on these studies, mixed-resolution frameworks applied for multiview systems offer great data-size reduction without incurring in significant quality degradation in 3-D video applications. However, it is interesting to recover high-frequency content of the blurred views, to reduce visual strain due to long-term exposure and to make the system suitable for free-viewpoint television. In this paper, we present a novel super-resolution technique, in which low-resolution views are enhanced with the aid of high-frequency content from neighboring full-resolution views, and the corresponding depth information for all views. Occlusions are handled by checking the consistency between views. Tests for synthetic and real image data in stereo and multiview cases are presented, and results show that significant objective quality gains can be achieved without any extra side information.

J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in Proc. of IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR), 2008
This paper addresses the problem of generating a super resolution (SR) image from a single low-resolution input image. We approach this problem from the perspective of compressed sensing. The low-resolution image is viewed as down sampled version of a high-resolution image, whose patches are assumed to have a sparse representation with respect to an over-complete dictionary of prototype signal atoms. The principle of compressed sensing ensures that under mild conditions, the sparse representation can be correctly recovered from the downsampled signal. We will demonstrate the effectiveness of sparsity as a prior for regularizing the otherwise ill-posed super-resolution problem. We further show that a small set of randomly chosen raw patches from training images of similar statistical nature to the input image generally serve as a good dictionary, in the sense that the computed representation is sparse and the recovered high-resolution image is competitive or even superior in quality to images produced by other SR methods.

N. Atzpadin, P. Kauff, and O. Schreer, "Stereo analysis by hybrid recursive matching for real-time immersive video conferencing," IEEE Trans. Circuits Syst. Video Technol., Special Issue on Immersive Telecommunications, vol. 14, no. 3, pp. 321–334, 2004.
Real-time stereo analysis is an important research area in computer vision. In this context, we propose a stereo algorithm for an immersive video-conferencing system by which conferees at different geographical places can meet under similar conditions as in the real world. For this purpose, virtual views of the remote conferees are generated and adapted to the current viewpoint of the local participant. Dense vector fields of high accuracy are required in order to guarantee an adequate quality of the virtual views. Due to the usage of a wide baseline system with strongly convergent camera configurations, the dynamic disparity range is about 150 pixels.

Considering computational costs, a full search or even a local search restricted to a small window of a few pixels, as it is implemented in many real-time algorithms, is not suitable for our application because processing on full-resolution video according to CCIR 601 TV standard with 25 frames per second is addressed—the most desirable as a pure software solution running on available processors without any support from dedicated hardware. Therefore, we propose in this paper a new fast algorithm for stereo analysis, which circumvents the window search by using a hybrid recursive matching strategy based on the effective selection of a small number of candidates. However, stereo analysis requires more than a straightforward application of stereo matching. The crucial problem is to produce accurate stereo correspondences in all parts of the image. Especially, errors in occluded regions and homogenous or less structured regions lead to disturbing artifacts in the synthesized virtual views. To cope with this problem, mismatches have to be detected and substituted by a sophisticated interpolation and extrapolation scheme.

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 8, August 2017

C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," in Proc. of IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR), 2011

Many computer vision tasks can be formulated as labeling problems. The desired solution is often a spatially smooth labeling where label transitions are aligned with color edges of the input image. We show that such solutions can be efficiently achieved by smoothing the label costs with a very fast edge-preserving filter. Our main contribution is to show that with such a simple framework state-of-the-art results can be achieved for several computer vision applications. In particular, we achieve 1) disparity maps in real time whose quality exceeds those of all other fast (local) approaches on the Middlebury stereo benchmark, and 2) optical flow fields which contain very fine structures as well as large displacements. To demonstrate robustness, the few parameters of our framework are set to nearly identical values for both applications. Also, competitive results for interactive image segmentation are presented. With this work, we hope to inspire other researchers to leverage this framework to other application areas.

A. K. Jain, A. E. Robinson, and T. Q. Nguyen, "Comparing perceived quality and fatigue for two methods of mixed resolution stereoscopic coding," IEEE Trans. Circuits Syst. Video Technol., vol. 24, no. 3, pp. 418–429, 2014.

Mixed resolution stereoscopic coding exploits the perceptual phenomenon of binocular suppression by encoding each view of a stereo pair at a different resolution. Many investigations have shown that a large reduction in bandwidth can be gained for a relatively small compromise in image quality by this asymmetric representation. However, the viability of such an encoding method depends on the subjective response to viewing such videos at length. While methods for mixed resolution coding have been developed on the presumption of a certain visual fatigue response, none have actually examined it. In this paper, we address this shortcoming in three experiments comparing two methods of binocular suppression processing. The first two experiments reveal subjects' preferences in terms of overall quality between the two methods for short exposures, and the third experiment examines the fatigue resulting from 10-min exposures to mixed resolution encoded videos.

J. Kowalczuk, E. Psota, and L. Perez, "Real-time temporal stereo matching using iterative adaptive support weights," in Proc. of 2013 IEEE International Conference on Electro/Information Technology (EIT), 2013

Stereo matching algorithms are nearly always designed to find matches between a single pair of images. A method is presented that was specifically designed to operate on sequences of images. This method considers the cost of matching image points in both the spatial and temporal domain. To maintain real-time operation, a temporal cost aggregation method is used to evaluate the likelihood of matches that is invariant with respect to the number of prior images being considered. This method has been implemented on massively parallel GPU hardware, and the implementation ranks as one of the fastest and most accurate real-time stereo matching methods as measured by the Middlebury stereo performance benchmark.

J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," IEEE Trans. Image Processing, vol. 19, no. 11, pp. 2861–2873, 2010.

In this paper a new algorithm for single-image super-resolution based on sparse representation over a set of coupled low and high resolution dictionary pairs is proposed. The sharpness measure is defined via the magnitude of the gradient operator and is shown to be approximately scale invariant for low and high resolution patch pairs. It is employed for clustering low and high resolution patches in the training stage and for model selection in the reconstruction stage. A pair of low and high resolution dictionaries is learned for each cluster. The sharpness measure of a low resolution patch is used to select the appropriate cluster dictionary pair for reconstructing the high resolution counterpart. The sparse representation coefficients of low and high resolution patches are assumed to be equal. By multiplying the high resolution dictionary and the sparse coding coefficient of a low resolution patch, the corresponding high resolution patch is reconstructed. Simulation results in terms of PSNR and SSIM and visual comparison, indicate the superior performance of the proposed algorithm compared to the leading super-resolution algorithms in the literature over a set of natural images in sharp edges and corners.

C. Liu and D. Sun, "A bayesian approach to adaptive video super resolution," in Proc. of IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR), 2011, pp. 209–216.

Although multi-frame super resolution has been extensively studied in past decades, super resolving real-world video sequences still remains challenging. In existing systems, either the motion models are oversimplified, or important factors such as blur kernel and noise level are assumed to be known. Such models cannot deal with the scene and imaging conditions that vary from one sequence to another. In this paper, we propose a Bayesian approach to adaptive video super resolution via simultaneously estimating underlying motion, blur kernel and noise level while reconstructing the original high-res frames. As a result, our system not only produces very promising super resolution

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 8, August 2017

results that outperform the state of the art, but also adapts to a variety of noise levels and blur kernels. Theoretical analysis of the relationship between blur kernel, noise level and frequency wise reconstruction rate is also provided, consistent with our experimental results.

III. SYSTEM ARCHITECTURE

A. Technology Introduction

For given an asymmetric stereoscopic video, without loss of generality, it is assumed that the left view is of full-resolution and the right view is of low-resolution. The goal of the proposed technique is to obtain a full-resolution video of the right view, by exploiting the plentiful details information of the left full-resolution video. The exploiting step builds up on the calculation of view correspondence using depth information. This method models the video super-resolution and stereo matching in a unified energy function. An alternating optimization method is used to minimize the energy function.

B. PROPOSED METHODOLOGY

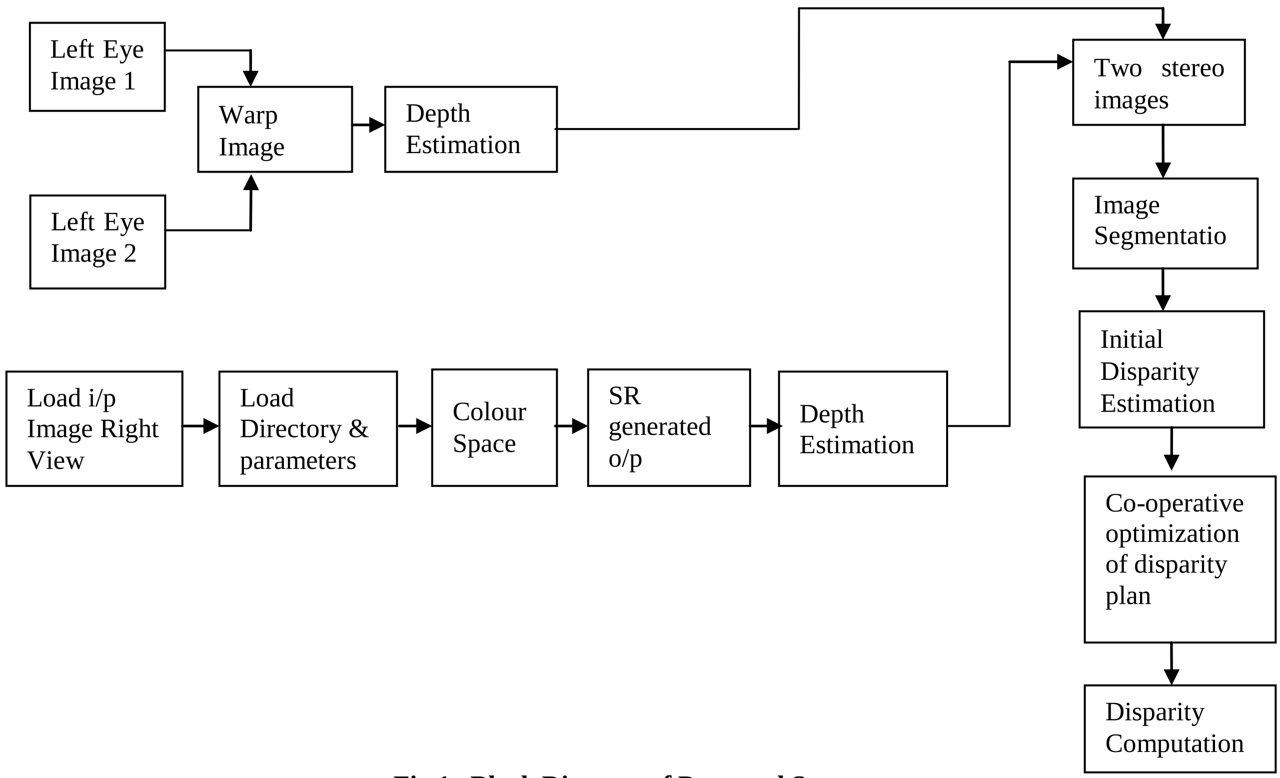

Fig 1. Block Diagram of Proposed System

C. Algorithm

The flowchart of the proposed method is shown in figure 1. It consists of two parts: depth estimation and video super resolution. We alternately update these two steps until stable results are obtained.

New depth-based super resolution method is presented in this work, which leverages the high-frequency information of the left full-resolution view for enhancing the resolution of the right view. This method belongs to the category of reconstruction-based methods, which set up an energy function and then minimize it to obtain the optimal solution. The energy function usually consists of a data term and some other constraint terms. Data term for energy function is mathematically as follows

$$E_{data} = \left\| SKI_N^R - I_N^R\,(low) \right\|_2^2,$$

where S is a down-sampling operator and K is a blurring operator, I_N^R is a variable referring to the expected full resolution right view of the Nth frame and I_N^R (low) is the initial low-resolution input of the Nth frame. || . ||2 denotes the Euclidean norm.

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 8, August 2017

In addition, the high-frequency information of the left full resolution view can be used to enhance the resolution of right view, since they share many scene points. Therefore, once the correspondence between the left and right views is obtained, a mapping term can also be added to the energy function. Similar disparity based pixel mapping strategy is also applied.

$$E_{map} = \sum_{(m,n)\in\Lambda} c_{mn} \left\| I_N^R(m,n) - I_N^L\left(m, n + D_N'(m,n)\right) \right\|_2^2$$

where Λ is the pixel index set of the image grid, I_N^L is the Nth frame of left full-resolution view, $D_N^{'}$ denotes the stereo correspondence of the Nth frame (depth map[1] of I_N^R relative to I_N^L). It can be obtained from the corresponding left view depth map D_N of I_N^L relative to I_N^R, and linear interpolating is used to deal with the non-integer case. C_{mn} is a binary confidence value about $D_N^{'}$ (m, n). It is necessary since the depth map may be not accurate, especially in the occlusion regions and non-overlapping regions of two views. In this paper, we determine C_{mn} by measuring the similarity (mean square error, MSE) between the local patch centered at I_N^R (m, n) and the local patch centered at I_N^L (m,n+ $D_N^{'}$ (m,n)).

Besides the above observation about the point-to-point mapping between two views, there is another useful observation about natural image, nonlocal prior. This nonlocal prior is based on such an observation that the image content is likely to repeat itself within some neighbourhood. This self-similarity of natural image is beneficial for solving super-resolution problem, because it means that we can exploit the redundant information hidden in the full-resolution view. Leveraging the nonlocal prior, we enforce an additional nonlocal constraint between the left view and right view under the guidance of stereo correspondences. The explicit form of this nonlocal regularization term is:

$$E_{nonlocal} = \sum_{(m,n)\in\Lambda} c_{mn} \sum_{(p,q)\in\Omega_{nr}\left(m,n+D_N'(m,n)\right)} w_{mn,pq} \left\| TI_N^R(m,n) - TI_N^L(p,q) \right\|_2^2 .$$

Minimizing the above energy function relies on the calculation of stereo correspondence. The result of stereo matching algorithm depends on the co-occurrence of distinct details in both views. The more distinct details these two views share, the more reliable the matching result is. However, since the mixed-resolution videos are only inputs, the result of directly matching the left view with the interpolated right view may not meet the expectation. The details of right view should be restored for obtaining are liable depth map. So, the calculation of stereo correspondence and the super-resolution together, and propose a unified function as follows:

$$E_{SR} = E_{data} + \lambda_1 E_{map} + \lambda_2 E_{nonlocal} + \lambda_3 E_{depth},$$

where λ_1, λ_2, and λ_3 are regularization parameters, E_{depth} is the depth energy function.

A region-based stereo matching algorithm using cooperative optimization is proposed in [19]. This method can achieve high-quality depth map with relatively high efficiency. In this paper, we extend this method to the stereoscopic video case. By using the temporal consistency of depth information in stereoscopic video, this extension method can obtain temporally consistent depth maps. A total energy function, which can be decomposed into the sum of several sub target energy functions. Mathematically

$$E_{depth} = \sum_{i\in\Lambda_{seg}} E^i,$$

where Λ seg is the index set of regions, E^i is the energy function of the i^{th} region R^i the explicit form of every sub target E^i. Here, we mainly concentrate on four aspects: data energy, occlusion energy, smoothness energy, and temporal consistency energy. Mathematically, the energy function of the i^{th} region $R^i$is defined as follows:

$$E^i = E^i_{data} + E^i_{occlusion} + E^i_{smooth} + E^i_{consistency}.$$

DOI10.17148/IJARCCE.2017.6822

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 8, August 2017

The first term is the data term. It evaluates the validity of the depth at the position (m, n) in region R^i by calculating the colour difference between two corresponding pixels. The first term is the data term. It evaluates the validity of the depth at the position (m, n) in region R^i by calculating the colour difference between two corresponding pixels. The second term is the occlusion energy. It imposes a fixed penalty on occlusion pixels. The third term is the smoothness energy. It is necessary since depth varies a little in a small neighbourhood. This term imposes a fixed penalty on the boundary pixels of each region, whose values are different from its neighbours. a consistency energy term that accounts for the penalty on temporal inconsistency. For a video, temporal consistency means the correspondence between adjacent frames, i.e., a scene point, which appears in several adjacent frames should share the same depth values. This correspondence can be calculated according to the forward and backward optical flow field.

Considering all the energy functions together, the explicit form of the unified energy function for depth-based super-resolution is as follows

$$\begin{aligned} E_{SR} = & \left\| SKI_N^R - I_N^R(low) \right\|_2^2 + \lambda_1 \sum_{(m,n)\in\Lambda} c_{mn} \left\| I_N^R(m,n) - I_N^L\left(m, n + D_N^{\prime}(m,n)\right) \right\|_2^2 + \\ & \lambda_2 \sum_{(m,n)\in\Lambda} c_{mn} \left(\sum_{(p,q)\in\Omega_{nr}\left(m,n+D_N^{\prime}(m,n)\right)} w_{mn,pq} \left\| TI_N^R(m,n) - TI_N^L(p,q) \right\|_2^2 \right) + \\ & \lambda_3 \sum_{i\in\Lambda_{seg}} \left(\begin{array}{l} \sum_{(m,n)\in V_R^i,(p,q)\in V_L^i} \left\| I_N^R(m,n) - I_N^L(p,q) \right\|_{\infty} + \left(\left|Occ_L^i\right| + \left|Occ_R^i\right| \right) \lambda_{occ} + \left|B_{discon}^i\right| \lambda_s + \\ \sum_{k\in\{\pm k_0,\ldots,\pm 1\}} \sum_{(m,n)\in R^i} \left\| D_N(m,n) - D_{N+k}\left(w_{N,N+k}(m,n)\right) \right\|_2 \end{array} \right) \end{aligned}$$

Minimizing the above energy function, we can obtain the optimal solution for I_N^R and D_N. It can be described as the following optimization problem:

$$\begin{aligned} \left\{ I_N^{R*}, D_N^* \right\} &= \underset{\{I_N^R, D_N\}}{\arg\min}\ E_{SR} \\ &= \underset{\{I_N^R, D_N\}}{\arg\min}\ E_{data} + \lambda_1 E_{map} + \lambda_2 E_{nonlocal} + \lambda_3 \sum_{i\in\Lambda_{seg}} E^i. \end{aligned}$$

Directly optimizing the above energy function is difficult. An alternating optimization technique is used to solve the above optimization problem. The unified energy function is optimized with respect to one variable while keeping the other one fixed. Set of advanced optimization techniques so that the proposed method can effectively handle this challenging problem.

Step 1: Optimizing D_N

This step optimizes energy function to obtain $D_N^{(t+1)}$ given the previous super-resolved result $I_N^{R(t)}$. Co-operative optimization technique is used to solve this optimization problem. The principle of cooperative optimization is decomposing a complex target into some comparatively simple sub targets, and optimizing these sub targets individually by keeping the consistent common parameters. While optimizing, local window-based matching method is used to determine the initial depth map.

Step 2: Optimizing I_N^R

In this step, the energy function is optimized to obtain $I_N^{R(t+1)}$ under the guidance of the previous depth estimate $D_N^{(t+1)}$. In other words, when the super-resolution part of the proposed method is used to obtain $I_N^{R(t+1)}$, keeping variable D_N^R in the energy function as a constant, i.e., $D_N^{(t+1)}$.

Step 3: Parameter Setting

The flowchart of the proposed method is shown in Fig. 3.1.The low-resolution videos are obtained by applying a blurring and down-sampling process. Here, a 3×3 Gaussian blur kernel is used with a standard deviation 1, as well as a down-sampling operator that samples pixels from odd rows and columns. Andin the following experiments, the bilinear interpolation method is used to obtain the initial full-resolution right view. The parameters in cooperative optimization technique are set the same with [19]. The nonlocal neighborhood is set to be11×11. The patch radius of the vectorization patch extraction operator T is 1. The regularization parameters λ_1 and λ_2 in are set to 0.005 empirically. Since alternately I_N^R and D_N is optimized, the setting of $\lambda_3$is trivial. Thus, It is just set it to 1.

SOFTWARE REQUIREMENTS: MATLAB

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 8, August 2017

IV. RESULTS AND DISCUSSION

PNG : LR/HR Size and PSNR/SSIM of Bicubic Interpolation Comparison for Proposed Algorithm for Different Images

Image	LR Frame Size	HR Frame Size	PSNR(dB)	SSIM(dB)
Parkinglot.png	128X128	256X256	29.181915	0.867931
Yellowstone.png	256X256	512X512	25.038400	0.815538
Parkinglot.png	400X300	800X600	36.170336	0.964322
Globe.png	512X512	1024X1024	38.134612	0.995621
Globe.png	640X480	1280X960	39.135092	0.996338



Fig.(ii) Input Image Fig. (iii) Bicubic Fig. (iv) Sparse

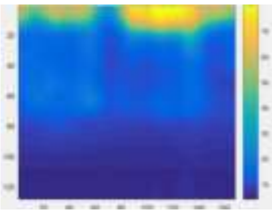
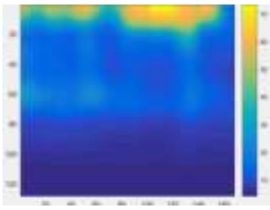

Fig.(v) Depth of Bicubic Fig. (vi). Depth of Sparse



Fig.(vii) Left Frame Fig(ix).Depth Estimation Fig.(viii) Right Frame

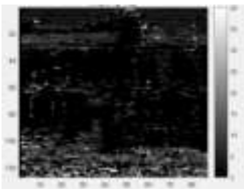
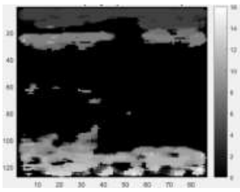
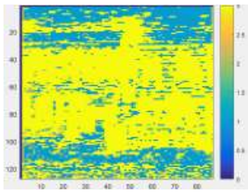

Fig.(x) Disparity Map Fig.(xi) Median Filtered Fig.(xii) Optical Flow Disparity Map

Advantages:

1. Inspired by suppression theory, mixed-resolution approach, in which low-resolution and full-resolution images are jointly used, has been applied to reduce the amount of data to be compressed.
2. Both objective and subjective experimental results verify that the proposed algorithm can achieve high quality depth and super-resolution results while preserving good 3D visual experience.

Applications:

1. They can bring new user experiences to a set of applications, such as mobile 3D TV, free-viewpoint video, and immersive teleconferencing.
2. It can also provide a high-quality depth map, which can be applied in many video applications, such as multiview video coding and video semantic segmentation

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 8, August 2017

V. CONCLUSION

In this work, we have proposed a novel method for reconstructing the low-resolution view and simultaneously estimating depth information for asymmetric stereoscopic video. The proposed method models these two problems as a unified energy function and then minimizes it by
Using an alternating optimization technique. The alternating steps interact with each other to improve the depth result as well as the super resolution result. Objective experiments convince that the proposed method can achieve high-quality depth and super resolution result In addition, subjective experiments show the superiority of the proposed method in terms of visual "sharpness". The computational cost of the proposed method mainly concentrates on three aspects: calculation of the optical flow field, stereo matching, and nonlocal operation in the super-resolution part.

ACKNOWLEDGMENT

This research was supported by my guide **Prof. Rohita Patil.** Without the full support and encouragement of my guide, this thesis would not have been completed in time.
We thank our HOD **Dr. S. K. Shah** who provided insight and expertise that greatly assisted the research.
My heartfelt gratitude to our respected Principal **Dr. A. V. Deshpande** and Vice Principal **Dr. K.R. Borole** for his encouragement throughout the course.
My sincere thanks to all other teaching and nonteaching staff members and colleagues who have helped me directly or indirectly in completing this thesis.
I thank my parents for wishes and moral support during the thesis and all concerned for helping and encouraging me in this honest pursuit.
I am grateful for the many useful comments and suggestions provided by reviewers, which have been resulted significant improvements in this thesis.

REFERENCES

[1] H. Brust, A. Smolic, K. Mueller, G. Tech, and T. Wiegand, "Mixed resolution coding of stereoscopic video for mobile devices," in Proc. of 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video, 2009
[2] P. Aflaki, W. Su, M. Joachimiak, D. Rusanovskyy, M. M. Hannuksela, H. Li, and M. Gabbouj, "Coding of mixed-resolution multiview video in 3d video application," in Proc. of IEEE Int. Conf. Image Processing (ICIP), 2013
[3] D. Garcia, C. Dorea, and R. de Queiroz, "Super resolution for multiview images using depth information," IEEE Trans. Circuits Syst. Video Technol., vol. 22, no. 9, pp. 1249–1256, sept. 2012
[4] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in Proc. of IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR), 2008
[5] N. Atzpadin, P. Kauff, and O. Schreer, "Stereo analysis by hybrid recursive matching for real-time immersive video conferencing," IEEE Trans. Circuits Syst. Video Technol., Special Issue on Immersive Telecommunications, vol. 14, no. 3, pp. 321–334, 2004.
[6] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," in Proc. of IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR), 2011
[7] A. K. Jain, A. E. Robinson, and T. Q. Nguyen, "Comparing perceived quality and fatigue for two methods of mixed resolution stereoscopic coding," IEEE Trans. Circuits Syst. Video Technol., vol. 24, no. 3, pp. 418–429, 2014.
[8] J. Kowalczuk, E. Psota, and L. Perez, "Real-time temporal stereo matching using iterative adaptive support weights," in Proc. of 2013 IEEE International Conference on Electro/Information Technology (EIT), 2013
[9] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," IEEE Trans. Image Processing, vol. 19, no. 11, pp. 2861–2873, 2010.
10] C. Liu and D. Sun, "A bayesian approach to adaptive video super resolution," in Proc. of IEEE Int. Conf. Computer Vision and Pattern Recognition (CVPR), 2011, pp.209–216.