Range Imaging and Video Generation using Generative Adversarial Network

Anderson Stephanie

Natural and Applied Sciences, Duke Kunshan University, Jiangsu Province, China.

andersondukes@hotmail.com

Abstract – Latency, high temporal pixel density, and dynamic range are just a few of the benefits of event camera systems over conventional camera systems. Methods and algorithms cannot be applied directly because the output data of event camera systems are segments of synchronization events and experiences rather than precise pixel intensities. As a result, generating intensity photographs from occurrences for other functions is difficult. We use occurrence camera-based contingent deep convolutional connections to establish images and videos from a variable component of the occasion stream of data in this journal article. The system is designed to replicate visuals based on spatio-temporal intensity variations using bundles of spatial coordinates of occurrences as input data. The ability of event camera systems to produce High Dynamic Range (HDR) pictures even in exceptional lighting circumstances, as well as non-blurry pictures in rapid motion, is demonstrated. Furthermore, because event cameras have a transient response of about 1 s, the ability to generate very increased frame rate video content has been evidenced, conceivably up to 1 million arrays per second. The implementation of the proposed algorithms are compared to density images recorded onto a similar gridline in the image of events based on the application of accessible primary data obtained and synthesized datasets generated by the occurrence camera simulation model.

Keywords - Generative Adversarial (GA), High-Dynamic Range (HDR), Stacking-Based Time (SBT), Stacking Based Events (SBE)

I.   INTRODUCTION

Bio-inspired biosensors, known as event camera systems, receive visual data at the same time as the visual system does. While conventional cameras broadcast intensity images at a rate that is constant, even sensors transfer the information regarding the changes of intensity in the structure of asynchronous occurrences that provide space-time dimensions. They offer a number of benefits over conventional cameras, including millisecond responsiveness, great temporal resolution (about 1 sec), and a broad bandwidth [1]. Most current algorithms cannot be applied successfully to events sensors because their outputs are sequencing of asynchronous occurrences across time rather than real intensity pictures. Although incident cameras have recently been proven to be adequate for certain applications such as 6-DoF positioning applications and image restoration, the ability to produce intensity pictures from occurrences for other object classification recognition, tagging, and SLAM would have been very beneficial. In fact, event cameras are said to provide all of the information required to rebuild pictures or a complete video stream in theory. This claim, nevertheless, has never been proven. We addressed the issue of yielding pixel intensities from occurrences and then further leverage the benefits of event camera systems to make maximum High Dynamic Range (HDR) image and video files with minimal movement artifacts that is incredibly significant when strength to rapid sequence and excessive illumination changes is essential, such as in autonomous vehicles.

To our understanding, this is the first effort to translate pure occurrences to HDR images and higher-frame videos per second video, demonstrating, which occurrence sensors can potentially generate quality non-blurred pictures and movies even in rapid movement and severe lighting situations. We showcase an occurrence-based domains translations system that, when compared to active pixel sensor (APS) panels and other existing techniques, produces higher-quality pictures from occurrences. Two novel and action event stacking techniques rely on switching over the occasion flow, Stacking Based Time (SBT), including aspects of palletization based on the occurrences. Using the layering methodologies, it is evident that a video with approximately 1,000.000 frames in a single second may be generated. We perform extensive tests and evaluation/comparisons to validate the robustness of the suggested techniques. Real records from the DAVIS combined event and illumination cameras, a dynamically and active-pixel vision sensors, were utilized in the studies. The device's pixel grid-line of occurrences and brightness are in the same place, reducing the number of additional stages of corrections and distortion required to align two pictures. To train a general framework for occurrence-to-image and the translation of videos, we showcase an open set of data including more than images that are 17K captured by camera. We also used the incident camera simulator to create a synthetic dataset with 17K pictures for testing. Section II reviews the past literature texts regarding the aspect of range imaging and video generation using the Generative Adversarial (GA) network. Section III focuses on the research methodology for the analysis. Section IV focuses on a critical analysis of the paper. Section V provides a discussion of the experiments. Section VI concludes the research.

II.   LITERATURE REVIEW

Y. Wang and H. Jiang in [2] focused their research on the reconstruction of intensity-images for different occurrences. Researchers' work, in which cyclically linked regions termed maps were used to understand intensities and optical flow, was one of the first efforts at optically analyzing or recreating the intensity picture from occurrences. To follow the camera, the researchers utilized pure occurrences on rotation-only images and created a super-resolution realistic composite of the environment using stochastic filtration.

According to R. Kaushik and J. Xiao in [3], intensities pictures were recovered in the noisy environments using a patch-based minimal thesaurus on both generated and real-world data. In contrast to earlier spinning methods, P. Ghasemi and M. Ghafoori in [4] went a step beyond by recreating the intensity picture as well as the movement field for generic mobility. Meanwhile, scientists developed a discretization denoising methodology that filters inbound events repeatedly. They reconstructed the picture by guiding the events via a manifold of times and dates. Scientists suggested the observations and computations on the event cameras with RGBW polarizing filters. For recreating the intensity picture, they proposed a naïve and computational approach. The techniques described above did produce high-intensity pictures, mostly of pure events. The recon design, on the other hand, was not naturalistic.

P. Shedligeri and K. Mitra in [5] have unveiled a hybrid technique for creating photorealistic pictures that combines intensity photos and events. Three autoencoders are used in their approach. This technique works well for typically lit situations, but it fails to reconstruct HDR images in severe lighting conditions because it only applied the occurrence datasets to identify the 6-DoF posture.

Deep learning was investigated by researchers. Though deep learning has yet to be widely deployed occurrence-based visions, many researches have shown that it can perform well with data sets. Researchers trained a convolutional neural network (CNN) to regulate the driving of a predatory robot using both data points and APS pictures. Other techniques of steered predictions for self-driving vehicles have been investigated, including utilizing clear occurrence and integrating APS pictures in an end-to-end manner. A stacking spatial LSTM system, on the other hand, was suggested in [6] that delocalize 6-Dof positions from the occurrences, and the vision flow forecasts based on the application of supervised decoder and encoder infrastructures that have been recommended in [7], which makes use of the learning algorithms to produce pseudo labels that identify objects in the ego movement. By retraining a CNN on APS pictures, the pseudo categories are translated to the event image. In addition, as stated in the preceding part, authors in [8] presented the merging of event data with APS images, which used autoencoders to generate photorealistic images. We would be the first to use deep convolutional networking to occurrence data, to our understanding.

In fact, conditional GANs (cGANs) are used to precondition image interpretation. However, no subjective study has been done on the efficacy of cGANs on event data. cGANs have previously been used to predict images from a normal map, forecast future frames, and generate images from sparse annotation. The distinction between utilizing GANs conditional and unconditional for image-to-image translations is that unrestricted GANs depend heavily on the limiting lost functions to regulate the conditioning output. In the frame source image, cGANs have been effectively used to extracting features, with these implementations mostly focusing on transforming images from a single representation to another based on the controlled environment.

Furthermore, it necessitates output-input pairs from the graphical tasks whereas assuming the connections of domains. cGANs are yet to be studied subjectively and empirically in the area of event vision. As a result, we want to investigate whether cGANs can be used to rebuild images from event data. Nevertheless, since the basic method to frame-based image retrieval differs from the occurrence-based approach, we showcase an approach for deep learning to complete this task and fundamentally utilize the merits of the occurrence sensors, e.g. the zero latency, high-resolution and frame interpolation. The suggested approach is then evaluated subjectively and numerically using actual and generated data. Section III presents an evaluation of the methodology used for this research.

III.   METHODOLOGY

We use presently existing deep neural networks, including cGANs, as proposed options for events vision to rebuild HDR and greater spatial precision pictures and movies from events. cGANs are procedural algorithms that train a mapping from the input picture x to the image pixel y, G: x, zy. An antagonistically trained differentiator, D, trains the generators G to generate output that cannot be distinguished from the source pictures. The goal is to reduce the gap between the underlying data and the generator's outputs while increasing the number of observations from the determiner. Image-to-image translations has been shown using cGANs like as Pix2Pix and CycleGANs, resulting in ground-breaking results. The primary advantage of cGANs is that they don't need customizing the loss function for particular tasks, and they can adjust their learnt loss to the dataframe where they're trained. Nevertheless, since event data differs significantly from that utilized in conventional cGAN-based vision techniques, we offer novel ways for providing off-the-shelf artificial neural input.

Stacking Events

Every occurrence is indicate as (p, t, v, u) as tuple within the camera events whereas v and u represent the dimensional pixels, t represents the time of the events, and p, which is the same as 1 represents the events' polarization that might be the direction of illumination various (p equals to zero for no events).

On Fig 1, these occurrences are represented as a stream. We synced APS pictures and asynchronous occurrences between two successive APS frames based on the frames per second of the intensity camera. New formulations of event logs are needed to compute occurrence information into networks. To present enough occurrence data for picture

Journal of Biomedical and Sustainable Healthcare Applications 1(1)(2021)

reconstructing, one easy approach is to formulate the three-dimensional event volumes as p (t, v, u) over a particular duration of time. The dimensions of the 3D volumes is (n, h, w), whereby h and w represent the density of the pixel of the event camera and n is equal to td/t. Whenever the temporary resolution of the event camera is represented by t and the timescales by td, the size of the image depth is n, h, w. This is the same as providing the network with an n-channel picture input. This visualization keeps track of all event details. The issue is that there are so many channels. When td is reset to 10 ms, for instance, n is approximately 10K that is a very high number given that an event camera's temporal resolution is approximately 1 s. As a result, we create the three-dimensional occurrence volume with small n by stacking and merging the occurrence within a minimal timeframe. Occurrence stacking might be achieved in different means, but it integrates sacrificing the temporal data of the occurrences.

Stacking based Time (SBT)
This methodology integrates that streaming occurrences between the timeframe references of two distinct intensities of pixel from the occurrence camera denoted by t. Not each event, nonetheless, is aggregative into the same framework. Instead, the temporal timeframe of the occurrence stream has been separated into n equal-scale segments, and the monochrome photographs denoted by n, SI p (v, u), 1 = (1, 2, n) are produced by integrating the occurrences within each interval of time (I) tn. SIp (v, u) is the overall assumption of polarization readings (p) at v, u. The grayscale frameworks are stacked to create a single stack Sp (v, u, 1 = Sip (v, u), 1 = 1, 2,..n that is integrated within the network. As stated earlier on, this approach of stacking loses its temporal data on occurrence within the timeframe t n. Nonetheless, the stacking saves some temporal information as a sequence of frameworks numbered 1 through n resultantly; the greater n enables more temporal information storage. Fig 2 indicates the occurrence layered and integrated. The stacking can be considered as pseudo color framework whenever n is equal to 3 (focusing on integrating the frames FC, FB, and FA within the same stack as shown in Fig 1. With respect to the time presented as the case manifolds within the middle segment of Fig 2, the APS images positions is about the part of the third-red rectangles approximately 0.03 seconds beginning from zero time on the three-dimensional format (the framework rate on the APS photo is approximately 32 FPS.

Stacking Based Events (SBE)
However, due to the event camera, SBT has an intrinsic restriction: no events whereas the environment or camera remains. When the occurrence data inside the timeframe is inadequate for image restoration, it is possible to get effective HDR images. This is true for the 5th and 4th frames of occurrence stream in Fig 1.

Another critical concern is because there are various occurrences in a single timeframe, as showcases in the 3rd timeframe. SBE is linked closely with the condition of time-asynchronous incident sensor, and can possibly alleviate the limitations of SBT. This approach, as visualized in Fig 1, establishes a frame by incorporating the occurrences with respect to the amount of incoming occurrences. The 1st Ne events are integrated into the first frame, the following Ne events into Fig 1, etc., to formulate a single n frame stacks. The network therefore receives the n frame stack that has an overall occurrence of nNe. This approach ensures that the required occurrence dataset is available to reestablish the images based on the Ne figures. The panels FH, FG, FF, FE in Fig 2 signifies 4Ne, 3Ne, 2Ne and Ne respectively with respect to side views. Because we count the occurrence number with a specific timeframe, we can possibly change the occurrences amount in every screen including single stacks.

All through this article, the occurrence polarities (plus, minus) is expressed by two major color itemsets: (Red (+), Blue (+) and (Green (+), Cyan (+). The yellow highlighting time is used to illustrate two forms of stacks (SBT – left and SBE – right) in the basic three-dimensional format.
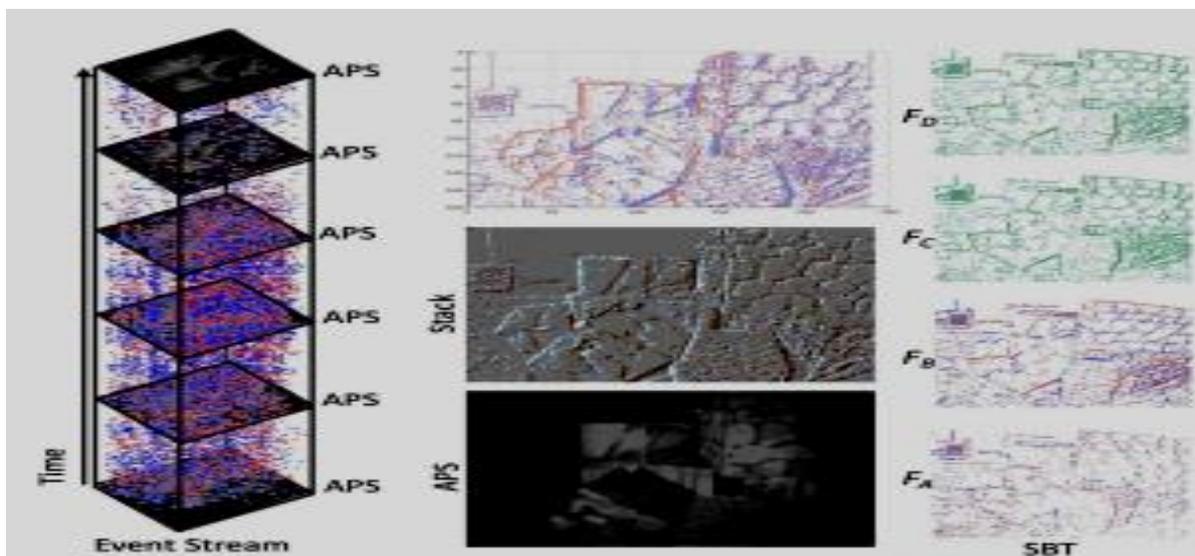
Fig 1. The incident channel and stack generation by SBT and SBE
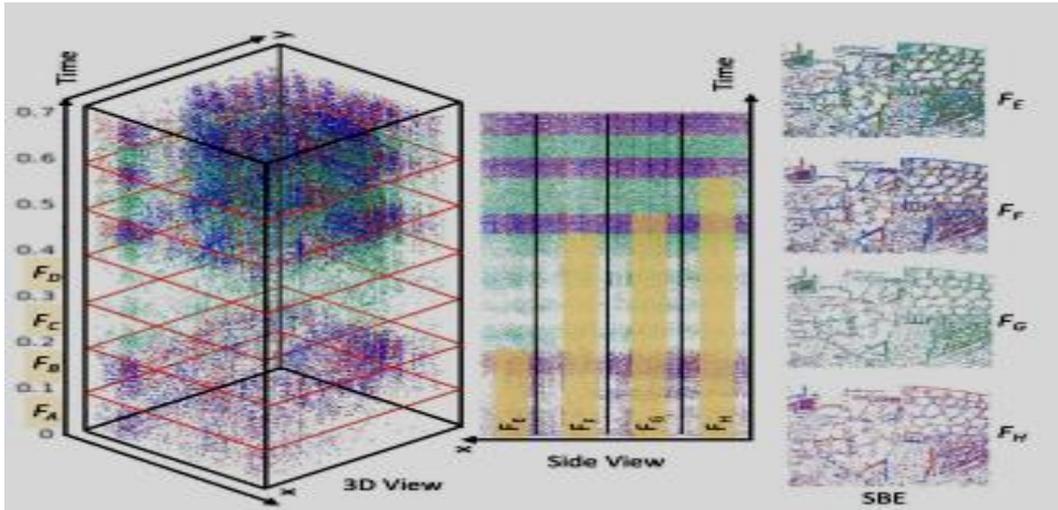
Fig 2. The 3D perspective and its side profile are colored with (Red, Blues) and (Green, Cyan) every 5,000 occurrences "HDR squares" series provided all of the photos and displayed data; for better viewing

Stacking of Video Restorations
SBT and SBE may be used to reassemble video from occurrences using the suggested networks, and the data stream of the resultant video can be modified in both approaches by altering the degree of time change between two consecutive event stacking utilized as an output within the scheme. Whenever the occurrences in the timeframe interval denoted by It, (I focused on generating one input stacks for images I; I within the video; occurrences in the intervals of time I t′, I + ts with the duration transition ts can be applied to produce the upcoming input stacks for images I + ts. The output rate of frame for the video is now I ts. It is fundamental to note that two stacking with 't' have a considerable overlapping timeframe I 't', I. The chronological consistency for surrounding frames is automatically maintained if t ′ > ts. It is possible to accomplish about 1,000,000 FPS videos within the periodic constancy using an event sensor with a time resolution of around one second.

IV.   CRITICAL ANALYSIS

Network Architectures
We discuss our classifier and producer in this article, which were inspired by [9]. Fig 3 and Fig 4 provide further information about the structures, particularly the size of every level.

Design of the Producer
The heart the event to the image transition is assessing how to change the variation case inputs to the HDR results with the data, which shares systemic photo elements such as blobs, vertices, and margins. For the image to image access tasks, the encoder and decoder framework is common utilized as a network. To retrieve a translation output, the input is downsampled using the framework before being upsampled. Because there is so much high-frequency critical data from events information transmitted through the networks in the event-to-image translations issue, it is probable that specific characteristics of events will be lost throughout the procedure, causing noise in the outputs. As a result, we examine the methods where we extend the "U-net" network topology in [10] by adding attention mechanisms. The number of hidden layers and inputs/outputs are shown in detail in Fig. 3.

Classifier Structure
 Structure of classifier in [11] is the source of our network. Our network design is shown in detail in Fig 4. Our discriminator may be thought of as a way to reduce the pattern classification loss between intensities picture and events.
     Using a mathematical approach, the objective element is illustrated as:

$$L_{c\,GAN}(G,D) = E_{e,g}[\log D\,(g,e)] + E_{e,e}[\log\,(1 - D(e\,G\,(\,e, \in)] \qquad (1)$$

     Whereby "e" denotes the original occurrence, "g" represents the generative picture, while "ǫ" shows the Gaussian disturbance as the input to generators. In the meantime, G focuses on the variation of the pictures from the occurrences, and D minimizes it. In this case, for the purpose of regularization, L1 utilized in the blurred shrinking as:

$$L_{L1}(G) = E_{e,e}[g - G(e, \in)\|1\|] \qquad (2)$$

The L1 type is focused on making the discriminator more integrated on the architectural frequency of the produced images from their occurrences. The main purpose is to evaluate the overall losses from the "occurrence-to-images" translations as:

$$G^* = \text{ag} \min_G \max_G [L_{cGAN}(G, D) + \lambda L_{L1}(G)] \tag{3}$$

whereby λ represents the parameters to effectively adjust the rate of learning. With the disturbance "ǫ", the network might study a mapping aspect from the occurrence "e".
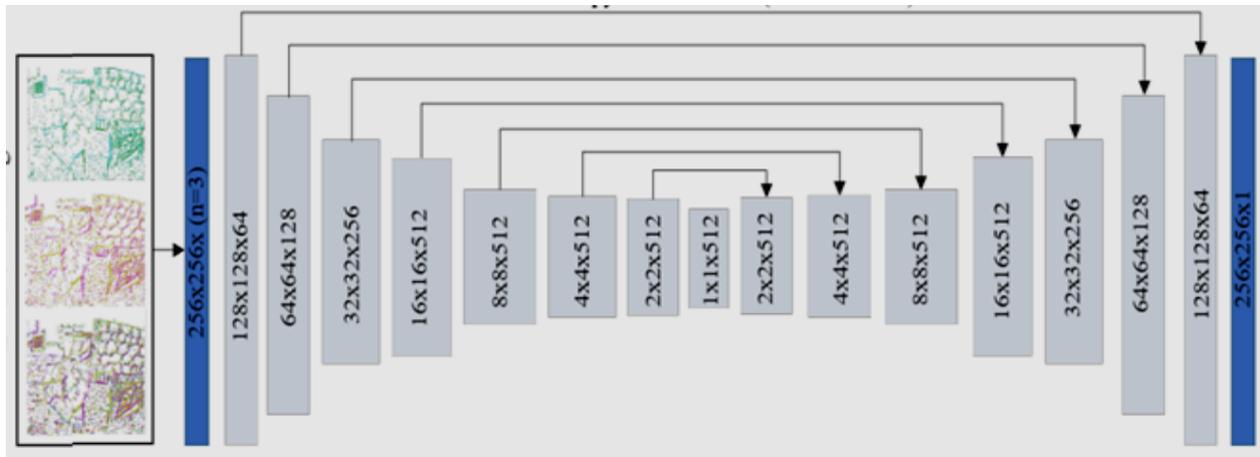
Fig 3. The network of the generator

The U-Network system, which takes inputs with the various 256 by 256 by n dimensions (n = 3 for instance), succeeded by the gray box that corresponds to te channel feature map. The overal channel number is represented within every box. The values (in the first two bozes) from indicate the dimensions of the filters whereas the final number shows the filters' number.
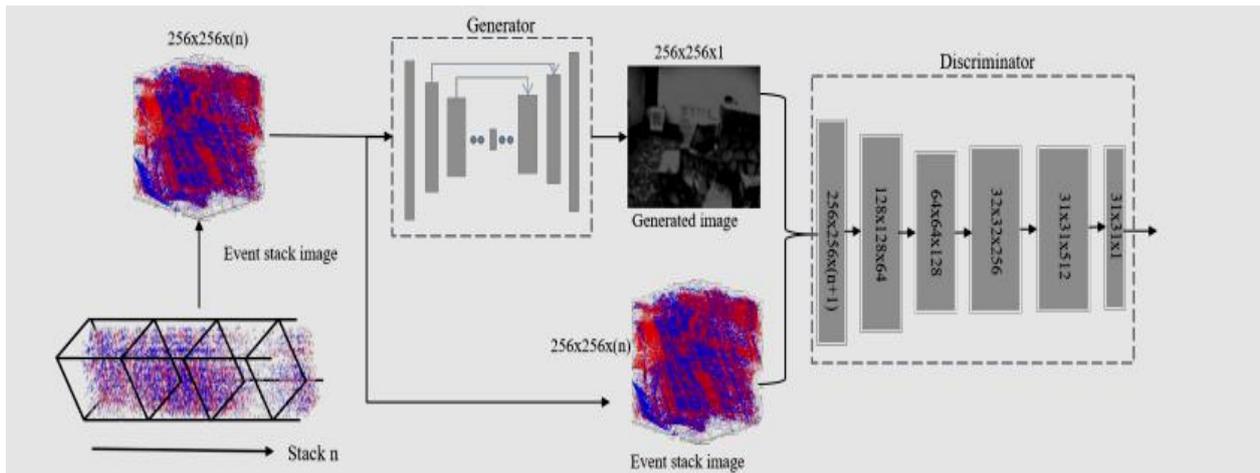
Fig 4. The projected model with the discriminator and generator network

The network is the same as the PatchGAN that considers two pictures (initial APS and the picture produced by the generators from the occurrences (see Fig 4). The classifier constructs the circumstance of attribute layouts from the generator's last stack and determines whether the produced photograph satisfies the circumstance of database transmission from the occurrence to frequency to "g" that can complement the event-based dispersion and aid in the production of more predetermined outcomes.

Processing of Datasets
Our learning and testing databases are created using three different approaches. The first collection of datasets is created to contains numerous real-world situations. We also create the second session of the dataset on our own for different learning and testing reasons, as well as for later public release. The records were taken using a DAVIS camera and include a large number of situations. ESIM, an open-source incident camera simulation, generates the third kind of datasets. Many distinct

interior and exterior views were recorded with numerous revolutions and conversion of the DAVIS camera in the actual data. Our classification model consists of sets of stack events as described as well as APS images from both actual-world cases and ESIM ground truth images. To train the model with actual information, we cautiously arrange the learning data to prevent the system from obtaining incorrect APS frame characteristics. APS images are affected by the movement artifacts in a rapid motion and integrate a restrained vibrant range amounting to a loss of data. Because our obligation is to generate HDR images with minimal blur by completely leveraging the merits of occurrence sensors, typically utilizing real APS images as underlying dataset is not a considerable approach to train networks. Resultantly, the occurrences dedicated to the training datasets' white and black segments are removed from the inputs, permitting networks to learn to structure pictures from events. Furthermore, APS pictures are categorized as non-blurred or blurred depending on BRISQUE ratings (which will be described later) and human assessment, and distorted APS pictures are not included in the training set of data. The sequence of modelling is mostly formulated based on the application of ESIM, in which occurrences are produced as a digital camera travels in all dimensions to record various situations in supplied pictures. The APS images are immediately counted as the underlying data for picture restoration since the occurrences and APS pictures are produced in a supervised simulated environment. As a result, for the generated dataset, the abovementioned learning information modification is not needed.

V.   EXPERIMENTATIONS

Observations and experimentations

We run extensive tests on the dataset as well as a comparative database containing three actual sequences (Face, leaping, and ball). We build training data of approximately 60K event layers with matching APS picture pairings depending on exact date stamps, and we evaluate our approach on both regular and HDR scenarios. We selected 1,000 APS or ground reality pictures with associated signal layers for evaluation at arbitrary from both the actual and generated sets. Because actual databases do not integrate the ground truth images for validation and training, we use the APS images instead. Nevertheless, frame rate and a limited contrast ratio characterize the APS picture. As a consequence, training and assessing outcomes using APS pictures may not be the ideal option. As a consequence, we generate the training APS pictures, and evaluate the outcomes using the structural similarities (SSIM), features similarities (FSIM), and the no-reference grade measurement.

The Blind/Referenceless

Image Spatial Quality Evaluator (BRISQUE), which uses standardized chrominance coefficient to evaluate the genuineness in photographs, is used to achieve a comprehensive quality, particularly when evaluating service effectiveness of reconstruction of real - world datasets without underlying data. For different datasets produced using ESIM, on the other hand, high accuracy is linked with the associated rebuilt picture with the nearest timeframe. To assess non-HDR images and situations for which we have accurate contextual information, we just use Peak Signal to Noise Ration, FSIM, and SSIM.

SBT vs. SBE

Using our actual sets of data, we examine and contrast two event stack techniques: SBT and SBE. For learning, we utilized 17K events stack-APS picture pairings, with t set to 0.03s for SBT and 60K for SBE. The number of pixels (n) in one layer is fixed to 3 for each technique to vividly show the impact of the layering approach. For quantitative comparisons, Fig 5 displays recreated pictures on our real life datasets based on the application of SBT and SBE. Our techniques (both the SBT and SBE) are demonstrated to be capable of reconstructing pictures in a variety of patterns, and the produced visuals are the same to the APS images utilized as the ground truth. Our techniques were effective in recreating human forms, building appearances, and so forth. SBE outperforms SBT in most cases. The results obtained from utilizing SBE are shown in Table 1.
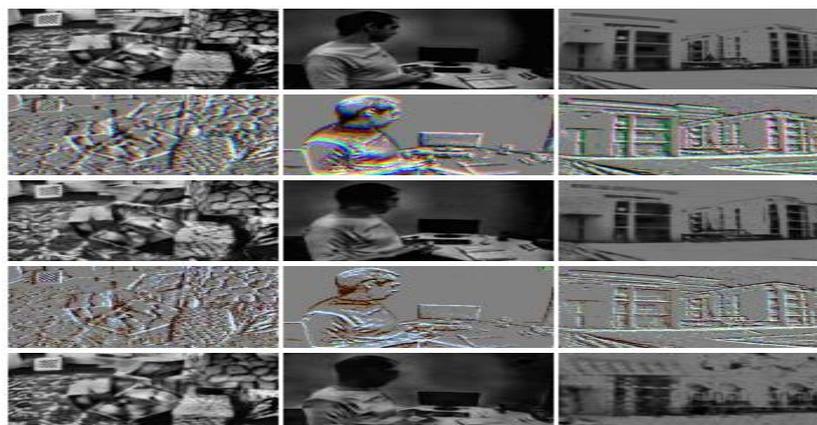
Fig 5. Reconstruction findings based on the input occurrence stack (visualized based on the pseudo colored image) on various actual-world patterns.

To the bottom from the top, APS pictures as the absolute truth, occurrence stack utilizing SBE, reconstructed pictures with SBE occurrence stack utilizing SBT, are restructured images based on the SBT application.

Table 1. Quantitative analysis of SBE on the actual sets of data

	SSIM	FSIM	BRISQUE
Ours (n=3)	0.7 +/- 0.2	0.9 +/- 0.1	37.8 +/- 4.9

Quantitative Assessment with Simulated Data
It is worth noting that the higher the SSIM and FSIM numbers in Table 1, the greater the production grade, since they simply show how close the pictures are to APS frames with motion artifacts and poor contrast ratio.

Statistical Assessment Using Model Results
We examine the capability of our method on the real life dataset that indicates that SBE is significantly resilient compared to SBT. As a result, we perform tests using SBE and demonstrate the resilience of our techniques using databases from ESIM, which may provide huge amounts of trustworthy data sources. Because the simulator generates noise-free APS images with occurences that similar for every imageset, the APS images might be utilized as underlying data, allowing statistical evaluation of the findings. Furthermore, whereas our technique is competent of combining any number of images (n) into a clump, we chose n = 1, 3 to investigate the impact of various channel counts. One stack may hold up to 60K events. With n = 1 and n = 3, Table 2 depicts the quantitative assessment of our approach. It is demonstrated that our approach performs better with n = 3 than with n = 1, demonstrating that collecting more images in one layer increases efficiency since it can retain more temporal features.

Fig 6. Reconstructed output from the input produced by ESIM

Utilizing three frames in a single track amounts in multiple robust reconstructions when compared to a single frame stack whereby images are disturbed because of occurrences that are over accumulated (n = 1).

Fig 7. Methodological comparison of images 1, 2, 3

Our methodlogy generates more information such as jumping pose, beard, face etc, including the natural gray variation in minimal textual segments.

Table 2. ESIM experiments (the simulator sets of data)

	SSIM	FSIM	PSNR(dB)
Ours (n=1)	0..7 +/- 0.2	0.8 +/- 0.1	20.5 +/- 2.9
Ours (n=3)	0.8 +/- 0.1	0.9 +/- 0.1	24.9 +/- 3.2

With more frames in a single stack produces more effective results. A few recovered pictures, as well as raw events stack and ground reality frames, are shown in Fig 6. One thing to note is that the n = 1 recreated face and the summit of the structure are somewhat deformed, which may be caused by too many occurrences collected in one single network.

References to Comparable Works

Reference to comparable works shows more difficult scenarios created in collaboration with the GT. In Fig 7, we examine the outcomes of manifold regularization (MR) and intensity estimation (IE) on the sequence (face, leaping, and ball) [12]. Because we are dealing with extremely volatile information, the additional film, which displays the whole series of numerous hundred images, provides a more compelling and clear description and findings. Because no ground-truth picture is provided for these episodes, we utilize the BRISQUE score to evaluate the results statistically. In Table 3, we make a comparison of our methodology (SBE, n=3) on the sequences (ball, leap, face) with IE and MR. The outcome is very impressive. The given figures are the norm and average variation of the BRISQUE metric used to all sequences of recovered images. In all cycles, our approach yields higher BRISQUE ratings.

Table 3. Comparison of our approach on patterns (face, leaping, and ball)

Sequences	Ball	Leaping	Face
Bardow	29.4 +/- 8.8	29.4 +/- 7.3	22.3 +/- 8.8
Munda	35.0 +/- 9.3	48.2 +/- 6.7	27.3 +/- 7.3
Ours (n=3)	39.2 +/- 3.5	48.3 +/- 2.2	48.3 +/- 3.1

VI.   CONCLUSION

We showed how the features of event camera systems may be used to effectively reconstitute the HDR non-blurred images and potentially maximize the rate of frame videos from the occurrences using our cGANs-based technique. For both image / video reconstructions from occurrences using the networks, we originally suggested two incentive item layering approaches (SBT and SBE). We then used tests centered on our databases of online accessible real-world series and simulations to demonstrate the benefits of using events camera systems to make high dynamic range photos and high frames per second films. To demonstrate the durability of our methodology, we contrasted our cGANs-based incident paradigm to other contemporary modeling techniques using publicly available datasets and found that our models outperform them. We also demonstrated that high dynamic range photos can be produced even in severe lighting circumstances, as well as non-blurred photographs in rapid movement.

References

[1]. D. Chrysostomou, G. Sirakoulis and A. Gasteratos, "A bio-inspired multi-camera system for dynamic crowd analysis", *Pattern Recognition Letters*, vol. 44, pp. 141-151, 2014. Doi: 10.1016/j.patrec.2013.11.020.

[2]. Y. Wang and H. Jiang, "Three-dimensional reconstruction of microscopic images using different order intensity derivatives", *Optical Engineering*, vol. 54, no. 2, p. 023103, 2015. Doi: 10.1117/1.oe.54.2.023103.

[3]. R. Kaushik and J. Xiao, "Accelerated patch-based planar clustering of noisy range images in indoor environments for robot mapping", *Robotics and Autonomous Systems*, vol. 60, no. 4, pp. 584-598, 2012. Doi: 10.1016/j.robot.2011.12.001.

[4]. P. Ghasemi and M. Ghafoori, "Holden in Search of Identity: Recreating the Picture of theFlâneur", *English Studies*, vol. 91, no. 1, pp. 74-88, 2010. Doi: 10.1080/00138380903355098.

[5]. P. Shedligeri and K. Mitra, "Photorealistic image reconstruction from hybrid intensity and event-based sensor", *Journal of Electronic Imaging*, vol. 28, no. 06, p. 1, 2019. Doi: 10.1117/1.jei.28.6.063012.

[6]. Preethi D. and N. Khare, "EFS-LSTM (Ensemble-Based Feature Selection With LSTM) Classifier for Intrusion Detection System", *International Journal of e-Collaboration*, vol. 16, no. 4, pp. 72-86, 2020. Doi: 10.4018/ijec.2020100106.

[7]. A. Jaad and K. Abdelghany, "Modeling urban growth using video prediction technology: A time-dependent convolutional encoder–decoder architecture", *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 5, pp. 430-447, 2020. Doi: 10.1111/mice.12503.

[8]. Z. GUO, Z. CHEN and J. ZHANG, "Event-based merging of partial behavior models", *Journal of Computer Applications*, vol. 30, no. 1, pp. 266-269, 2010. Doi: 10.3724/sp.j.1087.2010.00266.

[9]. M. Fairhurst and M. Mattoso Maia, "Performance comparisons in hierarchical architectures for memory network pattern classifiers", *Pattern Recognition Letters*, vol. 4, no. 2, pp. 121-124, 1986. Doi: 10.1016/0167-8655(86)90033-4.

[10]. S. Park and Y. Shin, "U-Net-Based Generative Adversarial Network", *Journal of the Institute of Electronics and Information Engineers*, vol. 58, no. 5, pp. 61-67, 2021. Doi: 10.5573/ieie.2021.58.5.61.

[11]. Y. Liu, L. Wang and M. Sun, "Efficient Heuristics for Structure Learning of k-Dependence Bayesian Classifier", *Entropy*, vol. 20, no. 12, p. 897, 2018. Doi: 10.3390/e20120897.

[12]. X. Tu et al., "MR image segmentation and bias field estimation based on coherent local intensity clustering with total variation regularization", *Medical & Biological Engineering & Computing*, vol. 54, no. 12, pp. 1807-1818, 2016. Doi: 10.1007/s11517-016-1540-7.