

# First Order Motion Model for Image Animation and Image Compression Using Wavelet Techniques

Sangita Kumari<sup>1</sup>, Prof. Nitesh Gupta<sup>2</sup>

<sup>1</sup>M. Tech. Scholar, <sup>2</sup>Guide and Assistant Professor

<sup>1,2</sup>Computer Science Engineering, NRI Institute of Information Science & Technology, Bhopal, India

## I. INTRODUCTION

These wavelet-based image/video compression algorithms (SPIHT and STW) are considered as refined versions of the seminal EZW algorithm. The 3D-Set Partitioning in hierarchical trees (3D-SPIHT) technique which is proposed by Kim and Pearlman is the extended form of SPIHT coding algorithm, in which the relationship among coefficients lying in different frequency bands is based on octal tree structure rather than quad-tree structure. The most enhanced image compression algorithm is the Adaptively Scanned Wavelet Difference Reduction (ASWDR) algorithm proposed by Walker. ASWDR technique adjusts the scanning order used by Wavelet Difference Reduction (WDR) algorithm, so as to predict locations of new significant values. The WDR method employs a fixed ordering of the positions of wavelet coefficients. Thus, ASWDR technique achieves high compression than WDR while retaining all of the important features of WDR such as low complexity, region of interest (ROI) capability and progressive SNR capability. The rate of compression achieved is largely determined by the encoding technique and the number of encoding loops used. Thus, in this paper the most powerful wavelet-based compression technique is identified by presenting a comparative study of the various approaches.

## II. PERFORMANCE MEASURES

When the compression is lossy in nature, the decompressed image may not be the same as the original image. Achieving a high compression ratio leads to a loss of more details in the image. The challenge of compression methods is to find the best compromise between a high compression ratio and a good perceptual result. The metrics used to compare the various image compression techniques are the Compression Ratio (CR), Mean Square Error (MSE), Peak Signal to Noise Ratio (PSNR), Maximum Error, L2-Norm Ratio and Bits Per Pixel (BPP).

*Compression Ratio:*

Compression Ratio measures the efficiency of the compression algorithm by computing the percentage of

compression achieved. The Compression Ratio CR, means that the compressed image is stored using only CR% of the initial storage size. It is defined as the number of elements in the compressed image divided by the number of elements in the original image.

*Peak Signal to Noise Ratio:*

Peak signal-to-noise ratio (PSNR) is a measure of the peak error in decibels. PSNR is meaningful only for data encoded in terms of bits per sample or bits per pixel. The PSNR is most commonly used as a measure of quality of reconstruction of lossy compression codecs. The signal in this case is the original data, and the noise is the error introduced by compression. When comparing compression codecs, it is used as an approximation to human perception of reconstruction quality, therefore in some cases one reconstruction may appear to be closer to the original than another, even though it has a lower PSNR (a higher PSNR would normally indicate that the reconstruction is of higher quality). One has to be extremely careful with the range of validity of this metric; it is only conclusively valid when it is used to compare results from the same codec (or codec type) and same content. The higher the PSNR, the better the quality of the reconstructed image. Typical values for lossy compression of an image are between 30 dB and 50 dB [2]. When the PSNR is greater than 40 dB, then the two images are indistinguishable. The PSNR can give an approximate index of image quality, but by itself it cannot make a comparison between the quality of two different images. It is possible, indeed, that an image with a lower PSNR might be perceived as an image of better quality compared to one with a higher signal to noise ratio.

## III. PROPOSED ANALYSIS

In this analysis the effect of changing the entropy encoding algorithms on the performance measures is analyzed. From the previous analysis it is clear that the best result is produced for number of encoding loops equal to 9. Thus, in this analysis the number of encoding loops is fixed as 9, the type of wavelet used is Haar and the encoding method is varied and its effect on compression parameters is

analyzed as shown in figure 2. Entropy Encoding Algorithms (Compression Algorithms) are divided in three categories: 1. Progressive Coefficients Significance Method (PCSM) 2. Coefficients Thresholding Method-1 (CTM-1) and 3. Coefficients Thresholding Method -2 (CTM-2). Under each category various algorithms are available and Table 1 shows various compression algorithms under each category. The compression algorithms are explained in detail in the following section.

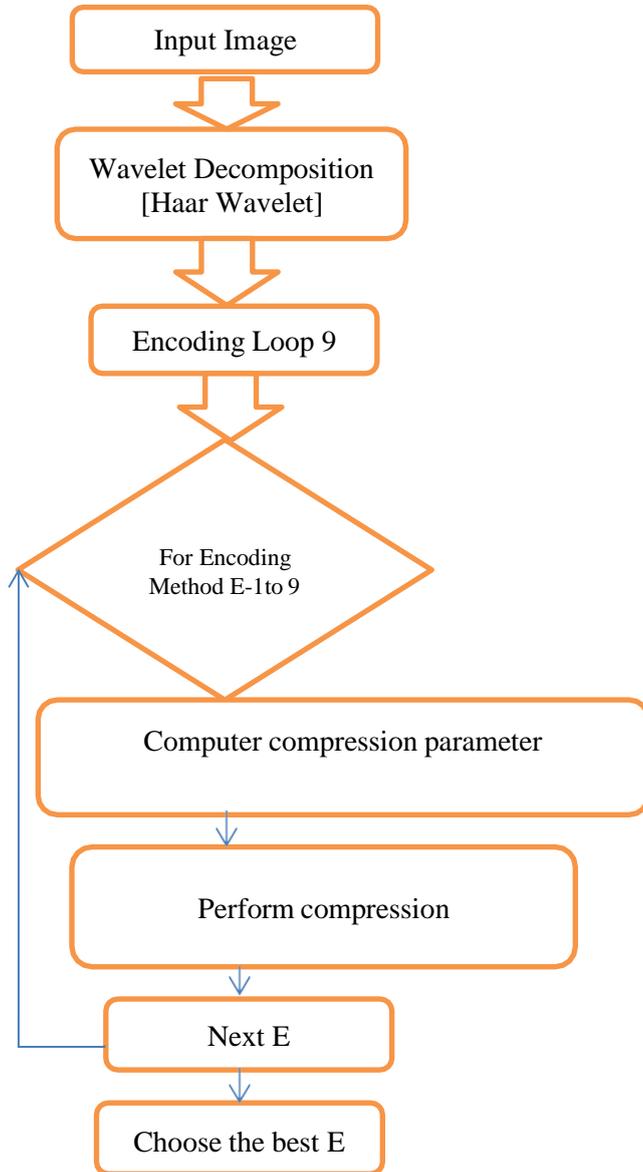


Fig. 2. Flow diagram of Proposed Analysis SPIHT

Pearlman were able to greatly improve the EZW algorithm, significantly increasing its compressive power. The SPIHT coder is a highly refined version of the EZW algorithm and is a powerful image compression algorithm that produces an embedded bit stream from which the best reconstructed images in the mean square error sense can be extracted at various bit rates. Some of the best results-highest PSNR values for given compression ratios -for a wide variety of images have been obtained with SPIHT.

Hence, it has become the benchmark state-of-the-art algorithm for image compression.

#### IV. SPATIAL-ORIENTATION TREE WAVELET

The Spatial Orientation Tree Wavelet (STW) employs a diverse approach in coding the information of zerotree. A zerotree have insignificant wavelet transform values at each of its locations for a given threshold T. Zerotree is a tree of locations in the wavelet transform with a root say  $[j, k]$ , and its descendants (children) located at  $[2j, 2k]$ ,  $[2j+1, 2k]$ ,  $[2j, 2k+1]$ , and  $[2j+1, 2k+1]$ , and each of their children, and so on. STW is more vigilant in its organization of coding outputs than the Embedded Zerotree Wavelet (EZW) [12] and SPIHT algorithm. In EZW, the root location is marked by encoding only one symbol for the output R or I as described in. Consequently, in EZW, the zerotrees provide narrow descriptions of the locations of insignificant values. The different approach used in STW is the use of a state transition model.

- Partial ordering by magnitude of 3 D wavelets transformed video with a 3d set partitioning algorithm
- Ordered bit plane transmission of refinement bits, and the SPIHT algorithm. It will be easier to explain SPIHT using the concepts underlying STW.
- The only difference between STW and EZW is that STW uses a different approach to encode the zero tree information. STW uses a state transition model. From one threshold to the next, the locations of transform values undergo state transitions.



Fig. 1 original black and white image



Fig. 2 Original image

The simulation results of image compression by applying the embedded zero tree Wavelet (EZW), Set Partitioning In Hierarchical Trees (SPIHT), Wavelet Difference Reduction (WDR), Spatial-orientated Tree Wavelet (STW),

Partitioning and Adaptively Scanned Wavelet Difference Reduction (ASWDR) algorithms various comparisons are obtained on the basis of PSNR and MSE and compression ratio (CR) values for the particular bit-per-pixel (BPP) ratio. For this purpose, we use the picture of pers. The original medical image is shown in Fig. and the compressed black and white images are shown in Figs.



Fig.3 E ZW



Fig.4 SPIHT



Fig. 5 STW



Fig 6 WDR



Fig.7 ASWR



Fig. 8 SPIHT

Table 1 and 2 show the values of PSNR and MSE for the different algorithms considered in this paper when CR and BP is approximately black and white image consider 1.3 and 0.3 respectively for TABLE 1 and for TABLE1 color image CR and BPP is 2.5 and 0.6 respectively.

Table 1. Pure-fractal image compression

| Exp # | Dpmain block size | PSNR  | CR    | Time(s) |
|-------|-------------------|-------|-------|---------|
| 1     | 2*2               | 52.2  | 6.33  | 256     |
| 2     | 4*4               | 41.10 | 26.11 | 58      |
| 3     | 8*8               | 34.12 | 130   | 13      |
| 4     | 16*16             | 30.22 | 692   | 1.99    |

Table 2 Numerical results of Wavelet-fractal image

compression algorithm

| Exp# | PSNR (db) | CR     | Time (s) |
|------|-----------|--------|----------|
| 5    | 29.82     | 333    | 105      |
| 6    | 36.66     | 24.22  | 726      |
| 7    | 31.88     | 101.21 | 78.66    |
| 8    | 32.55     | 65     | 12.21    |
| 9    | 37.15     | 17.2   | 156.33   |

Table 3 Overall results of implemented fractal methods

| Technique                  | PSNR  | CR    |
|----------------------------|-------|-------|
| Semi lossless pure Fractal | 52    | 6.11  |
| Lpsy Quality pure Fractal  | 30.23 | 690   |
| Wavelet fractal tree       | 36    | 16.99 |
| Wavelet fractal            | 31    | 333   |

Table 4 Compression eye image

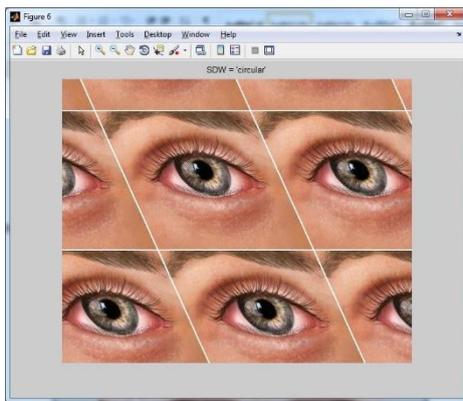
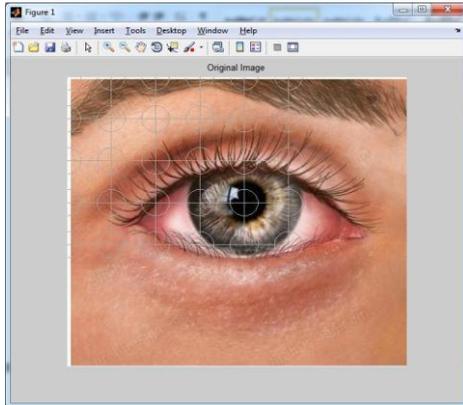
| Algorithm | MSE(%)  | PSNR(db) |
|-----------|---------|----------|
| EZW       | 18.1991 | 35.5303  |
| SPIHT     | 11.9975 | 37.3399  |
| STW       | 18.0313 | 35.5705  |
| WDR       | 18.1991 | 35.5303  |
| ASWR      | 19.1991 | 35.5303  |
| SPIHT_3D  | 11.9983 | 37.3396  |

Table 5 Original image

| Algorithm | MSE(%)  | PSNR(db) |
|-----------|---------|----------|
| EZW       | 10.1915 | 38.0484  |
| SPIHT     | 6.2820  | 40.1498  |
| STW       | 9.9473  | 38.1537  |
| WDR       | 10.1915 | 38.0484  |
| ASWR      | 10.1915 | 38.0484  |
| SPIHT_3D  | 6.2820  | 40.1498  |

In this paper, we have implemented and compared techniques for image compression. These algorithms are Embedded Zerotree Wavelet (EZW), Set Partitioning In Hierarchical Trees (SPIHT), Wavelet Difference Reduction (WDR), Spatial-orientated Tree Wavelet (STW), 3D-Set Partitioning In Hierarchical Trees (3D-SPIHT) and Adaptively Scanned Wavelet Difference Reduction (ASWDR). With the help of these algorithms each image is compressed and then decompressed. For the purpose to compare image quality, we consider MSE and PSNR as quality parameters. MAXLOOP is selected for compression algorithms on the basis of CR and BPP. We select MAXLOOP by keeping two things in mind that we require a low compression ratio and a better.

Calculated performance, comparisons amongst the algorithms are carried out. For a specific value of CR and BPP the results of SPIHT technique are best among all these techniques. It has low MSE and high PSNR values. By the help of these algorithms, we sustain good reproduction of the images as well as compression and also, we can preserve the image quality. In future, many methodological aspects like scale parameters, choice of the mother wavelet, threshold values etc of the wavelet technique will always require further investigations and can lead for enhanced outcome.



It is only conclusively applicable when it is used to compare results from an equivalent codec (or codec type) and same content. It is most simply outlined via the mean squared error (MSE) which for two  $m \times n$  images I and K wherever one in every of the images is taken into account a noisy approximation of the other is outlined.

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2$$

$$PSNR = 10 \cdot \log_{10} \left( \frac{MAX_1^2}{MSE} \right)$$

$$PSNR = 20 \cdot \log_{10} \left( \frac{MAX_1}{\sqrt{MSE}} \right)$$

Here,  $MAX_1$  is the maximum possible pixel value of the image. When the pixels are represented using 8 bits per sample, this is 255. More generally, when samples are represented using linear PCM with B bits per sample,  $MAX_1$  is  $2^B - 1$ . For color images with three RGB values per pixel, the definition of PSNR is the same except the

MSE is the sum over all squared value differences divided by image size and by three. For color images the image is transformed to a different color space and PSNR is reported alongside each channel of that color space.

*Compression Ratio (CR) and Bit-Per-Pixel (BPP):*

Compression Ratio (CR) provide the measure of achieved compression is given by the and the Bit-Per-Pixel (BPP) ratio. BPP CR and represent Bul. equivalent information. CR indicates that the compressed image is stored using CR% of the initial storage size while BPP is the number of bits used to store one pixel of the image. The initial BPP is 8 for a grey scale image. The initial BPP is 24 for a true color image, because 8 bits are used to encode each of the three colors (RGB color space). Confront of compression methods is to find the best compromise between a low compression ratio and a perceptual result.

*Peak Signal to Noise Ratio:*

The higher the PSNR, the better the quality of the compressed, or reconstructed image.

$$PSNR = 10 \log_{10} \frac{R^2}{MS}$$

Where,

- $I(x, y)$  is the original image.
- $I'(x, y)$  is the approximated version (which is actually the decompressed image)
- M, N are the dimensions of the images. (M=Width, N=height)

**ANALYSIS:** As shown in the above drawn comparison chart the PSNR value for the image-II by using wavelet difference reduction (WDR) is larger), so the chance of error is lesser. While, the PSNR value for the image-IV by using spatial time domain wavelet (STW) is lesser, so the chance of error is higher.

*Mean Square Error:*

The lower the value of MSE, the lower is the error.

$$MSE = \frac{1}{MN} \sum_{y=1}^M \sum_{x=1}^N [I(x, y) - I'(x, y)]^2$$

Where, R is the maximum fluctuation in the input image data type. For example, if the input image has a double-precision floating-point data type, then R is 1. If it has an 8-bit unsigned integer data type, R is 255, etc.

*Analysis:*

As shown in the above drawn comparison chart the MSE value for the image-II by using wavelet difference reduction (WDR) is lesser, so the chance of error is lesser. While, the MSE value for the image-IV by using spatial time

domain wavelet (STW) is larger, so the chance of error is higher.

*Comparative analysis of compression ratio for with respect to wavelets vs images compression ratio:*

The following formula is used to find the value of Compression Ratio:

$$CR = \frac{\text{Original Data}}{\text{Compressed Data}}$$

Or,

$$CR = \frac{\text{Actual BPP}}{\text{Reduced BPP}}$$

*Analysis:*

As shown in the above drawn comparison chart the CR value for the image-I by using spatial time domain wavelet (STW) is lesser, so the image will take less space for storage. While, the CR value for the image-IV by using wavelet difference reduction (WDR) is larger, so the image will take larger space for storage.

## V. RESULT

In this report, the results of four different wavelet-based medical image compression techniques are compared. The effects of different values like PSNR, MSE, BPP & CR are examined. The results of the different wavelet like EZW, WDR, SPIHT & STW are compared by using four parameters such as PSNR, MSE, BPP & CR values from the reconstructed image. These compression algorithms provide a better performance in picture quality at low bit rates. These techniques are successfully tested in many images. WDR technique provides high PSNR and low MSE values when compared to the EZW, STW & SPIHT technique.

## REFERENCES

- [1]. mr.jons lee Image Compression with Different Types of image using Wavelets technics IEEE Transactions on Communication, Control and Computing Technologies,
- [2]. S.P.Raja & Dr. A. Suruliandi, "Performance Evaluation on EZW & WDR Image Compression Techniques," IEEE Transactions on Communication, Control and Computing Technologies, vol.5, no.1, pp. 661-664, 2010.
- [3]. K.Nagamani & AG Ananth , "Study of ezw compression techniques for high and low resolution satellite imageries," IEEE Transactions on Computing, Communication and Networking Technologies, vol.1, pp.1-7, July.2010.
- [4]. S.P.Raja & Dr. A.Suruliandi, "Analysis of Efficient Wavelet based Image Compression Techniques," IEEE Transactions on Computing, Communication and Networking Technologies, pp.1-6, 2010.
- [5]. Zhenghua-Shu, Guodong-Liu, Lixin-Gan, Lin Zhang, "An Image Coding Combines
- [6]. H.264 Standard and Contourlet Transform," IEEE Transactions on Comm. Systems, Networks and Applications, pp.240-243, 2010.
- [7]. Li Huifang, Li Moa, "New Method Of Image Compression Based On Quantum Neural Network," IEEE Transactions on Information Science & Management Engineering, pp.567-570, 2010.
- [8]. Mohammad Babaeizadeh, Chelsea Finn, Dumitru Erhan, Roy H Campbell, and Sergey Levine. Stochastic variational video prediction. In *ICLR*,2017.
- [9]. Guha Balakrishnan, Amy Zhao, Adrian V Dalca, Fredo Durand, and John Guttag. Synthesizing images of humans in unseen poses. In *CVPR*,2018.
- [10]. Aayush Bansal, Shugao Ma, Deva Ramanan, and Yaser Sheikh. Recycle-gan: Unsupervised video retargeting. In *ECCV*,2018.
- [11]. Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3d faces. In *SIGGRAPH*, 1999.
- [12]. Adrian Bulat and Georgios Tzimiropoulos. How far are we from solving the 2d & 3d face alignment problem? (and a dataset of 230,000 3d facial landmarks). In *ICCV*,2017.
- [13]. Chen Cao, Qiming Hou, and Kun Zhou. Displaced dynamic expression regression for real-time facial tracking and animation. *TOG*,2014.
- [14]. Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *CVPR*,2017.
- [15]. Caroline Chan, Shiry Ginosar, Tinghui Zhou, and Alexei A Efros. Everybody dance now. In *ECCV*, 2018.
- [16]. Hamdi Dibekliog˘lu, Albert Ali Salah, and Theo Gevers. Are you really smiling at me? spontaneous versus posed enjoyment smiles. In *ECCV*,2012.
- [17]. Frederik Ebert, Chelsea Finn, Alex X Lee, and Sergey Levine. Self-supervised visual planning with temporal skip connections. In *CoRL*,2017.
- [18]. Patrick Esser, Ekaterina Sutter, and Björn Ommer. A variational u-net for conditional appearance and shape generation. In *CVPR*,2018.
- [19]. Chelsea Finn, Ian Goodfellow, and Sergey Levine. Unsupervised learning for physical interaction through video prediction. In *NIPS*, 2016.
- [20]. Zhenglin Geng, Chen Cao, and Sergey Tulyakov. 3d guided fine-grained face manipulation. In *CVPR*, 2019.
- [21]. Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *NIPS*, 2014.
- [22]. Artur Grigorev, Artem Sevastopolsky, Alexander Vakhitov, and Victor Lempitsky. Coordinate-based texture inpainting for pose-guided image generation. In *CVPR*, 2019.
- [23]. Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional

- adversarial networks. In CVPR, 2017.
- [24]. Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. In NIPS, 2015.
- [25]. Tomas Jakab, Ankush Gupta, Hakan Bilen, and Andrea Vedaldi. Unsupervised learning of object landmarks through conditional image generation. In NIPS, 2018.
- [26]. Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In ECCV, 2016.
- [27]. Diederik P Kingma and Max Welling. Auto-encoding variational bayes. In ICLR, 2014.
- [28]. Yahui Liu, Marco De Nadai, Gloria Zen, Nicu Sebe, and Bruno Lepri. Gesture- to-gesture translation in the wild via category-independent conditional maps. ACM MM, 2019.
- [29]. A. Nagrani, J. S. Chung, and A. Zisserman. Voxceleb: a large-scale speaker identification dataset. In
- [30]. INTERSPEECH, 2017.
- [31]. Junhyuk Oh, Xiaoxiao Guo, Honglak Lee, Richard L Lewis, and Satinder Singh. Action-conditional video prediction using deep networks in atari games. In NIPS, 2015.
- [32]. Joseph P Robinson, Yuncheng Li, Ning Zhang, Yun Fu, and Sergey Tulyakov. Laplace landmark localiza- tion. In ICCV, 2019.
- [33]. onneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, 2015.
- [34]. Masaki Saito, Eiichi Matsumoto, and Shunta Saito. Temporal generative adversarial nets with singular value clipping. In ICCV, 2017.
- [35]. Aliaksandra Shysheya, Egor Zakharov, Kara-Ali Aliev, Renat Bashirov, Egor Burkov, Karim Iskakov, Aleksei Ivakhnenko, Yury Malkov, Igor Pasechnik, Dmitry Ulyanov, Alexander Vakhitov, and Victor Lempitsky. Textured neural avatars. In CVPR, June 2019.
- [36]. Aliaksandr Siarohin, Stéphane Lathuilière, Sergey Tulyakov, Elisa Ricci, and Nicu Sebe. Animating arbitrary objects via deep motion transfer. In CVPR, 2019.
- [37]. Aliaksandr Siarohin, Enver Sangineto, Stéphane Lathuilière, and Nicu Sebe. Deformable gans for pose-based human image generation. In CVPR, 2018.
- [38]. Nitish Srivastava, Elman Mansimov, and Ruslan Salakhudinov. Unsupervised learning of video representations using lstms. In ICML, 2015.
- [39]. Hao Tang, Wei Wang, Dan Xu, Yan Yan, and Nicu Sebe. Gesturegan for hand gesture-to-gesture translation in the wild. In ACM MM, 2018.
- [40]. Hao Tang, Dan Xu, Wei Wang, Yan Yan, and Nicu Sebe. Dual generator generative adversarial networks for multi-domain image-to-image translation. In ACCV, 2018.
- [41]. Justus Thies, Michael Zollhofer, Marc Stamminger, Christian Theobalt, and Matthias Nießner. Face2face: Real-time face capture and reenactment of rgb videos. In CVPR, 2016.
- [42]. Sergey Tulyakov, Ming-Yu Liu, Xiaodong Yang, and Jan Kautz. Mocogan: Decomposing motion and content for video generation. In CVPR, 2018.
- [43]. Joost Van Amersfoort, Anitha Kannan, Marc’Aurelio Ranzato, Arthur Szlam, Du Tran, and Soumith Chintala. Transformation-based models of video sequences. arXiv preprint arXiv:1701.08435, 2017.
- [44]. Carl Vondrick, Hamed Pirsiavash, and Antonio Torralba. Generating videos with scene dynamics. In NIPS, 2016.
- [45]. Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Guilin Liu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. Video-to-video synthesis. In NIPS, 2018.
- [46]. Wei Wang, Xavier Alameda-Pineda, Dan Xu, Pascal Fua, Elisa Ricci, and Nicu Sebe. Every smile is unique: Landmark-guided diverse smile generation. In CVPR, 2018.
- [47]. Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In ACSSC, 2003.
- [48]. Olivia Wiles, A Sophia Koepke, and Andrew Zisserman. X2face: A network for controlling face generation using images, audio, and pose codes. In ECCV, 2018.
- [49]. Polina Zablotskaia, Aliaksandr Siarohin, Bo Zhao, and Leonid Sigal. Dwnet: Dense warp-based network for pose-guided human video generation. BMVC, 2019.
- [50]. Egor Zakharov, Aliaksandra Shysheya, Egor Burkov, and Victor Lempitsky. Few-shot adversarial learning of realistic neural talking head models. In ICCV, 2019.
- [51]. Yuting Zhang, Yijie Guo, Yixin Jin, Yijun Luo, Zhiyuan He, and Honglak Lee. Unsupervised discovery of object landmarks as structural representations. In CVPR, 2018.
- [52]. Long Zhao, Xi Peng, Yu Tian, Mubbasis Kapadia, and Dimitris Metaxas. Learning to forecast and refine residual motion for image-to-video generation. In ECCV, 2018.
- [53]. Michael Zollhöfer, Justus Thies, Pablo Garrido, Derek Bradley, Thabo Beeler, Patrick Pérez, Marc Stamminger, Matthias Nießner, and Christian Theobalt. State of the art on monocular 3d face reconstruction, tracking, and applications. In Computer Graphics Forum, 2018.