

FedCSIS_2023_paper_1464.pdf

FracRIFE: Fractional-Order Real-Time Intermediate Flow Estimation for Video Frame Interpolation

Malwina Kubas, Grzegorz Sarwas, Witold Czajewski

Institute of Control and Industrial Electronics

Warsaw University of Technology

ul. Koszykowa 75, 00-662 Warszawa, Poland

Email: {malwina.kubas.stud, grzegorz.sarwas, witold.czajewski}@pw.edu.pl

42

Abstract—Video frame interpolation, or VFI, is one of the critical tasks in image processing, and it can be applied to a wide range of problems. It enhances video sequences by seamlessly increasing the number of frames, resulting in improved video quality and smoothness. It is also useful for creating slow-motion videos and can be applied in video compression and reconstruction algorithms. However, increasingly popular real-time and mobile applications require quick and computationally light as well as high-quality solutions. This research paper introduces the FracRIFE algorithm, a modified version of the RIFE (Real-Time Intermediate Flow Estimation) method, where a novel fractional-order Lucas-Kanade optical flow solution replaces the original optical flow calculation based on the IFNet. Through benchmark evaluations and comparisons with leading algorithms in the field of VFI, the proposed method demonstrates superior processing speed compared to RIFE, while maintaining comparable quality. Moreover, the proposed fractional-order Lucas-Kanade variant surpasses its traditional first derivative predecessor in terms of quality, albeit it is slightly slower.

4

1. Introduction

Video frame interpolation, or VFI, is the task of synthesizing intermediate frames between two consecutive frames of a video. It is generally used to increase the frame rate and improve the visual quality and perception of videos. In particular, VFI can be applied to a wide range of problems, such as video compression [1], generating slow-motion videos [2], increasing the frame rate of (cloud-based) video games [3], novel view synthesis [4], missing frame restoration [5] or super-resolution [6]. Some of these applications require real-time performance, but they can effectively reduce the costs of constructing high-speed hardware or providing services to users with limited resources of bandwidth.

VFI algorithms need to cope with complex, large, and non-linear motions and illumination changes of the real-world scenes. One common approach to the VFI problem is based on a hybrid solution comprising initial warping of the input frames according to an estimated optical flow and final fusing and refining these warped frames with Convolutional Neural Networks (CNNs). Some of the recent flow-based

VFI algorithms have shown remarkable results [7], [8], [9], [10], [11]. Flow-based algorithms can be classified as back-warping (commonly used) and forward-warping (rarely used). They calculate bidirectional optical flow to generate an intermediate flow, which is computationally expensive process, and such solutions are not suitable for real-time applications.

In this paper, the authors propose a novel method for video frame interpolation based on the fractional-order optical flow algorithm proposed by González-Acuña *et al.* in [12]. This hybrid method is expected to be more efficient than methods that use neural network models for bidirectional optical flow estimation. The main contribution of this work is the feasibility study of fractional-order derivative application in the VFI domain: can it be helpful in developing a fast, high-quality method for video frame interpolation, which can be implemented on many devices and run in real-time? The authors analyzed the impact of derivative order on the algorithm outcome. They compared the achieved results with some of the best in this field: MEMC-Net [13], SepConv [14], Softmax Splatting [15], DA [9], RIFE [11] and FastRIFE [16].

The rest of this paper is organized as follows. In the next section, related works on flow-based VFI methods are introduced. The proposed algorithm and the mathematical formulations of fractional-order derivatives are described in section 3. Experimental results and conclusions are presented in the last two sections.

2. Related Work

According to a recent survey on video frame interpolation published in [17], the approaches can be divided into five general categories: flow-based, CNN-based, phase-based, GAN-based, and hybrid. The CNN-based methods are gaining interest and show promising results. Still, the hybrid approaches that merge deep network architectures with analytical solutions seem to bring together the best possible solutions.

Since 2012, when AlexNet [18] won the ImageNet Large Scale Visual Recognition Challenge and left all the other competitors far behind, CNNs have been attracting attention and gaining popularity among data scientists. Deep

learning solutions often outperform conventional methods in robustness, generalization and learning ability, especially in tedious and ill-defined computer vision tasks. One of the many successful applications of CNNs is optical flow estimation. The first breakthrough method was proposed by Long *et al.* in [19]. They developed a deep CNN, which can be trained without any supervision and is able to estimate the optical flow directly from input frames using the temporal coherency present in natural videos. This pioneering method provided admittedly blurry results, but it paved a path for other researchers in this area, showing promising potential for further development and application of deeper network structures.

Niklaus *et al.* improved the above idea in [14] by building and training an end-to-end network using only one pass through the encoder-decoder structure. Most solutions, however, employ additional sub-networks for optical flow estimation. Precise information about pixel transition results in more accurately interpolated frames and boosts benchmark metrics. This approach was used, e.g., in the MEMC-Net method, which combines both motion estimation, motion compensation, and post-processing by four sub-networks [13]. In [9] Bao *et al.* improved the MEMC-Net algorithm by incorporating pixel depth information into their DAIN method (Depth-Aware Interpolation). This additional depth information improved dealing with occlusions and large object motion by sampling closer objects better. As depth estimation is one of the most challenging tasks in computer vision, the whole process takes a noticeable amount of time and resources. To this end, the DAIN authors used a fine-tuned hourglass network [20] learned on the MegaDepth dataset [21]. The DAIN method estimates the bi-directional optical flow using an off-the-shelf solution called PWC-Net [10].

In [15] Niklaus and Liu proposed a forward warping operator, called Softmax Splitting, which works conversely to backward warping. It uses an importance mask and weighs all pixels in the input frame. Similarly as in the DAIN method, a fine-tuned PWC-Net for estimating optical flow was used.

Handling large complex motions and occlusions can be achieved by several different approaches. A coarse-to-fine strategy named Deep Voxel Flow was applied in [22], an advanced flow estimation architecture based on PWC-net was used in [8], occlusion mask calculation and adaptive blending were proposed in Super SloMo [7] and MEMC-Net [13], interpolation kernels for adaptive generation of output pixels from a large neighborhood were implemented in SFC-Conv [14].

Most state-of-the-art algorithms generate intermediate frames by combining bi-directional optical flows and additional neural network architectures to provide extra information about the scene, e.g., for estimating depth. This kind of computing is so time-consuming that it makes these methods less in real-time applications. Huang *et al.* proposed the Real-Time Intermediate Flow Estimation (RIFE) algorithm [11] where they used a simple neural network called IFNet for estimating optical flow and another network for gener-

ating interpolated frames called FusionNet. The IFNet comprises three IFBlocks, each built with six ResNet modules and operating on increasing image resolutions. This method generates high-quality output flow and runs six times faster than the PWCNet and 20 times faster than the LiteFlowNet [23]. After estimating the flow, coarse reconstructions are generated by backward warping, and then the fusion process is performed. This fusion process uses a context extractor and the FusionNet with an architecture similar to U-Net, both consisting of four ResNet blocks. In [16] Kubas and Sarwas proposed a modification to the RIFE algorithm called FastRIFE. This hybrid method uses classic analytical methods like Lucas-Kanade or Gunnar-Farneback for optical flow estimation. The proposed modification significantly improved the speed of the RIFE algorithm with only a minor deterioration of the interpolation quality. Hence, the whole process of inter-frame interpolation in the FastRIFE method can run in real time on many more devices.

In this paper, following the FastRIFE algorithm, the authors propose another hybrid solution that combines analytical methods for optical flow estimation with a neural network model for inter-frame interpolation. The optical flow will be estimated with a variant of the sparse Lucas-Kanade algorithm based on fractional-order derivatives.

3. Proposed solution

Optical flow is the pattern of apparent motion of image objects between two consecutive frames caused by the object or camera movement. It can be described by a two-dimensional vector defining the direction of pixel movement from one frame to another. This displacement is estimated based on several assumptions, e.g., the pixels' intensities of an object do not change between consecutive frames, or neighboring pixels share similar motions. One of the classical methods for optical flow estimation is the Lucas-Kanade algorithm [24]. It takes a 3×3 patch around corner points and assumes all nine pixels will have the same flow. The solution is calculated with the least-squares method. Since the publication of the paper mentioned above, many improved optical flow models [25], [26], [27] have been developed. Also, many other models that calculate the flow field using variation-based methods have been proposed. These models generalize the differential order. Two classes of generalizing the differentiation to the variational optical flow methods can be distinguished. One generalization deals with higher-order differentiations [28], and the other deals with fractional-order differentiation, a generalization of the integer-order differentiations [29].

3.1. Fractional-order differentiation

One of the most common formulas for calculating the fractional-order differ-integral of a function $f(x)$ in the time domain is expressed as:

$${}_a D_x^\alpha f(x) = \frac{1}{\Gamma(n-\alpha)} \frac{d^n}{dx^n} \int_a^x \frac{f^{(n)}(\tau)}{(x-\tau)^{\alpha-n+1}} d\tau, \quad (1)$$

where $\alpha \in \mathbb{R}$ (but in general, it can be an imaginary or even a complete complex number) is a fractional order of the differ-integral of the function $f(x)$ and for $n \in \mathbb{N} \cup \{0\}$ the following holds:

$$\begin{aligned} n-1 < \alpha \leq n & \text{ for } \alpha > 0, \\ n = 0 & \text{ for } \alpha \leq 0. \end{aligned} \quad (24)$$

As one can see, when $\alpha > 0$, the result of this function is equivalent to a fractional-order derivative for $\alpha < 0$ to a fractional-order integral and for $\alpha = 0$ to the function itself. The main property of the operator ${}_a D_x^\alpha$ is its linearity for the integer-order differentiation as well as for the fractional differentiation.

There are many different definitions of the fractional derivative. Although the Riemann-Liouville formula in the time domain and many other definitions are sometimes used in practice, the frequency domain definition is applied in this paper because it is easy to implement. The Fourier transform of the n -th derivative of $f(x)$ is:

$$\mathcal{F}(f^{(n)}(x)) = (j\omega)^n F(\omega). \quad (23)$$

The Fourier transform of the Riemann-Liouville fractional derivatives with the lower bound $a = -\infty$ is described in [30] as:

$$\begin{aligned} -\infty D_x^\alpha f(x) &= \frac{1}{\Gamma(n-\alpha)} \int_{-\infty}^x \frac{f^{(n)}(\tau) d\tau}{(x-\tau)^{\alpha+1-n}} d\tau \\ &= -\infty D^{\alpha-n} f^{(n)}(x) \end{aligned} \quad (39)$$

and an assumption is made here that $n-1 < \alpha < n$. After applying the Fourier transform to the equation (3) one arrives at:

$$\begin{aligned} \mathcal{F}(D^\alpha f(x)) &= (j\omega)^{\alpha-n} \mathcal{F}(f^{(n)}(t)) \\ &= (j\omega)^{\alpha-n} (j\omega)^n F(\omega) \\ &= (j\omega)^\alpha F(\omega). \end{aligned} \quad (3)$$

Therefore the pair of fractional derivative in the time domain and the frequency domain is:

$$D^\alpha f(x) \leftrightarrow (j\omega)^\alpha F(\omega), \quad \alpha \in \mathbb{R}^+. \quad (5)$$

For any two-dimensional function $g(x, y)$ absolutely integrable in $(-\infty, \infty) \times (-\infty, \infty)$ the corresponding 2-D Fourier transform is as follows [30]:

$$G(\omega_1, \omega_2) = \int g(x, y) e^{-j(\omega_1 x + \omega_2 y)} dx dy. \quad (48)$$

Therefore one can write the formula for the fractional-order derivatives as:

$$\begin{aligned} D_x^\alpha g &= \mathcal{F}^{-1}((j\omega_1)^\alpha G(\omega_1, \omega_2)), \\ D_y^\alpha g &= \mathcal{F}^{-1}((j\omega_2)^\alpha G(\omega_1, \omega_2)) \end{aligned} \quad (3)$$

where $\mathcal{F}^{-1}$ denotes the inverse of the 2-D continuous Fourier transform. This formula was extensively used in the proposed solution.

36

3.2. Fractional order optical flow

To compute the fractional-order optical flow, many authors, like Kumar *et al.* in [31] for example, use a fractional-order variational optical flow model proposed by Horn and Schunck in [32]. A generalization of this model for non-integer derivative is defined as:

$$\min_w \int_{\Omega} I_x u + I_y v + I_t)^2 + \lambda(|D^\alpha u|^2 + |D^\alpha v|^2) d\Omega, \quad (8)$$

where $D^\alpha := (D_x^\alpha, D_y^\alpha)^T$ denotes the Riemann-Liouville fractional derivative operator and $|D^\alpha u| = \sqrt{(D_x^\alpha u)^2 + (D_y^\alpha u)^2}$.

In this work, the authors used a generalization of the Lucas-Kanade optical flow method based on two main assumptions: color (or brightness in the case of intensity images) constancy, which allows comparing images pixel to pixel, and small motion [12]. The model proposed in this paper has the following form:

$$I_x^\alpha u + I_y^\alpha v + I_t = 0, \quad (34)$$

where I_x^α, I_y^α denote the Riemann-Liouville fractional derivative operator with order $\alpha \in [1, 2]$ in rows and cols, respectively, and I_t is the pixel's motion vector. For $\alpha = 1$, the proposed algorithm is reduced to the traditional Lucas-Kanade method.

3.3. Fractional-order RIFE

The proposed solution for video frame interpolation is based on the Real-time Intermediate Flow (RIFE) algorithm [11]. The authors replaced the IFNet part with analytical methods based on a fractional-order derivative. Then, new FusionNet and ContextNet models were fine-tuned on the proposed solution. The complete architecture is shown in Figure 1.

The authors decided to test the fractional-order Lucas-Kanade optical flow method for three values of $\alpha = 1.2, 1.5, 1.8$. As the L-K algorithm estimates sparse results, the authors used the Shi-Tomasi algorithm for feature point detection, which calculates corner quality measures in each pixel region using minimum eigenvalues. The parameters that were used in the fractional-order Lucas-Kanade algorithm are:

- search window size: 15×15 ,
- number of pyramids: 1 (single level),
- termination criteria: epsilon and criteria count,

and for the Shi-Tomasi detector:

- maximum number of corners: 100,
- minimum accepted quality of corners: 0.1,
- minimum possible distance between the corners: 10,
- computing block size: 7.

An example of optical flow and frame interpolation results is shown in Figure 1, and the speed/accuracy trade-off is presented in Figure 3.

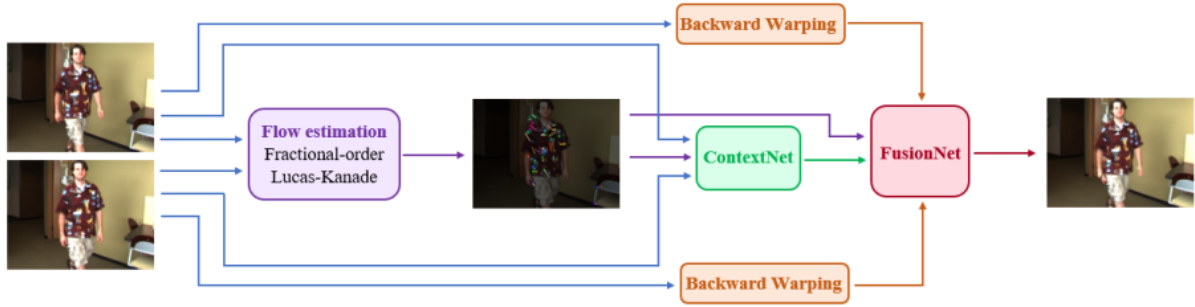


Figure 1. The architecture of the proposed video frame interpolation algorithm. The optical flow is estimated with analytical methods and further processed with fine-tuned neural network models from RIFE: ContextNet and FusionNet.

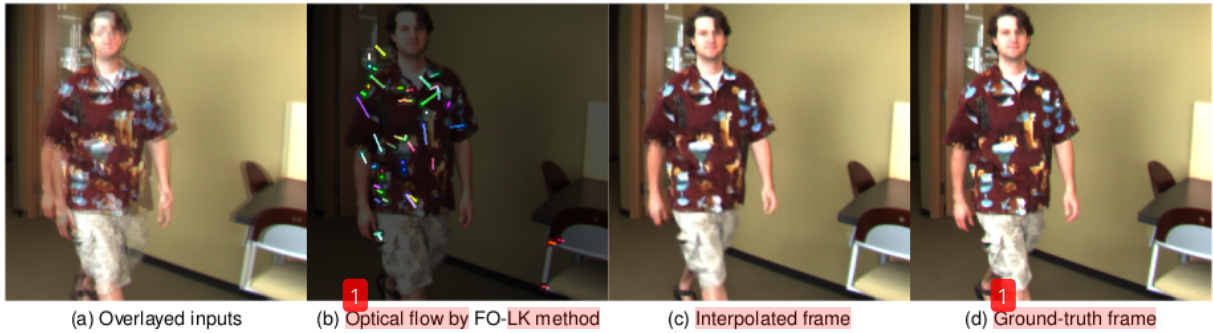


Figure 2. An example of a video frame interpolation results. The authors propose a simplified RIFE model with an optical flow estimated using a modified fractional-order Lucas-Kanade method.

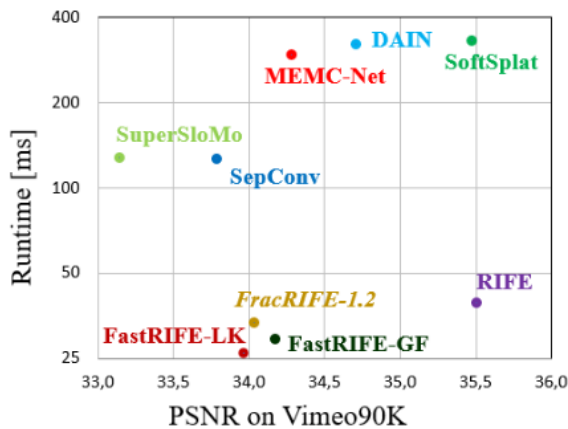


Figure 3. Speed and accuracy trade-off measured on an RTX 2080Ti GPU. The proposed model is compared on Vimeo90K test dataset with other frame interpolation methods: RIFE [11], FastRIFE-LK [16], FastRIFE-GF [16], Softmax Splatting [15], DAIN [9], MEMC-Net [13], SuperSloMo [7] and SepConv [14]. Please note the logarithmic runtime scale.

3.4. Runtime

Runtime comparison can be found in Table 1. The IFNet works much faster than any other state-of-the-art method,

but analytical methods can still calculate the optical flow in less time. Hence, the fractional-order Lucas-Kanade method is faster than IFNet but slower than the traditional Lucas-Kanade algorithm and Gunnar-Farneback method used in FastRIFE-GF. The results were measured again on an RTX 2080 Ti GPU.

TABLE 1. RUNTIME COMPARISON OF OPTICAL FLOW ESTIMATION ALGORITHMS

Method	Runtime [ms]
PWCNet	125
LiteFlowNet	370
IFNet	34
GF	9
LK	7
Ours - Frac-Order LK	13

4. Experiments

This section provides evaluation results and compares the 1st and the best methods with the proposed solution. The comparison is made based on quantitative and visual results using standard metrics and datasets as well as algorithms' runtime.

4.1. Learning strategy

The proposed solution was trained on the Vimeo90K dataset with optical flow labels generated by the LiteFlowNet [23] in PyTorch version [33]. The authors used the AdamW optimizer, the same as used in RIFE [11].

4.1.1. Loss function. To address the problem of training the IFNet module, Huang *et al.* proposed a solution of leakage distillation schema, which compares the network result with the output of the pre-trained optical flow estimation method. The training loss is a linear combination of reconstruction loss L_{rec} , census loss L_{cen} , and leakage distillation loss L_{dis} . The proposed solution was trained using the same loss function:

$$L = L_{rec} + L_{cen} + \lambda L_{dis}, \quad (10)$$

with the weight of leakage distillation loss 10 times smaller than the two others ($\lambda = 0.1$).

4.1.2. Training dataset. Vimeo90K is a video dataset containing 73,171 frame triplets with 448×256 pixels image resolution [10]. This dataset was used as a training set in all compared methods.

4.1.3. Computational efficiency. All analyzed algorithms are implemented in PyTorch [34] using CUDA, cuDNN, and CuPy [35]. The proposed solutions were implemented using OpenCV functions, and the RIFE model was fine-tuned on an RTX 2080 Ti GPU with 11GB of RAM for 150 epochs which took 15 days to converge.

4.2. Evaluation datasets

The following publicly available datasets were used for evaluation: Middlebury, UCF101, and Vimeo90K.

Middlebury contains two subsets: Other with ground truth interpolation results and Evaluation, which output frames are not publicly available [36]. It is the most widely used dataset for testing image interpolation algorithms. The resolution of frames is 640×480 pixels.

UCF101 provides 379 frame triplets from action videos in 101 categories [22], [37]. Image resolution is 256×256 pixels.

Vimeo90K was used for training, but it also contains a test subset: 3,782 triplets with 448×256 pixels image resolution [10].

4.3. Metrics

The following metrics were used to evaluate each algorithm: PSNR, SSIM, and IE:

PSNR (Peak Signal-to-Noise Ratio):

$$\text{PSNR}(i,j) = 10 \cdot \log_{10} \frac{255^2}{\text{MSE}(i,j)}. \quad (11)$$

SSIM (Structural Similarity, l - luminance, c - contrast and s - structure):

$$\text{SSIM}(i,j) = l(i,j) \cdot c(i,j) \cdot s(i,j). \quad (12)$$

IE (Interpolation Error) is the arithmetic average of a difference between the interpolated image and the ground truth frame:

$$\text{IE}(i,j) = |i - j|, \quad (13)$$

where i is the interpolated frame and j is the ground truth image [38]. On the UCF101 and Vimeo90K datasets, the PSNR and SSIM metrics were used (higher values denote better results). The IE metric was used on the Middlebury Evaluation and Other datasets (lower values denote better results). This set of benchmarks is used in most frame interpolation papers and datasets.

4.4. Quantitative comparison

Table 2 shows a comparison of the results obtained on datasets: Middlebury Other, UCF101 and Vimeo90K, red marks the best performance, blue marks second best score. The evaluation results on Middlebury benchmark are not available yet.

The proposed model performs comparably to other methods, scoring very similar results in both PSNR and SSIM benchmarks. The most significant gap is between the results of Interpolation Error on the Middlebury Other dataset. FracRIFE generates a little higher values on the Vimeo90K dataset than SepConv. Also, for $\alpha = 1.2$ FracRIFE achieved better results than FastRIFE - LK. Analytical methods like FastRIFE and FracRIFE achieved the worst result on the Middlebury dataset. The results in the UCF101 and Vimeo90K benchmarks are comparable with other algorithms. The FracRIFE algorithm, like FastRIFE, significantly reduces the number of model parameters.

Evaluation of fractional optical flow methods for different orders shows that the best results were achieved for $\alpha = 1.2$ and are better than inter-frame interpolation based on the traditional Lucas-Kanade method ($\alpha = 1$).

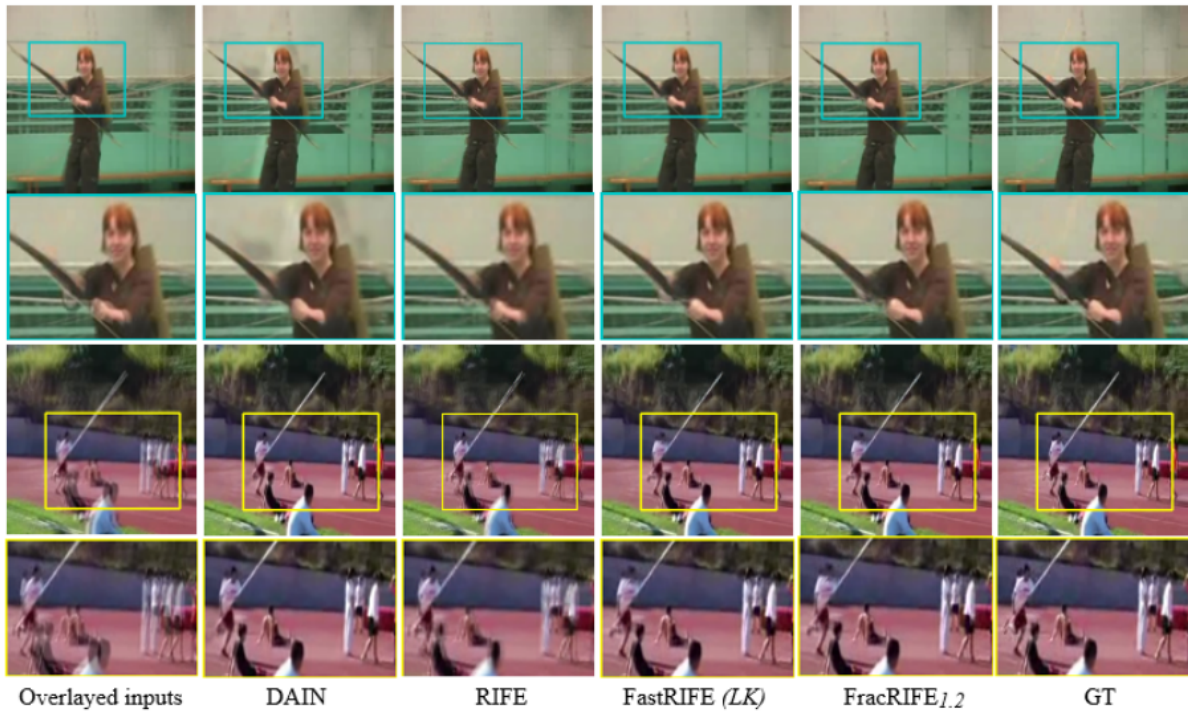
4.5. Visual Comparison

Figure 4 shows examples of frame estimation with comparisons made with the DAIN, RIFE, and FastRIFE(LK) methods. These images show dynamic scenes, which are more challenging to analyze for frame interpolation methods. The resolution of images is small (256×256), but the FracRIFE method for $\alpha = 1.2$ generated acceptable results without shadows found in the DAIN result. The worst smudging is visible in the RIFE results, which works better with high-resolution videos.

The results of high-resolution image interpolation are shown in Figure 5. FracRIFE generates a very noisy frame. Surprisingly, the image looks worse than that obtained with the traditional Lucas-Kanade interpolation method. It is probably caused by including low-resolution videos in the training process and setting constant parameters when calculating the optical flow. Most likely, higher-resolution images should be used in the training process, which will be addressed in future work.

1
TABLE 2. QUANTITATIVE COMPARISON ON THE MIDDLEBURY OTHER, UCF101 AND VIMEO90K TEST DATASETS

Method	Middlebury IE	UCF101		Vimeo90K		Parameters (million)
		PSNR	SSIM	PSNR	SSIM	
SepConv	2.27	34.78	0.967	33.79	0.970	21.6
MEMC-Net	2.12	35.01	0.968	34.29	0.970	70.3
DAIN	2.04	34.99	0.968	34.71	0.976	24.0
SoftSplat	1.81	35.10	0.948	35.48	0.964	7.7
RIFE	1.96	35.25	0.969	35.51	0.978	9.8
FastRIFE - GF	2.89	34.84	0.968	34.18	0.968	4.1
FastRIFE - LK	2.91	34.26	0.967	33.97	0.967	4.1
Ours: FracRIFE _{1.2}	2.92	34.72	0.967	34.01	0.967	4.1
Ours: FracRIFE _{1.5}	2.94	34.57	0.967	33.92	0.965	4.1
Ours: FracRIFE _{1.8}	2.93	34.69	0.967	33.93	0.966	4.1



1
Figure 4. Qualitative comparison on UCF101 dataset.

2
One important issue related to using the sparse optical flow method (Lucas-Kanade), highlighted by the authors of the FastRIFE algorithm, is overlapping feature points as visible spots. These spots are noticeable after zooming in on small-resolution images. The authors did not find this kind of behavior in the FracRIFE models.

1.6. Runtime comparison

Applying analytical methods for optical flow calculation and simplifying the RIFE model architecture made the proposed FracRIFE algorithm faster than any other state-

of-the-art algorithm based on neural network optical flow models. FracRIFE is a little bit slower than FastRIFE, but like its sibling, it has only 4 million parameters which means that these methods are the smallest neural network structures among all video frame interpolation methods. The runtime comparison is shown in Table 3.

Compared to other models, except for FastRIFE and RIFE, FracRIFE runs up to 10 times faster. For $\alpha = 1.2$, the proposed model has slightly better benchmark results from SepConv on the Vimeo90K dataset, but the runtime was considerably reduced. The replacement of the optical flow module in RIFE saved 25% of the time needed to interpolate

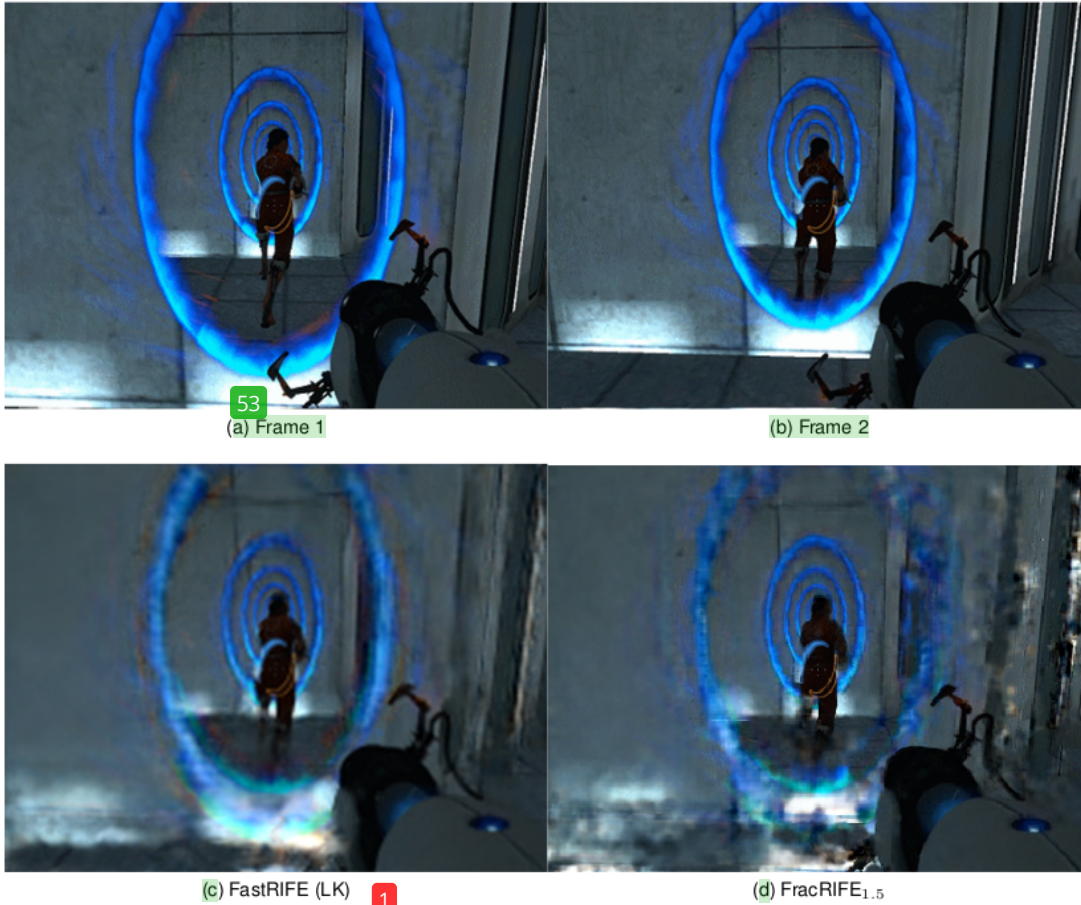


Figure 5. Visual comparison on HD frames.

TABLE 3. RUNTIME COMPARISON OF DESCRIBED METHODS TESTED ON 448×256 IMAGE (GPU: RTX 2080Ti)

Method	Inference time on 448×256 image [ms]	Time of learning one epoch [h]
SepConv	125	-
MEMC-Net	293	-
SoftSplat	329	-
DAIN	320	9
RIFE	39	4.5
FastRIFE - GF	29	2.4
FastRIFE - LK	26	2.4
Ours: FracRIFE _{α}	33	2.9

frames. As shown in Table 3, also the time of learning one epoch was significantly lessened.

5. Conclusion

In this paper, the authors proposed a hybrid solution for video frame interpolation. This novel approach combines

the fractional-order Lucas-Kanade optical flow method and two neural network architectures from the RIFE approach (FusionNet and ContextNet). This combination decreases the complexity of the RIFE algorithm in terms of optical flow estimation. FracRIFE uses a fractional-order optical flow method instead of additional neural network modules, which results in excellent runtime improvement with an

acceptable slight drop in the quality of output frames. The FracRIFE model for $\alpha = 1.2$ achieved better results than FastRIFE based on the standard LK method.

Similarly to FastRIFE, the novel approach proposed in this paper can reduce memory usage. Thanks to the small number of parameters, FracRIFE is able to generate output frames with the allocation of a max of 1.5 GB of RAM in HD videos. Low memory consumption with a short execution time makes the model appropriate for implementation on many devices, including mobile phones.

The FracRIFE algorithm shows that the application of fractional-order derivatives instead of first-order derivatives can improve the results of inter-frame interpolation. Because the experiments conducted up to date cannot indicate the optimal value of the fractional order, it is necessary to perform a deeper analysis on how to choose the best value of this parameter in future works. It is also important to examine another fractional-order optical flow method, namely the dense GF implementation. Future works should also focus on high-resolution videos as the presented results are clearly worse than those provided by RIFE in this domain. It can probably be achieved by including HD images in the training dataset or adjusting optical flow algorithm parameters on the fly based on video resolution or other criteria. The proposed method can stimulate future research on increasing the efficiency of video frame interpolation algorithms.

References

- [1] C.-Y. Wu, N. Singhal, and P. Krahenbuhl, "Video compression through image interpolation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 9 2018.
- [2] H. Men, V. Hosu, H. Lin, A. Bruhn, and D. Saupe, "Visual quality assessment for interpolated slow-motion videos based on a novel database," in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, 2020, pp. 1–6.
- [3] M. Usman, X. He, K.-M. Lam, M. Xu, S. M. M. Bokhari, and J. Chen, "Frame interpolation for cloud-based mobile video streaming," *IEEE Transactions on Multimedia*, vol. 18, no. 5, pp. 831–839, 2016.
- [4] J. Flynn, I. Neulander, J. Philbin, and N. Snavely, "Deep stereo: Learning to predict new views from the world's imagery," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 5515–5524.
- [5] T. Xia, Y. Ding, B. Yu, and X. Huang, "Patch-based frame interpolation for old films via the guidance of motion paths," in *Ninth International Conference on Graphic and Image Processing (ICGIP 2017)*, H. Yu and J. Dong, Eds., vol. 10615, International Society for Optics and Photonics. SPIE, 2018, pp. 1264 – 1271.
- [6] X. Xiang, Y. Tian, Y. Zhang, Y. Fu, J. P. Allebach, and C. Xu, "Zooming slowmo: An efficient one-stage framework for space-time video super-resolution," 2021.
- [7] H. Jiang, D. Sun, V. Jampani, M. Yang, E. Learned-Miller, and J. Kautz, "Super slo-mo: High quality estimation of multiple intermediate frames for video interpolation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9000–9008.
- [8] S. Niklaus and F. Liu, "Context-aware synthesis for video frame interpolation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1701–1710.
- [9] W. Bao, W. Lai, C. Ma, X. Zhang, Z. Gao, and M. Yang, "Depth-aware video frame interpolation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 3698–3707.
- [10] T. Xue, B. Chen, J. Wu, D. Wei, and W. T. Freeman, "Video enhancement with task-oriented flow," *International Journal of Computer Vision (IJCV)*, vol. 127, no. 8, pp. 1106–1125, 2019.
- [11] Z. Huang, T. Zhang, W. Heng, B. Shi, and S. Zhou, "Real-time intermediate flow estimation for video frame interpolation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2019, pp. 29–41.
- [12] R. G. González-Acuña, A. Dávila, and J. C. Gutiérrez-Vega, "Optical flow of non-integer order in particle image velocimetry techniques," *Signal Processing*, vol. 155, pp. 317–322, 2019.
- [13] W. Bao, W.-S. Lai, X. Zhang, Z. Gao, and M.-H. Yang, "Memc-net: Motion estimation and motion compensation driven neural network for video interpolation and enhancement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 3, pp. 933–944, 2021.
- [14] S. Niklaus, L. Mai, and F. Liu, "Video frame interpolation via adaptive separable convolution," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 261–270.
- [15] S. Niklaus and F. Liu, "Softmax splatting for video frame interpolation," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 5436–5445.
- [16] M. Kubas and G. Sarwas, "Fastrife: Optimization of real-time intermediate flow estimation for video frame interpolation," in *29th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision 2021 - WSCG 2021*, vol. 29, Pilsen, Czech Republic: Journal of WSCG, 2021.
- [17] A. Parihar, D. Varshney, K. Pandya, and A. Aggarwal, "A comprehensive survey on video frame interpolation techniques," *The Visual Computer*, pp. 1–25, 01 2021.
- [18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, ser. NIPS'12*. Red Hook, NY, USA: Curran Associates Inc., 2012, p. 1097–1105.
- [19] G. Long, L. Kneip, J. M. Alvarez, H. Li, X. Zhang, and Q. Yu, "Learning image matching by simply watching video," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., Cham: Springer International Publishing, 2016, pp. 434–450.
- [20] W. Chen, Z. Fu, D. Yang, and J. Deng, "Single-image depth perception in the wild," in *Advances in Neural Information Processing Systems 29*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and J. Garnett, Eds., Curran Associates, Inc., 2016, pp. 730–738.
- [21] Z. Li and N. Snavely, "Megadepth: Learning single-view depth prediction from internet photos," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2041–2050.
- [22] Z. Liu, R. A. Yeh, X. Tang, Y. Liu, and A. Agarwala, "Video frame synthesis using deep voxel flow," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 4473–4481.
- [23] T.-W. Hui, X. Tang, and C. C. Loy, "LiteFlowNet: A lightweight convolutional neural network for optical flow estimation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [24] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *In IJCAI81*, 1981, pp. 674–679.
- [25] T. Brox, A. Bruhn, N. Papenberger, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *Computer Vision – ECCV 2004*, T. Pajdla and J. Matas, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 25–36.
- [26] A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/kanade meets horn/schunck: Combining local and global optic flow methods," *International Journal of Computer Vision*, vol. 61, no. 3, pp. 211–231, 2005.
- [27] G. Farneback, "Two-frame motion estimation based on polynomial expansion," in *Image Analysis*, J. Bigun and T. Gustavsson, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 363–370.

- 27
- [28] W. Trobin, T. Pock, D. Cremers, and H. Bischof, "An unbiased second-order prior for high-acc motion estimation," in *Pattern Recognition*, G. Rigoll, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 396–405.
- [29] D. Chen, H. Sheng, Y. Chen, and D. Xue, "Fractional-order variational optical flow model for motion estimation," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1990, pp. 1–10, 2013.
- 44
- [30] I. Podlubny, *Fractional differential equations*. Academic press, 1999.
- [31] P. Kumar, S. Kumar, and B. Raman, "A fractional order variational model for the robust estimation of optical flow from image sequences," *Opt*, vol. 127, no. 20, pp. 8710–8727, 2016.
- 35
- [32] B. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 1, pp. 185–203, 1981.
- 5
- [33] S. Niklaus, "A reimplementation of LiteFlowNet using PyTorch," <https://github.com/sniklaus/pytorch-liteflownet>, 2019.
- 18
- [34] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," in *NIPS 2017 Autodiff Workshop: The Future of Gradient-based Machine Learning Software and Techniques*, 2017.
- 33
- [35] S. Chetlur, C. Woolley, P. Vandermersch, J. Cohen, J. Tran, B. Catanzaro, and E. Shelhamer, "Cudnn: Efficient primitives for deep learning," *International Journal of Computer Vision*, 2014.
- 11
- [36] S. Baker, S. Roth, D. Scharstein, M. J. Black, J. P. Lewis, and R. Szeliski, "A database and evaluation methodology for optical flow," in *2007 IEEE 11th International Conference on Computer Vision*, 2007, pp. 1–8.
- [37] K. Soomro, A. R. Zamir, and M. Shah, "Ucf101: A dataset of 101 human actions classes from videos in the wild," in *International Journal of Computer Vision*, 2012.
- 8
- [38] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

60%

SIMILARITY INDEX

PRIMARY SOURCES

1	hdl.handle.net Internet	1046 words — 19%
2	www.arxiv-vanity.com Internet	506 words — 9%
3	Grzegorz Sarwas, Slawomir Skoneczny. "FSIFT based feature points for face hierarchical clustering", 2018 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), 2018 Crossref	164 words — 3%
4	Duolikun Danier, David Bull. "Texture-aware Video Frame Interpolation", 2021 Picture Coding Symposium (PCS), 2021 Crossref	130 words — 2%
5	arxiv.org Internet	128 words — 2%
6	web.archive.org Internet	79 words — 1%
7	Muzammil Khan, Pushpendra Kumar. "A Vision Based Fractional Order TV-Model for Underwater Motion Estimation", 2021 IEEE Bombay Section Signature Conference (IBSSC), 2021 Crossref	64 words — 1%

8	export.arxiv.org Internet	63 words — 1%
9	Dali Chen, Hu Sheng, YangQuan Chen, Dingyü Xue. "Fractional-order variational optical flow model for motion estimation", Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2013 Crossref	55 words — 1%
10	Shaohua Liu, Minghao Li, Xiaona Zhang, Shuang Liu, Zhaoxin Li, Jing Liu, Tianlu Mao. "Image-based Rendering for Large-scale Outdoor Scenes with Fusion of Monocular and Multi-view Stereo Depth", IEEE Access, 2020 Crossref	54 words — 1%
11	www.creatis.insa-lyon.fr Internet	54 words — 1%
12	deepai.org Internet	49 words — 1%
13	Xiaodong Cun, Feng Xu, Chi-Man Pun, Hao Gao. "Depth Assisted Full Resolution Network for Single Image-based View Synthesis", IEEE Computer Graphics and Applications, 2018 Crossref	44 words — 1%
14	Anil Kokaram, Davinder Singh, Simon Robinson. "A Bayesian View of Frame Interpolation and a Comparison with Existing Motion Picture Effects Tools", 2020 IEEE International Conference on Image Processing (ICIP), 2020 Crossref	43 words — 1%
15	dspace.cvut.cz Internet	43 words — 1%

-
- 16 Phat Nguyen Huu, Tien Luong Ngoc, Quang Tran Minh. "Proposing Gesture Recognition Algorithm Using Two-Stream Convolutional Network and LSTM", 2020 IEEE Eighth International Conference on Communications and Electronics (ICCE), 2021
Crossref 42 words — 1%
-
- 17 manuscriptlink-society-file.s3.ap-northeast-1.amazonaws.com
Internet 42 words — 1%
-
- 18 yorkspace.library.yorku.ca
Internet 41 words — 1%
-
- 19 link.springer.com
Internet 39 words — 1%
-
- 20 Xianhang Cheng, Zhenzhong Chen. "A Multi-Scale Position Feature Transform Network for Video Frame Interpolation", IEEE Transactions on Circuits and Systems for Video Technology, 2019
Crossref 38 words — 1%
-
- 21 macsphere.mcmaster.ca
Internet 36 words — 1%
-
- 22 di.ku.dk
Internet 32 words — 1%
-
- 23 "Image Processing and Communications Challenges 10", Springer Science and Business Media LLC, 2019
Crossref 31 words — 1%
-
- 24 advancesindifferenceequations.springeropen.com
Internet 30 words — 1%

25 Lingtong Kong, Jinfeng Liu, Jie Yang. "Progressive Motion Context Refine Network for Efficient Video Frame Interpolation", IEEE Signal Processing Letters, 2022 29 words — 1%
Crossref

26 repository.dl.itc.u-tokyo.ac.jp 28 words — 1%
Internet

27 Grbić, Saša, Martin Urschler, Thomas Pock, Horst Bischof, and David R. Haynor. "<title>Optical flow based deformable volume registration using a novel second-order regularization prior</title>", Medical Imaging 2010 Image Processing, 2010. 25 words — < 1%
Crossref

28 Piotr Skulimowski, Mateusz Owczarek, Pawel Strumillo. "Door detection in images of 3D scenes in an electronic travel aid for the blind", Proceedings of the 10th International Symposium on Image and Signal Processing and Analysis, 2017 25 words — < 1%
Crossref

29 cio.mx 24 words — < 1%
Internet

30 Wenchao Su, Yanyan Wang, Fei Su, Zhicheng Zhao. "Micro-Expression Recognition Based on the Spatio-Temporal Feature", 2018 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 2018 20 words — < 1%
Crossref

31 www.iitr.ac.in 19 words — < 1%
Internet

32 epdf.pub 18 words — < 1%
Internet

33	mit.imt.si Internet	18 words — < 1%
34	www.mdpi.com Internet	18 words — < 1%
35	Keito Suzuki, Masaaki Ikehara. "Residual Learning of Video Frame Interpolation Using Convolutional LSTM", IEEE Access, 2020 Crossref	17 words — < 1%
36	www.hindawi.com Internet	16 words — < 1%
37	www.grafiati.com Internet	15 words — < 1%
38	ras.papercept.net Internet	14 words — < 1%
39	core.ac.uk Internet	13 words — < 1%
40	sail.usc.edu Internet	12 words — < 1%
41	www.tib.eu Internet	12 words — < 1%
42	Zehua Fan, Feng Zhu, Lei Li, Xiaoyang Tan. "Multi-scale Intermediate Flow Estimation for Video Frame Interpolation", 2022 IEEE 34th International Conference on Tools with Artificial Intelligence (ICTAI), 2022 Crossref	11 words — < 1%
43	Bin Zhu, Lian-Fang Tian, Qi-Liang Du, Qiu-Xia Wu, Farisi Zeyad Sahl, Yao Yeboah. "Adaptive dual	10 words — < 1%

fractional-order variational optical flow model for motion estimation", IET Computer Vision, 2019

Crossref

-
- | | | |
|----|--|-----------------|
| 44 | citeseerx.ist.psu.edu
<small>Internet</small> | 10 words — < 1% |
|----|--|-----------------|
-
- | | | |
|----|--|-----------------|
| 45 | scholarcommons.scu.edu
<small>Internet</small> | 10 words — < 1% |
|----|--|-----------------|
-
- | | | |
|----|---|-----------------|
| 46 | www.theses.fr
<small>Internet</small> | 10 words — < 1% |
|----|---|-----------------|
-
- | | | |
|----|--|----------------|
| 47 | "Pattern Recognition and Computer Vision",
Springer Science and Business Media LLC, 2018
<small>Crossref</small> | 9 words — < 1% |
|----|--|----------------|
-
- | | | |
|----|--|----------------|
| 48 | Yan-Shan Zhang, Feng Zhang, Bing-Zhao Li, Ran Tao. "Fractional domain varying-order differential denoising method", Optical Engineering, 2014
<small>Crossref</small> | 9 words — < 1% |
|----|--|----------------|
-
- | | | |
|----|---|----------------|
| 49 | congresodeinvestigacion.tec.mx
<small>Internet</small> | 9 words — < 1% |
|----|---|----------------|
-
- | | | |
|----|--|----------------|
| 50 | Alexiadis, A.. "Extension of the method of moments for population balances involving fractional moments and application to a typical agglomeration problem", Journal of Colloid And Interface Science, 20040801
<small>Crossref</small> | 8 words — < 1% |
|----|--|----------------|
-
- | | | |
|----|---|----------------|
| 51 | Lecture Notes in Computer Science, 2010.
<small>Crossref</small> | 8 words — < 1% |
|----|---|----------------|
-
- | | | |
|----|--|----------------|
| 52 | Y. Matsushita, E. Ofek, Weina Ge, Xiaoou Tang, Heung-Yeung Shum. "Full-frame video stabilization with motion inpainting", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006 | 8 words — < 1% |
|----|--|----------------|

-
- 53 researchchain.net
Internet 8 words — < 1%
-
- 54 smartech.gatech.edu
Internet 8 words — < 1%
-
- 55 vdoc.pub
Internet 8 words — < 1%
-
- 56 Anil Singh Parihar, Disha Varshney, Kshitija Pandya, Ashray Aggarwal. "A comprehensive survey on video frame interpolation techniques", The Visual Computer, 2021
Crossref 7 words — < 1%
-
- 57 Fenghui Yao. "Real-Time Multiple Moving Targets Detection from Airborne IR Imagery by Dynamic Gabor Filter and Dynamic Gaussian Detector", EURASIP Journal on Image and Video Processing, 2010
Crossref 7 words — < 1%
-
- 58 "Computer Vision – ECCV 2018", Springer Science and Business Media LLC, 2018
Crossref 6 words — < 1%
-
- 59 Lecture Notes in Computer Science, 2015.
Crossref 6 words — < 1%
-
- 60 Richard Szeliski. "A Database and Evaluation Methodology for Optical Flow", 2007 IEEE 11th International Conference on Computer Vision, 10/2007
Crossref 6 words — < 1%
-
- 61 Xinhua Liu, Tianyu You, Xiaolin Ma, Hailan Kuang. "An Optimization Model for Human Activity Recognition Inspired by Information on Human-Object

62

Zhewei Huang, Tianyuan Zhang, Wen Heng, Boxin Shi, Shuchang Zhou. "Chapter 36 Real-Time Intermediate Flow Estimation forVideo Frame Interpolation", Springer Science and Business Media LLC, 2022

6 words — < 1%