Jian Jin (金鉴)

Tel: +86 18519643859

E-mail: jianj008@gmail.com or jian.jin@ntu.edu.sg

Personal page: https://jianjin008.github.io

BIO

I am a Research Fellow in Prof. Weisi Lin's group in Alibaba-NTU Singapore Joint Research Institute, Nanyang Technological University (NTU). I received my Ph.D. degree from Beijing Jiaotong University (BJTU), supervised by Prof. Yao Zhao and Prof. Chunyu Lin. I was a Joint Ph.D. Student from Sep. 2016 to Sep. 2018 in the School of Engineering Science, Simon Fraser University (SFU), supervised by Prof. Jie Liang. I was fortunate to work closely with Prof. Ming-Ting Sun (UW), Prof. Bing Zeng (UESTC), and Prof. Ce Zhu (UESTC).

EDUCATION

- Ph.D. in School of Computer and Information Technology, Beijing Jiaotong University, P. R. China, Sep. 2014 – Jun. 2019
- M.E. in School of Electrical Engineering, Taiyuan University of Science and Technology, P. R. China, Sep. 2011 - Jun. 2014
- B.E. in School of Economics and Management, Taiyuan University of Science and Technology, P. R. China, Sep. 2007 - Jun. 2011

ACADEMIC RESEARCH EXPERIENCE

- Project from Alibaba, visual (image/video/feature) quality assessment, May 2020 – now, Alibaba-NTU Singapore Joint Research Institute, Nanyang Technological University.
- Project from Alibaba, visual data compression for human and machine, standards and applications, May 2020 – now, Alibaba-NTU Singapore Joint Research Institute, Nanyang Technological University.
- Project from Alibaba, real-time communication video quality assessment, and applications, Nov. 2019 – May 2020, Alibaba-NTU Singapore Joint Research Institute, Nanyang Technological University.
- Doctoral Graduate Research Program, image coding and digital watermarking, Oct. 2018 – Jun. 2019, Beijing Jiaotong University.
- Visiting Ph.D. Research, supported by China Scholarship Council (CSC) Scholarship, (1) view synthesis distortion estimation modeling for 3-D video coding, (2) machine learning-based image/video coding and (3) automatic driving technique, Sep. 2016 – Sep. 2018, school of engineering science, Simon Fraser University.
- Research Assistant, working with Prof. Jie Liang, Jan. 2017 – May 2018, school of engineering science, Simon Fraser University:

- Teaching Assistant for ensc424 “Multimedia Communications Engineering”, working with Prof. Jie Liang, Sep. 2017 – Dec. 2017, School of Engineering Science, Simon Fraser University, Canada
- Doctoral Graduate Research Program, objects, and microcube based 3-D video coding focus on view synthesis distortion analysis, Sep. 2015 – Aug. 2016, Beijing Jiaotong University.
- Doctoral Graduate Research Program, objects, and microcube based 3-D video coding focus on the theoretical study of view synthesis. Feb. 2015 – Aug. 2015, Beijing Jiaotong University.
- Doctoral Graduate Research Program, image, and video coding focus on the theoretical study of image and video coding. Sep. 2014 – Jan. 2015, Beijing Jiaotong University.

RESEARCH INTERESTS

- *Perceptual modeling for human and machines*, including Just Noticeable Difference modeling for human visual system/deep machine vision.
- *Visual data compression for human and machines*, including feature coding, deep codec, entropy modeling, rate distortion optimization.
- *Image/video/feature quality assessment*.
- *2-D/3-D Image and Video Coding*, including traditional image/video coding, depth information quantization, depth video coding, virtual view synthesizing, view synthesis distortion estimation modeling, and perceptual video coding.
- *3-D Scene Reconstruction*, including image-based rendering, graph rendering, and camera calibration.
- *Deep Learning*, including basic neural networks designing, and common platforms using, such as TensorFlow, Pytorch, Keras.

PROFESSIONAL SKILLS

- C/C++, MATLAB, Python (Tensorflow, Pytorch), LaTeX, Windows Office.

AWARDS (*latest two years*)

- 2019 Excellent Doctoral Dissertation Award of Chinese Institute of Electronics (Nov. 2019, only **26** Ph.D.s, which are in the **electronics domain** and **graduated in recent three years**, selected this award in China)
- Outstanding Graduates Awards of Beijing (Jun. 2019, only 3 Ph.D. got this award in BJTU)
- Outstanding Graduates Awards of BJTU (Jun. 2019)
- President Scholarship of BJTU (Apr. 2017 **the greatest honor for the graduate students at BJTU**)
- Annual Best paper of Institute of Information Science, Beijing Jiao Tong University (Jan. 2017)
- The Hanergy Li Jianing Scholarship (Nov. 2016 Rank 1)
- China National Scholarship (Oct. 2016 Rank 1)
- China Scholarship Council (CSC) Scholarship (Jun. 2016)
- Excellent Paper Award of BJTU (Jun. 2016)
- Ph.D. Innovation Fund of BJTU (Dec. 2015)

- Miyoshi outstanding graduate student of BJTU (Nov. 2015)
- First-class Ph.D. academic scholarship of BJTU (Oct. 2015)

PUBLICATION LIST

[1] **Jian Jin**, Xingxing Zhang, Weisi Lin, Jie Liang, Yao Zhao, Chunyu Lin, "Layer-Oriented View Synthesis Distortion Estimation with Learned Weights", *IEEE Transactions on Image processing* (***TIP***), 2021. *(Preparing)*

[2] **Jian Jin**, Xingxing Zhang, Xin Fu, Huan Zhang, Weisi Lin, Jian Lou, Yao Zhao, "Just Noticeable Difference for Deep Machine Vision", *IEEE Transactions on Circuits and Systems for Video Technology* (***TCSVT***), 2021.

[3] Geyang Wang, Yao Zhao, Chunyu Lin, Meiqin Liu, and **Jian Jin**, "*Dually Octagonal Projection for 360 Video with Less-Distortion Introduced*", *International Conference on Signal Processing (**ICSP**)*, vol. 1, pp. 246-251. IEEE, 2020.

[4] Chao Yao, Jimin Xiao, **Jian Jin**, Xiaojuan Ban, "Edge Orientation Driven Depth Super-Resolution for View Synthesis", *International Conference on Image and Graphics (**ICIG**)*, 2020.

[5] **Jian Jin**, Jie Liang, Yao Zhao, Chunyu Lin, Chao Yao, Lili Meng, "Pixel-level View Synthesis Distortion Estimation for 3-D Video Coding", *IEEE Transactions on Circuits and Systems for Video Technology* (***TCSVT***), 2019.

[6] **Jian Jin**, Jie Liang, Yao Zhao, Chunyu Lin, Chao Yao, Anhong Wang, "A Depth-Bin-Based Graphical Model for Fast View Synthesis Distortion Estimation", *IEEE Transactions on Circuits and Systems for Video Technology* (***TCSVT***) , 2019, 29(6): 1754 - 1766.

[7] **Jian Jin**, Anhong Wang, Yao Zhao, Chunyu Lin, and Bing Zeng, "Region-aware 3D-warping for DIBR," *IEEE Transactions on Multimedia* (***TMM***), 2016, 18(6): 953-966.

[8] **Jian Jin**, Yao Zhao, Chunyu Lin, and Anhong Wang, "An Accurate and Efficient Nonlinear Depth Quantization Scheme," *2015 Pacific Rim Conference on Multimedia—PCM* (***PCM 2015***), Korea, pp. 390-399, Aug. 2015.

[9] **Jian Jin**, Anhong Wang, Yao Zhao, and Chunyu Lin, "A fast region-level 3D-warping method for depth-image-based rendering", *2015 IEEE International Workshop on Multimedia Signal Processing* (***MMSP 2015***), Xiamen, pp. 1-6, Oct. 2015.

[10] Lijun Zhao, Anhong Wang, Bing Zeng, **Jian Jin**, "Scalable Coding of Depth Images with Synthesis-Guided Edge Detection," KSII Transactions on Internet and Information Systems, 2015, 9(10): 4108-4125.

[11] Zhiwei Xing, Anhong Wang, **Jian Jin**, Yingchun Wu, "Synthesis-Aware Region-Based 3D Video Coding," 2015 Pacific Rim Conference on Multimedia—PCM (PCM 2015), Korea, pp. 400-409, Aug. 2015.

PROFESSIONAL ACTIVITIES

- **Membership:**

- IEEE

- **Technical Program/Review Committee Member:**
  - National Conference on Image and Graphics (NCIG 2020)
  - IEEE International Conference on Visual Communications and Image Processing (VCIP 2018)
  - Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2018)
- **Reviewer for:**
  - IEEE Transactions on Image Processing
  - IEEE Transactions on Circuits and Systems for Video Technology
  - IEEE Transactions on Multimedia
  - Neurocomputing
  - Multimedia Systems
  - EURASIP Journal on Image and Video Processing
  - International Journal of Pattern Recognition and Artificial Intelligence
  - IEEE International Conference on Multimedia & Expo (ICME 2020)
  - IEEE International Symposium on Circuits and Systems (ISCAS 2019)
  - IEEE International Workshop on Multimedia Signal Processing (MMSP 2019)
  - IEEE International Conference on Visual Communications and Image Processing (VCIP 2019, 2018)
  - Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2018, 2017)

RESEARCH ACCOMPLISHMENT

My research accomplishment to date includes four directions:

Peceptual modeling for human visual system and/or deep machine vision; Visual data (image/video/feature) compression for human and machines; Visual (image/video/feature) quality assessment; 3-D video coding and rendering; For perceptual modeling, I mainly focus on the just noticeable difference modeling for human and machines. Besides, unlike traditional JND modeling for signal level, we are trying to develop the feature level ones. Visual quality assessment, I focus on the blend quality assessment for the user generated contents and also try to develop metrics for feature quality assessment. For visual data compression for human and machines, I devote to mine the relations and differences between signal reconstruction and machine visual analysis tasks, and develop a new paradigm for visual compression for human and machines. For 3-D video coding, I focus on depth information preprocessing to achieve significant compression for depth video. Besides, I have developed the view synthesis distortion estimation algorithm to serve for the rate-distortion optimization during 3-D video coding. For 3-D video rendering, I have proposed a region-aware 3-D warping algorithm to speedup view synthesis processing.

Just noticeable difference modeling for deep machine vision

As an important perceptual characteristic of the Human Visual System (HVS), the Just Noticeable Difference (JND) has been studied for decades with image and video processing (e.g., perceptual visual signal compression). However, there is little exploration on the existence of JND for the Deep Machine Vision (DMV), although the DMV has made great strides in many machine vision tasks. In this paper, we take an initial attempt, and demonstrate that the DMV has the JND, termed as the DMV-JND. We then propose a JND model for the image classification task in the DMV. It has been discovered that the DMV can tolerate distorted images with average PSNR of only 9.56dB (the lower the better), by generating JND via unsupervised learning with the proposed DMV-JND-NET. In particular, a semantic-guided redundancy assessment strategy is designed to restrain the magnitude and spatial distribution of the DMV-JND. Experimental results on image classification demonstrate that we successfully find the JND for deep machine vision. Our DMV-JND facilitates a possible direction for DMV-oriented image and video compression, watermarking, quality assessment, deep neural network security, and so on.

The main contributions in this research are summarized as follows.

- To the best of our knowledge, our work is the first to demonstrate that the DMV has the JND. Besides, we also propose the first algorithmic framework to model the DMV-JND.
- The proposed DMV-JND model, achieved via unsupervised learning with our DMV-JND-NET, is capable of generating the DMV-JND distorted image with average PSNR of only 9.56dB for the DMV.
- A semantic-guided redundancy assessment strategy is introduced toward the reasonability of the generated DMV-JND, by restraining its magnitude and spatial distribution.
- Reducing the noise from the resultant DMV-JND to zero, the DMV maintains the RCA throughout the process; This demonstrates that the DMV has the same homogeneous property as the HVS-JND.

Deep blendness image quality assessment for user generated contents

To be continue …

A novel paradigm of image compression for human and machines: combining high-level sementic and low-level signal

To be continue …

Depth information quantization

As known, depth information exists as floating distance data, when firstly captured by the depth sensor. In view of storage and transmission, it is necessary to be quantized into several depth layers. Generally, it is mutually contradictory between the efficiency of depth quantization and the accuracy of view synthesis. Since 3D-warping rounding calculation exists during view synthesis, depth changes within a certain range will not cause different warped position. This phenomenon provides a good way to quantize depth data more efficiently. However, 3D-warping rounding calculation can also bring additional view synthesis distortion, if the warped-interval and image-resolution-interval are misaligned. Hence, to achieve efficient depth quantization without introducing additional view

synthesis distortion, an accurate and efficient nonlinear-depth quantization scheme (AE-NDQ) is presented in which the alignment between warped-interval and image-resolution-interval is taken into consideration during the depth quantization. Experimental results show, compared with the efficient nonlinear-depth-quantization (E-NDQ), AE-NDQ needs almost the same bits to represent the depth layers but maintains more accurate on view synthesis. For the traditional 8-bits nonlinear-depth-quantization (NDQ), AE-NDQ needs less bits to represent the depth layers, while has the same accuracy of the synthesized view.

Depth-bin-based graphical model for fast view synthesis distortion estimation

During 3-D video communication, transmission errors such as packet loss could happen to the texture and depth sequences. View synthesis distortion will be generated when these sequences are used to synthesize virtual views according to the depth-image-based rendering method. Depth-value-based graphical model (DVGM) has been employed to achieve the accurate packet-loss-caused view synthesis distortion estimation (VSDE). However, the DVGM models the complicated view synthesis processes at depth-value level, which costs too much computation and is difficult to be applied in practice. In this work, a depth-bin-based graphical model (DBGM) is developed where the complicated view synthesis processes are modeled at depth-bin level so that it can be used for the fast VSDE with 1-D parallel camera configuration. To this end, several depth values are fused into one depth bin, and a depth-bin-oriented rule is developed to handle the warping competition process. Then, the properties of depth bin are analyzed and utilized to form the DBGM. Finally, a conversion algorithm is developed to convert the per-pixel input depth value probability distribution into the depth-bin format. Experimental results verify that our proposed method is 8 to 32 times faster and requires 17% to 60% less memory than the DVGM, with exactly the same accuracy.

In this work, the main contributions are listed as follows.

- The concept of depth bin is firstly defined. At the same time, a depth-bin-oriented warping competition rule is developed.
- The DBGM is developed, which is the first work to formulate the complicated view synthesis process at depth bin level to simplify the VSDE.
- The properties of depth bin are studied and utilized to optimize the DBGM further.
- A conversion of probability distribution between depth bin and depth value is developed so that it can be used to integrate the RODE method into the DBGM directly.

Pixel-level view synthesis distortion estimation for 3-D video coding

Recently, region-based 3-D video coding has been proposed. However, existing view synthesis distortion estimation (VSDE) methods are performed at the frame level. To guide the rate-distortion optimization process of region-based 3-D video coding schemes, we propose the first pixel-level VSDE (PL-VSDE) method. We first give the definition of the pixel-level view synthesis distortion. To estimate it, a backward prediction method is then developed, which starts from the pixels of interest (POIs) in the virtual view and finds their corresponding pixels in the reference view via a coarse-to-fine approach, denoted as coarse-to-fine backward prediction (CFBP) method. Additionally, the CFBP fully considers the details of 3-D warping, the rounding operation and the warping competition in view

synthesis, leading to improved accuracy of the prediction. Besides, a table-lookup method and a warping property are introduced to speed up the CFBP. After integrating the CFBP into the PL-VSDE, we can estimate the view synthesis distortion at the pixel level. Our method is carried out pixel-by-pixel independently, which is friendly for parallel processing. Experimental results demonstrate that our proposed method has significant advantages in both accuracy and efficiency compared with state-of-the-art frame level VSDE methods.

The main contributions of the paper are listed as follows:

- We carefully analyze the depth-change-caused view synthesis distortion, which shows that the accuracies of the traditional depth-change-caused VSDE algorithms are not satisfactory, since they only use 3-D warping to roughly propagate the depth-change-caused distortion from reference view to the warped view. To improve their accuracies, both the rounding operation and warping competition should be considered as well.
- We propose a pixel-level view synthesis distortion estimation method (PL-VSDE) to better serve the emerging region-based 3-D video coding schemes. Our method is based on a backward prediction, which starts from a pixel in the virtual view, and find its corresponding pixel in the reference view via a coarse-to-fine approach. Besides, it also considers the details in 3-D warping, rounding operation, and warping competition.
- We use a table-lookup method and the warping property to further speed up the CFBP. Besides, our method can be implemented pixel-by-pixel independently, which is friendly for parallel processing.

Region-aware 3-D warping for DIBR

In 3-D video (3DV) applications, depth-image-based rendering (DIBR) has been widely employed to synthesize virtual views. However, this approach is performed in a frame-based way, meaning each whole frame is dealt with and the characteristics of different regions in the frame are ignored. As a result, redundant pixels in some regions are abused during the subsequent warping and blending stage. We propose a region-aware 3-D warping approach for DIBR in which warped frames are reasonably divided beforehand so that only the indispensable regions are used. With the proposed scheme, it is possible to avoid noneffective and repeated pixels during the warping stage. In addition, the blending process is also saved. The experimental results show that compared to the state-of-the-art VSRS3.5 and VSRS-1D-fast algorithms, our approach can achieve significant computation savings without sacrificing synthesis quality.