# Hardware-Friendly Laplacian-Based Multi-Focus Image Fusion in DCT Domain for Visual Sensor Network

**SOBHAN KANTI DHARA** [1], (Graduate Student Member, IEEE),
**DEBASHIS SEN** [1], (Senior Member, IEEE), AND
**M. N. S. SWAMY** [2], (Life Fellow, IEEE)

[1]Department of Electronics and Electrical Communication Engineering, IIT Kharagpur, Kharagpur 721302, India
[2]Department of Electrical and Computer Engineering, Concordia University, Montreal, QC H3G 1M8, Canada

Corresponding author: M. N. S. Swamy (swamy@encs.concordia.ca)

**ABSTRACT** Visual sensor network (VSN) requires a multi-focus image or video frame fusion technique involving focus measure computation in the DCT-domain to generate an all-in-focus image. Such techniques are implemented on resource-constrained on-board systems requiring hardware-friendly implementations. In this article, we first show that components of the Laplacian matrix are related to the discrete cosine transform (DCT) basis. The relation is that the eigenvalues of the Laplacian with proper boundary condition form the diagonal elements of the diagonal matrix generated by the DCT operation on the Laplacian. Exploiting this relation, we propose a focus measure which works on the DCT coefficients reflecting the spatial-domain Laplacian operation. Certain simplifications allow our focus measure computation through hardware-friendly integer multiplication and summation, where matrix multiplication involves just N scalar multiplications for an $N \times N$ 2D signal. Finally, we propose an approach which suitably fuses multi-focus images or video frames in DCT based image or video coding framework through detection of properly focused area and neighborhood consistency analysis. We show that our proposed approach is hardware-friendly, computationally simple, and is fast enough for VSN. Through experimental results, we show that our approach outperforms the relevant state-of-the-art in multi-focus image fusion for VSN both quantitatively and subjectively. We also show that our approach is effective in comparison to the state-of-the-art and a few latest generic multi-focus image fusion techniques in terms of quantitative and subjective evaluations.

**INDEX TERMS** Anti-diagonal matrix, discrete cosine transform, multi-focus image fusion, visual sensor network.

## I. INTRODUCTION

Visual sensor network (VSN) is an intelligent system [1], [2] which generates huge sensory data from the environment through geographically distributed camera nodes and then collaboratively processes [3] to send useful information for different applications related to surveillance, monitoring, etc. [4]–[6]. The cameras have limited depth of field, and hence, they capture reliable (focused) information of objects only at their focus ranges and blur the rest of the scene [7]. The on-board system for the required collaborative processing to produce all-in-focus images/video frames, is resource-constrained [8], [9], allowing very simple

The associate editor coordinating the review of this manuscript and approving it for publication was Sudhakar Radhakrishnan.

hardware-friendly processing of the huge captured data [10]. Therefore, generating an all-in-focus image/video frame in real-time, especially in a wireless sensor network [11], [12] is a challenging task.

To handle and transmit huge visual data, VSN adopts an image/video compression or video coding technique. Many use discrete cosine transform (DCT) in view of its efficient energy compaction property [13]–[15]. Thus, a simple multi-focus image fusion technique that works and generates an all-in-focus image/frame in DCT domain would be preferable for VSN, as it can be directly incorporated in the DCT based compression framework. Many such approaches have been proposed [10], [16]–[23], which are discussed in Section III. Among these, a few attempts have been made to formulate DCT domain equivalent of efficient spatial domain

focus measures that utilize spatial variance [18] or spatial frequency [19] in a local spatial neighbourhood. Other attempts have also been made in DCT domain that represent operations in spatial neighborhods [20]–[23], which are discussed in Section III as well. For example, Amin-Naji *et al.* [20] proposed the use of variance and energy of Laplacian response in DCT domain for multi-focus image fusion in VSN. Though these approaches produce state-of-the-art results, they are comparatively expensive as detailed in Section VI-C. But, such a solution requires floating point squaring and multiplication operations, which are not hardware-friendly like addition.

It is straightforward to assume that when an area of an image is not focused, it is blurred. Such an area contains substantially more low frequency content than high frequency one. On the other hand, when an area (except smooth areas) is focused, it is comparatively sharper, contributing to high frequency content. Therefore, different operators which compute spatial frequencies can provide focus measures. Gradient is one such operator to measure image sharpness indicating spatial frequency. It's equivalent in DCT domain has been used in [19] to get spatial frequency based focus measure. Laplacian operator is another well known operator [24], which is good at determining the sharpness in an image. A plethora of focus measures in spatial domain based on Laplacian operator have been proposed and are found to be effective [7], [25]–[28]. Subbarao *et al.* [26], [27] have proposed the energy of the Laplacian of an image to measure the focus for an auto focusing application. Nayar *et al.* [25] have proposed the sum of modified Laplacian (SML) of an image as an efficient focus measure. Malik *et al.* [7] have proposed an auto focus algorithm, where SML of an image is used to determine the focus information.

Inspired by the aforesaid requirement of hardware friendliness, in this article, we formulate a computationally simple hardware-friendly multi-focus image fusion approach for VSN using a non-trivial equivalent of the spatial domain Laplacian operator in the DCT domain.

The major contributions of the paper are as follows:

1) It is shown that an $N \times N$ Laplacian matrix (formed using Laplacian operator with proper boundary condition) is related to the $N \times N$ DCT matrix in such a way that the eigenvalues of the Laplacian matrix form the diagonal elements of the diagonal matrix generated by DCT operation on the Laplacian matrix.

2) The above property of the Laplacian matrix is exploited to propose a focus measure, which can be directly computed on the DCT coefficients such that it requires $N$ fixed point multiplications. In comparison, a generic matrix multiplication requires $N^3$ floating point multiplications.

3) An approach, which fuses the multi-focus images or video frames in DCT based image or video coding framework using our Laplacian based focus measure, is proposed. The proposed approach is not only computationally less expensive and hardware-friendly,

but also produces state-of-the-art results of multi-focus image fusion for VSN.

The paper is organized as follows. Section II shows how the properties of the Laplacian with proper boundary condition can be exploited to provide a computationally simple solution in the DCT domain, which can be used in multi-focus fusion for VSN. Section III describes the related literature on multi-focus fusion for VSN. Section IV proposes hardware-friendly computationally simple focus measure for VSN. Section V describes the step-by-step procedure of the proposed multi-focus image fusion method. Comparative results of multi-focus image fusion for VSN using different techniques including our proposed one are presented in Section VI. Finally, Section VII concludes the paper.

## II. LAPLACIAN RESPONSE THROUGH EFFICIENT DCT DOMAIN COMPUTATION

In this section, we first discuss the relation of $N \times N$ Laplacian operator with DCT. Later, we mathematically show how such a relation can be useful for developing a hardware-friendly computationally simple solution in the DCT domain to get the Laplacian response.

### A. LAPLACIAN RESPONSE FOR A BLOCK BASED OPERATION

Let us first consider a digital double derivative having the smallest length, $[-1 \ 2 \ -1]$, which is obtained using the digital single derivative $[-1 \ 1]$ twice at consecutive locations. We shall then extend the operator $[-1 \ 2 \ -1]$ to get a 2D spatial Laplacian operator for block based operation. The digital double derivative (1D Laplacian) operation for digital 1D signal $F$ having $N$ samples is as follows

$$\frac{d^2 f_x}{dx^2} = -f_{x-1} + 2f_x - f_{x+1} \qquad (1)$$

As the operation at the signal sample $x$ is basically high pass filtering, its response provides the high frequency content at that signal sample. If we want to extend the Laplacian response defined at 1 to $N$ samples at a time, we can represent it in matrix form as follows

$$R = PF \qquad (2)$$

where $R$ is an $N \times 1$ column vector, $P$ is an $N \times N$ matrix defined

$$P = \begin{bmatrix} \times & \times & 0 & \cdots & \\ -1 & 2 & -1 & 0 & \cdots \\ & & \cdots & & \\ \cdots & 0 & -1 & 2 & -1 \\ & \cdots & 0 & \times & \times \end{bmatrix} \qquad (3)$$

and $F$ is a column vector of size $N \times 1$ given by

$$F = [f(0) \quad . \quad . \quad . \quad f(N-1)]^T \qquad (4)$$

As the 1D Laplacian operator of the smallest length requires one preceding sample and one succeeding sample for computation at a signal sample, the values of $R$ at $r(0)$ and $r(N-1)$

cannot be computed. Hence, we use '$\times$' in the $\boldsymbol{P}$ operator to indicate the same, which essentially means that the values at '$\times$' location would not be employed. As stated in [29], the operator $\boldsymbol{P}$ is related to DCT basis function.

If $\boldsymbol{F}$ is a 2D signal of size $N \times N$, for a location $(x, y)$, we have

$$\frac{\partial^2 f_{x,y}}{\partial x^2} = -f_{x-1,y} + 2f_{x,y} - f_{x+1,y}$$
$$\frac{\partial^2 f_{x,y}}{\partial y^2} = -f_{x,y-1} + 2f_{x,y} - f_{x,y+1} \qquad (5)$$

The combination $\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is known as the Laplacian operator. For a 2D signal we will get two responses, one for row-wise operation ($\boldsymbol{R_r}$) and another for column-wise operation ($\boldsymbol{R_c}$) as follows

$$\boldsymbol{R_r} = \boldsymbol{PF}$$
$$\boldsymbol{R_c} = \boldsymbol{FP^T} \qquad (6)$$

where $\boldsymbol{F}$ (the signal) is an $N \times N$ matrix. As mentioned already, '$\times$' in $\boldsymbol{P}$ remains unemployed. However, depending on the underlying purpose, certain values at '$\times$' can be more suited than others. In our case, we require the following

- The values at '$\times$' must yield high frequency upon operation on $\boldsymbol{F}$, as $\boldsymbol{P}$ is meant to provide high frequency content.
- The values at '$\times$' must allow efficient computation of $\boldsymbol{R_r}$ and $\boldsymbol{R_c}$ in the DCT domain, that is, on the DCT coefficients of $\boldsymbol{F}$ (See Subsection II-B).

To meet the said requirements, we propose the following operator $\boldsymbol{L}$ for an $N \times N$ signal block because of its interesting properties, which will be shown in the discussion that follows.

$$\boldsymbol{L} = \begin{bmatrix} 1 & -1 & 0 & \dots & \\ -1 & 2 & -1 & 0 & \dots \\ & & \dots & & \\ \dots & 0 & -1 & 2 & -1 \\ & \dots & 0 & -1 & 1 \end{bmatrix} \qquad (7)$$

It is straightforward that $\boldsymbol{L}$ is obtained by replacing $[\times \ \times]$ of $\boldsymbol{P}$ in the first row by $[1 \ {-1}]$ and in the last row by $[{-1} \ 1]$, both being digital derivatives. The operator $\boldsymbol{L}$, which contains the digital derivatives and 1D Laplacian, yields high frequency content of the signal in the relevant direction. We see that, $\boldsymbol{L} = \boldsymbol{L^T}$ and $\boldsymbol{L}$ can be defined by singular value decomposition as follows

$$\boldsymbol{L} = \boldsymbol{U} \Lambda \boldsymbol{V^T} \qquad (8)$$

where both $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal matrices and $\lambda_{i,j} = \lambda_{i,i}\delta_{i,j}, i, j = 1, 2, \dots, N, \delta_{i,j}$ being the Kronecker delta and $\lambda_{i,j}$ the singular values of $\boldsymbol{L}$. Here, $\boldsymbol{U}$ and $\boldsymbol{V}$ are such that $\lambda_{i,i} > \lambda_{k,k}, \forall i < k, (i, j) = 1, 2, \dots, N$. In our case, $\boldsymbol{U} = \boldsymbol{V}$ so that $\boldsymbol{UV^T} = \boldsymbol{I}, \boldsymbol{V^T U} = \boldsymbol{I}$. Let us represent $u_{i,j}$ as $i^{th}$ row and $j^{th}$ column element of matrix $\boldsymbol{U}$. Then, we can represent $i^{th}$ row vector $\boldsymbol{U_i^r}$ as $[u_{i,1} \ u_{i,2} \ \dots \ u_{i,n} \ \dots \ u_{i,N}]$ and $i^{th}$ column vector $\boldsymbol{U_i^c}$ as $[u_{1,i} \ u_{2,i} \ \dots \ u_{n,i} \ \dots \ u_{N,i}]^T$. Similarly for $N \times N$ DCT

matrix $\boldsymbol{D}$, we can represent the element at $i^{th}$ row and $j^{th}$ column as $d_{i,j}$, and $i^{th}$ row vector $\boldsymbol{D_i^r}$ as $[d_{i,1} \ d_{i,2} \ \dots \ d_{i,n} \ \dots \ d_{i,N}]$ and $i^{th}$ column vector $\boldsymbol{D_i^c}$ as $[d_{1,i} \ d_{2,i} \ \dots \ d_{n,i} \ \dots \ d_{N,i}]^T$. The matrices $\boldsymbol{U}$ and $\boldsymbol{D}$ are related as follows

$$\boldsymbol{D_i^r U_j^c} = \begin{cases} 1 \ or \ -1, & j = N + 1 - i \\ 0, & elsewhere \end{cases} \qquad (9)$$

The 0 value indicates that the corresponding eigenvectors are orthogonal to DCT basis vectors, which is the case except at the index $(i, N + 1 - i)$. At index $(i, N + 1 - i)$, we get a value 1 when an eigenvector and a DCT basis vector are the same, and a value $-1$ when they have a $180°$ phase difference. Due to the above properties, $\boldsymbol{E}$ is an anti-diagonal matrix where $\boldsymbol{E} = \boldsymbol{DU}$. This implies $\boldsymbol{E^T} = \boldsymbol{U^T D^T}$. As, $\boldsymbol{U} = \boldsymbol{V}$, we can write $\boldsymbol{E^T} = \boldsymbol{V^T D^T}$. Therefore, $\boldsymbol{E}$ is an anti-diagonal matrix which has values $1/-1$ at index $(i, N+1-i)$. Transpose of an anti-diagonal matrix is another anti-diagonal matrix. So, $\boldsymbol{E}$ and $\boldsymbol{E^T}$ are two anti-diagonal matrices. Based on these, we can show that $\boldsymbol{A} = \boldsymbol{EE^T}$ is a diagonal matrix whose diagonal element $a_{i,i} = e_{i,N+1-i} \times e_{N+1-i,i}^T$. The corresponding proof is provided in Appendix A. Applying (9) and the above result for $a_{i,i}$, we can say, $\boldsymbol{EE^T} = \boldsymbol{I}$. Now, applying DCT to both sides of (8), we have

$$DCT(\boldsymbol{L}) = \boldsymbol{DU}\Lambda \boldsymbol{V^T D^T} \qquad (10)$$

As $\boldsymbol{E} = \boldsymbol{DU}$ and $\boldsymbol{E^T} = \boldsymbol{V^T D^T}$, we can rewrite the above expression

$$DCT(\boldsymbol{L}) = \boldsymbol{E}\Lambda \boldsymbol{E^T} \qquad (11)$$

In the above expression, $\Lambda$ is a diagonal matrix. As $\Lambda$ is diagonal matrix and $\boldsymbol{E^T}$ is an anti-diagonal matrix, we can show that $\boldsymbol{B} = \Lambda \boldsymbol{E^T}$ is an anti-diagonal matrix where the anti-diagonal element $b_{i,N+1-i} = \lambda_{i,i} \times e_{i,N+1-i}^T$. The corresponding proof is provided in Appendix B. Now, let us denote $DCT(\boldsymbol{L}) = \boldsymbol{E}\Lambda \boldsymbol{E^T}$ as $\boldsymbol{M}$, where $\boldsymbol{E}$ and $\boldsymbol{E^T}$ are two anti-diagonal matrices and $\Lambda$ is a diagonal matrix. Then, applying $\boldsymbol{EE^T} = \boldsymbol{I}$, we can say that $\boldsymbol{M}$ is a diagonal matrix whose diagonal element $m_{i,i}$ is given by

$$m_{i,i} = \lambda_{N+1-i,N+1-i} \qquad (12)$$

The corresponding proof is provided in Appendix C. Again, $\lambda_{i,j} = \lambda_{i,i}\delta_{i,j}, i, j = 1, 2, \dots, N$, where $\lambda_{i,i} > \lambda_{k,k}, \forall i < k, (i, j) = 1, 2, \dots, N$. So, we can say that the DCT operation on $\boldsymbol{L}$ generates a diagonal matrix $\boldsymbol{M}$ whose diagonal terms are the eigenvalues of $\boldsymbol{L}$ arranged from minimum to maximum starting from the first diagonal element.

### B. EFFICIENT DCT DOMAIN COMPUTATION OF LAPLACIAN FOR A BLOCK BASED OPERATION

Any filtering operation on 2D signal (image block), like the ones used in the computation of our focus measure, requires row-wise and column-wise operations as shown in expression (6). In matrix representation, the row-wise operation can be expressed as $\boldsymbol{C} = \boldsymbol{AB}$, where any element $c_{i,j}$ is

computed as $\sum_{k=1}^{N} a_{ik}b_{kj}$. We require $N$ multiplications and $N-1$ additions for $N^2$ such elements. Now, to compute the additive response $\sum_{i=1}^{N}\sum_{j=1}^{N} c_{i,j}$ (as used later for focus measure computation in Section IV), we require $N^2-1$ additions. So, the computation of $\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N} a_{ik}b_{kj}$ requires $N^3$ multiplications and $N(N-1)+(N^2-1)$ additions. Further, we will have the same amount of computation for the column-wise operation and additive response computation. Multiplication operations, which are not hardware-friendly like additions, are always preferred to be minimal in number. The focus measure in [19], computed by the filtering operation on image block, reduces the computations required by one matrix operation, thus involving $N^3$ multiplications instead of $2N^3$. Our choice of operator $\boldsymbol{L}$ which is defined in (7) is such that DCT($\boldsymbol{L}$) is a diagonal matrix as shown in (12). For an image $\boldsymbol{F}$, let us denote $\boldsymbol{F^D} = DCT(\boldsymbol{F})$. To compute the additive response from $\boldsymbol{T^1} = DCT(\boldsymbol{LF})$ we require $N^2$ multiplications due to the operator $\boldsymbol{L}$. So, there is a reduction in the number of multiplications from $N^3$ to $N^2$. As our DCT domain focus measure is an additive response on matrix multiplication operation, we can further reduce the required computations. The additive response $\sum \boldsymbol{T^1} = \sum DCT(\boldsymbol{LF}) = \sum DCT(\boldsymbol{L})DCT(\boldsymbol{F})$ which is again equal to $\sum \boldsymbol{MF^D}$. As shown in expression (12), $\boldsymbol{M}$ is a diagonal matrix having $m_{i,i} = \lambda_{N+1-i,N+1-i}$. So $i^{th}$ row of $\boldsymbol{T^1}$ is $i^{th}$ row of $\boldsymbol{F^D}$ multiplied with $\lambda_{N+1-i,N+1-i}$. The additive response for $i^{th}$ row of $\boldsymbol{T^1}$ is $\lambda_{N+1-i,N+1-i}\sum_{j=1}^{N} f_{i,j}^D$. Therefore, additive response of $\boldsymbol{T^1}$ is as follows

$$\sum_{i=1}^{N}\sum_{j=1}^{N} t_{i,j}^1 = \sum_{i=1}^{N} \lambda_{N+1-i,N+1-i}\sum_{j=1}^{N} f_{i,j}^D \qquad (13)$$

Similarly, the additive response of $\boldsymbol{T^2} = DCT(\boldsymbol{FL})$ is computed as follows

$$\sum_{i=1}^{N}\sum_{j=1}^{N} t_{i,j}^2 = \sum_{i=1}^{N} \lambda_{N+1-i,N+1-i}\sum_{j=1}^{N} f_{j,i}^D \qquad (14)$$

It is clear that the above solutions reduce the number of multiplications from $2N^3$ to $2N$, making it more hardware-friendly and less computationally expensive.

## III. RELATED WORKS ON MULTI-FOCUS IMAGE FUSION FOR VSN

Resource-constrained VSN system demands a simple multi-focus image fusion solution generating an all-in-focus image/frame from different multi-focus images/frames captured from geographically distributed camera nodes. As mentioned before, a DCT domain hardware-friendly, fast, but computationally less intensive solution, is preferable which can be incorporated into image compression or video coding framework.

Tang [16] has proposed two methods in the DCT domain, DCT+Average (DCTav) and DCT+Contrast (DCTcm) for multi-focus image fusion. DCT+Average is the simpler one, where all the DCT coefficients are averaged to generate the fused all-in-focus image. This technique does not exploit the relationship between focused and defocused regions based on the DCT coefficients. Degree of focus which is not used in DCTav, is a measure of sharpness that reflects the high frequency components of an image block in the DCT domain. Considering the importance of focus, a contrast based on the mean amplitude over a frequency band in an image block is proposed. In the fused image, the DC coefficients are averaged, whereas the AC coefficients corresponding to larger contrast value are chosen. Including the above proposals, a number of novel approaches have been proposed and listed in [17], where the author has shown multiple ways to exploit the DCT coefficients for image fusion. In DCT+AC-Max (DCTma), the fused image is generated by choosing the AC coefficients of larger magnitude among the input images. For the DC coefficient in the fused image, the DC values of the input images are averaged. In DCT+Contrast-Modified (DCTch) and DCT+AC-Max-Modified (DCTah), the DC along with the lowest frequency coefficients are averaged in the fused image. The rest of the coefficients from input images are chosen based on high contrast value for DCTch and based on the magnitude of AC coefficients for DCTah. In DCT+Energy-Max (DCTe), the author has proposed the energy of the frequency band of the DCT coefficients as the measure, and the DCT coefficients corresponding to higher frequency band energy are selected in the fused image. But none of the above proposed techniques include spatial information of the image.

Haghighat *et al.* [18], [30] have shown that the variance in a local spatial neighbourhood is the same as the variance of normalized DCT coefficients. Such a formulation generates the result as in the spatial domain without doing spatial domain computations. Based on this, the authors have proposed two technqiues DCT+Variance (DCTv) and DCT+Variance+CV (DCTvc) [18]. The former approach fuses images based on the variance of normalized DCT coefficients, whereas the approach DCTvc includes a consistency verification in DCTv [18] for fusion. Cao *et al.* [19] formulated the computation of the spatial frequency in the DCT domain and used it for multi-focus image fusion. Based on the spatial frequncy in the DCT domain, they proposed DCT+SF(DCTs) and DCT+SF+CV (DCTsc) [19]. The former approach fuses images based on the spatial frequency in DCT domain, whereas the latter includes a consistency verification [19] in DCTs for fusion. Later Vakaimalar *et al.* [21] suggested Min-Max normalization on the DCT coefficients while performing fusion based on spatial frequency. Amin-Naji *et al.* [22] proposed a multi-focus image fusion technique which fuses the images based on the geometric mean of the largest five eigenvalues obtained through the singular value decomposition (SVD) on the input blocks in DCT domain. In [20], the authors

proposed four approaches, namely DCT+EOL, DCT+VOL, DCT+Corr and DCT+Corr_Eng for multi-focus fusion. In DCT+EOL and DCT+VOL, the authors proposed the use of the energy of Laplacian and the variance of Laplacian for focus measure computation. Focus measure based on the correlation coefficient between image blocks in DCT domain is shown to perform well in multi-focus image fusion [23]. The authors in [20] proposed DCT+Corr and DCT+Corr_Eng, which use the correlation coefficient and the energy correlation coefficient between source blocks and artificially blurred blocks as the focus measure. Further, considering existing consistency verification in the approaches, the authors showed that their techniques were superior in performance. Moreover, the authors also suggested repeated consistency verification for better consistency, and therefore, better performance.

## IV. PROPOSED FOCUS MEASURE

In an all-in-focus system, the crucial part is to fuse the most informative and relevant information, that is, the focused areas in multiple images. An appropriate focus measure determines the focused area based on sharpness quantification [25]. As mentioned already, the sum of modified Laplacian (SML) is such an efficient focus measure which is known to perform well.

The application of $\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ on a 2D signal $F$ is known as the Laplacian operation, which involves double derivatives on $F$ in both the $x$ and $y$ directions. If such an operation generates responses of opposite polarities in the two directions, the overall response gets diminished. As the magnitude of the responses in the two directions quantify the amount of focus, Nayar *et. al.* [25] proposed the sum of modified Laplacian (SML) to measure focus as follows

$$SML(x, y) = \left|\frac{\partial^2 f_{x,y}}{\partial x^2}\right| + \left|\frac{\partial^2 f_{x,y}}{\partial y^2}\right| \quad (15)$$

We extend it for use with $N \times N$ 2D signal and denote the measure as $SML_L$ defined using our $N \times N$ $L$ operator. Note that $SML_L$ is the additive response of $R_r$ and $R_c$, which are defined in (6), where $P$ is to be replaced by $L$. In this case (using $L$) the responses ($R'_r$ and $R'_c$) are as follows

$$R'_r = LF$$
$$R'_c = FL^T \quad (16)$$

where $R'_r$ represents $\frac{\partial^2 F}{\partial x^2}$ and $R'_c$ represents $\frac{\partial^2 F}{\partial y^2}$ for $N \times N$ block signal $F$. Now, let us define SML using $L$ operator for $N \times N$ block in DCT domain as $SML_L^{DCT}$ given by

$$SML_L^{DCT} = \sum |DCT(R'_r)| + |DCT(R'_c)| \quad (17)$$

where the summation $\sum(.)$ is over all the matrix elements yielding a single value. combining (16), (17), and $L = L^T$, we have

$$SML_L^{DCT} = \sum |DCT(LF)| + |DCT(FL)| \quad (18)$$

Now $DCT(LF) = DCT(L)DCT(F)$. As $DCT(L)$ is a diagonal matrix having non-negative diagonal values, we will have $|DCT(R'_c)| = |DCT(F)|DCT(L)$ and $|DCT(R'_r)| = DCT(L)|DCT(F)|$. Therefore, $SML_L^{DCT}$ reduces to

$$SML_L^{DCT} = \sum DCT(L)|DCT(F)| + |DCT(F)|DCT(L) \quad (19)$$

Now, $DCT(LF)$ is $T^1$ in (13) and $DCT(FL)$ is $T^2$ in (14). Combining expressions of (13), (14) and (19), we have,

$$SML_L^{DCT} = \sum_{i=1}^{N} \lambda_{N+1-i,N+1-i} \sum_{j=1}^{N} |f_{i,j}^D|$$
$$+ \sum_{i=1}^{N} \lambda_{N+1-i,N+1-i} \sum_{j=1}^{N} |f_{j,i}^D|$$
$$= \sum_{i=1}^{N} \lambda_{N+1-i,N+1-i} \sum_{j=1}^{N} |f_{i,j}^D| + |f_{j,i}^D| \quad (20)$$

From the above expression, we see that the focus measure is nothing but a weighted sum of the absolute values of the DCT block coefficients, where the weights are the eigenvalues of $L$. From our formulation in (20), we infer that SML based focus measure computed in the spatial domain can be directly represented in the DCT domain. Moreover, as discussed in Section II-B, the formulation of (20) reduces the hardware-unfriendly multiplication operations from $2N^3$ to just $N$.

Fixed point operations are always hardware-friendly over floating point operations. To consider fixed point opearation in the DCT domain, Binary DCT [31], optimized Integer DCT [32], and efficient Integer DCT [15], [33] are employed in compression/coding framework. Inspired by this, we propose to compute the focus measure on the integer valued DCT coefficients. To further improve on hardware-friendliness, we modify expression (20) and propose the following as our focus measure (FM)

$$FM = \sum_{i=1}^{N} \lambda^o_{N+1-i,N+1-i} \sum_{j=1}^{N} |f^o_{i,j}| + |f^o_{j,i}| \quad (21)$$

where $\Lambda^o$ and $F^o$ are the nearest integer values of $\Lambda$ and $F^D$, respectively. To summarize, our focus measure is hardware-friendly for the following reasons.

1) For an $N \times N$ block, the focus measure computation requires $N$ integer multiplications compared to $N^3$ floating point multiplications required in DCT+SF [19] and DCT+SF+CV [19].
2) In our formulation of (21), the weights and modified DCT coefficients are all integer valued.

Equation (21) can be represented as follows

$$FM = \sum_{i=1}^{N} \sum_{j=1}^{N} (m^o_{i,i} + m^o_{j,j}) |f^o_{i,j}| = \sum M^o |F^o| \quad (22)$$

where $M^o$ is the diagonal matrix which is the nearest integer values of $M$ defined in (12). The diagonal terms of $M^o$ are $m^o_{i,i} = \lfloor(\lambda_{N+1-i,N+1-i})\rceil$, where $\lfloor\rceil$ indicates rounding to nearest integer. $W^o$ has the following properties which are beneficial for focus computation [10]

1) Weight corresponding to the DC element ($f^o_{1,1}$), that is $w^o_{1,1} = \lambda_{N,N} = 0$, as $L$ is a positive semi-definite matrix.

2) Due to the round off operation to the nearest integer, weights corresponding to the lower frequency components (in $F^o$) become 0. For example, if we consider a signal block size $8 \times 8$ for processing, we have $w^o_{1,2} = \lfloor\lambda_{N-1,N-1}\rceil = 0$, $w^o_{2,1} = \lfloor\lambda_{N-1,N-1}\rceil = 0$ and $w^o_{2,2} = \lfloor 2 \times \lambda_{N-1,N-1}\rceil = 0$. So, $w^o_{i,j} = 0$, for $i \leq 2$ and $j \leq 2$, which correspond to lower frequencies with $w^o_{i,j}$ for $i > 2$ and $j > 2$ corresponding to high frequencies.

3) All the weights are integer-valued, and weights either increase or remain the same (monotonically non-decreasing function) with increasing frequency.

As mentioned already, it is well understood that when an image area is not focused, it is blurred. Such areas contain substantially more low frequency content than high frequency ones. On the other hand, when an area (except smooth area) is focused, it is comparatively sharper, comprising of high frequency content. Our focus measure exactly takes these aspects into account. First two properties of $W^o$ basically suggest that the influence of blurriness is discarded, and the third property shows that sharpness is included with the high frequencies getting higher weights.

## V. PROPOSED MULTI-FOCUS IMAGE FUSION

As mentioned earlier, VSN adopts image compression or video coding for huge visual data to be transmitted. Figure 1 shows a schematic diagram of an all-in-focus system for VSN [16], [18], [19], where the framework considers DCT-based compression technique on $8 \times 8$ image/video frame blocks. Similar to [18], [19], let us consider the JPEG encoder framework, where the input image is divided into $8 \times 8$ non-overlapping blocks. Then, for an 8-bit image, levels in the image are shifted from [0-255] to [-128-127], after which block DCT transformation is performed. The DCT coefficients are quantized with a pre-defined quantization matrix. The quantized DCT coefficients are then scanned in a zigzag fashion for the application of run-length coding. Finally, the coded DCT coefficients are sent in the form of a bitstream to the receiving end. At the receiving end, the order of the processes is reversed, that is, decoding of the bitstream, dequantization, inverse DCT operation, and level shifting operation in that order are carried out. In such a compression framework for VSN application [18], [19], a multi-focus image fusion operation on DCT coefficients is the best suited one.

For simplicity, we will explain the fusion of two images/ video frames, and an extension to multiple images/video



**FIGURE 1.** A process flow diagram of a general (e.g. JPEG) compression framework combined with multi-focus image fusion. Part of the figure enclosed in dashed block is the added module required for multi-focus image fusion.

frames is straightforward. Our proposed approach consists of three modules, namely, estimation of a decision map for
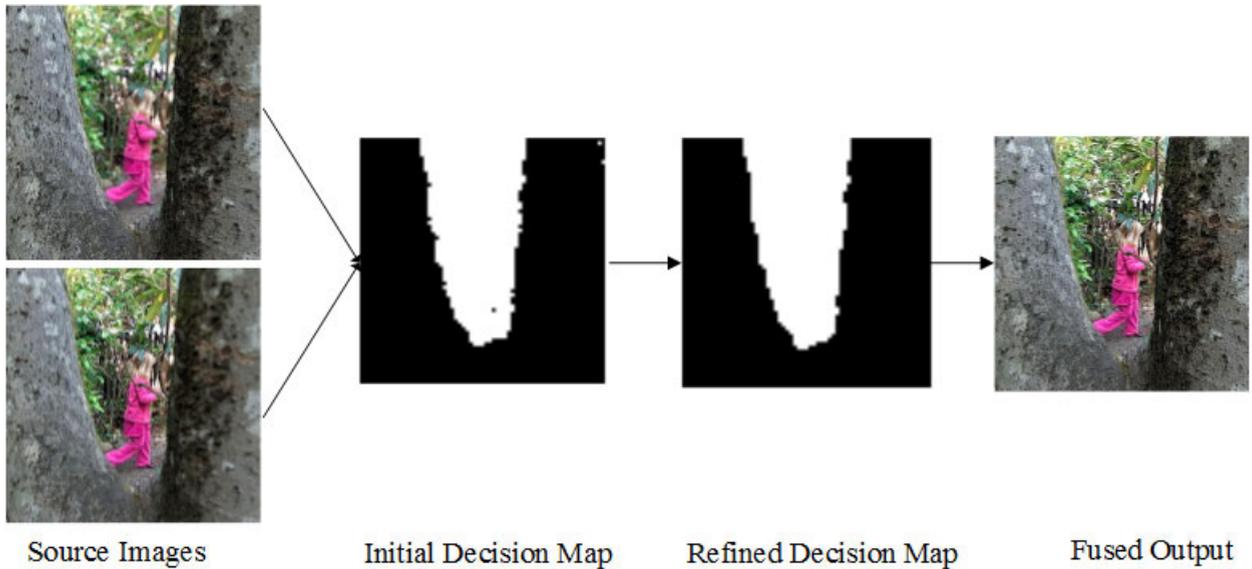
**FIGURE 2.** Initial and refined decision maps generated in our fusion approach to get the fused output.

fusion through our proposed focus measure, refinement of the decision map based on a spatial neighborhood relation, and finally, the fusion based on the refined decision map. The modules of the proposed approach are elaborated below.

### A. COMPUTATION OF FOCUS MEASURE

The received encoded bitstream of the two images are first decoded and then dequantized. For the input images, say $A$ and $B$, let us consider the $(x, y)^{th}$ block as $A_{x,y}$ and $B_{x,y}$, respectively. Then, our focus measure (FM) is computed for both the blocks using (21). Let us denote the resulting FM values as $FM^A_{x,y}$ and $FM^B_{x,y}$ for $A_{x,y}$ and $B_{x,y}$ blocks, respectively. Note that a higher value of FM for a block means that it is more focused (less out of focus).

### B. ESTIMATION OF INITIAL DECISION MAP

In the decision map, for each block, a decision is assigned in the form of $+1/-1$, where $+1$ indicates that at the time of fusion, the block from $A_{x,y}$ needs to be selected, whereas $-1$ indicates $B_{x,y}$ needs to be selected. We generate the initial decision map ($D^M$) as follows,

$$d^M_{x,y} = \begin{cases} +1, & FM\_A_{x,y} \geq FM\_B_{x,y} \\ -1, & FM\_A_{x,y} < FM\_B_{x,y} \end{cases} \quad (23)$$

Note that the above decision rule does not allow any provision of ambiguous case, that is, a case where there may be confusion of belongingness to a particular class. Moreover, the map $D^M$ takes the decision based only on the focus measure. So, this initial decision map may suffer from spatial inconsistencies. For example, it may happen that a block belongs to a particular class, whereas most of its surrounding blocks belong to the other class. Thus, a refinement is required, considering local spatial consistency.

### C. REFINED DECISION MAP

Inconsistencies in the initial decision map may lead to generation of unwanted artifacts in the output image. The common consistency verification which is adopted in a few recent DCT based all-in-focus system is majority filter [10], [18], [19], [34]. In it, the majority in the neighborhood decides the class belongingness of the center being operated on. Sometimes [18] such an operation is carried twice sequentially for better consistency in the decision map, and therefore, we also do so but in a different manner. However in our case, we do a moving summation operation on the initial decision map to generate a modified decision map followed by another moving summation operation on the sign values of the modified decision map to generate the final refined decision map. Therefore, the refined decision map $R^M$ is obtained as follows

$$r^M_{i,j} = \sum_{x=i-m}^{i+m} \sum_{y=j-n}^{j+n} d^M_{x,y} \quad (24)$$

$$r^M_{i,j} = \sum_{x=i-m}^{i+m} \sum_{y=j-n}^{j+n} sign(r^M_{x,y}) \quad (25)$$

where we consider, the neighboring (8-neighbor) blocks around the current central block of operation to incorporate the local spatial influence. The above operations in our refined decision map are computationally simple as elaborated in Section VI-C. Figure 2 shows examples of the decision maps.

### D. FUSION

The inconsistencies in the initial decision map have now been handled by the refinement operation. Then, based on the refined decision map, we fuse the images choosing a block

either from **A** or **B**. So, the fusion is as follows,

$$O_{i,j} = \begin{cases} A_{i,j}, & r_{i,j}^M > 0 \\ B_{i,j}, & r_{i,j}^M < 0 \end{cases} \quad (26)$$

Finally, the resulting DCT coefficients are quantized and coded, after which they are sent in form of the bitstream to receiving end, as mentioned earlier.

## VI. EXPERIMENTAL RESULTS AND DISCUSSION

To analyze the effectiveness of our approach, we compare the proposed approach with existing technique applicable for VSN such as DCTav [16], [17], DCTcm [16], [17], DCTma [17], DCTe [17], DCTah [17], DCTch [17], DCTv [18], DCTvc [18], DCTs [19] and DCTsc [19], DCT+SVD [22], DCT+SVD+CV [22], DCT+EOL [20], DCT+EOL+CV [20],DCT+VOL [20], DCT+VOL+ CV [20], DCT+Corr_Eng [20] and DCT+Corr_Eng+ CV [20]. For our approach, we present results using only the initial decision map (Proposed) and using the refined decision map (Proposed+CV). We further compare our approach with a few latest generic multi-focus image fusion techniques that represent the state-of-the-art such as MFCNN [37], MFGFDF [38], MFRW [39], MADCNN [40] and ECNN [41].

First, we perform quantitative and subjective evaluation with multi-focus image fusion techniques for VSN. For quantitative evaluation, we perform no-reference evaluation, where outputs are evaluated with standard no-reference measures, as the ground truths are not available. Then, we perform a computation time evaluation on three image sizes with recent state-of-the-art techniques. For subjective evaluation, we present the visual comparison of the output images generated by the different techniques. We further present the initial and refined decision maps generated by the recent state-of-the-art techniques and our approach. Finally, we discuss the hardware friendliness and computational simplicity of the proposed approach, which make it suitable for VSN. In addition, we evaluate our approach quantitatively and qualitatively comparing it to the state-of-the-art and the latest generic multi-focus image fusion techniques.

### A. QUANTITATIVE EVALUATION

As the ground truths of actual multi-focus images are not available, most of the techniques rely on evaluating the algorithm based on standard no-reference measures [28], [42]–[49]. Therefore, we evaluate the performance of the approaches on the standard multi-focus image database provided by the authors in [35] and Lytro [36] based on following standard no-reference quality measures for image fusion:

- FMI ($Q_{FMI}$) [43], [44]:
  FMI of Haghighat et al. is an improved version of the widely accepted mutual information based measures for image fusion given by [50] and [51]. FMI measures the amount of information transferred from each of the input images to the fused image.

- Xydeas and Petrovic measure ($Q^{AB/F}$) [45]:
  Xydeas and Petrovic have proposed a no-reference quality measure, which includes the universal quality index of Wang and Bovik [52]. The measure computes the amount of gradient information (edge) transferred from each of the input images to the fused image. The measured value ranges from 0 to 1, with 0 indicating the worst result and 1 the best fused result.

- Pellia and Heijmans measure ($Q^W$) [46]:
  Pellia and Heijmans have proposed a no-reference quality measure, which also includes the universal quality index of Wang and Bovik [52]. The measure computes the amount of salient information transferred from each of the input images to the fused image without introducing distortion. The measure also includes similarity index [53] and human visual system sensitive edge information. Higher the value of the measure, better is the fused output.

- Yang et al. ($Q^Y$) [47]
  Yang et al. have proposed an image fusion quality measure, where complementary or conflicting regions of the source images are selected from redundant regions and structural similarity feature of SSIM is applied separately. The measure is shown to be consistent with human visual evaluations. Higher the value of the measure, better is the fused output.

- Chen and Blum ($Q^{CB}$) [48]
  Chen and Blum have proposed an HVS based measure, which computes the amount of contrast features transferred from each of the input images to the fused image. Higher the value of the measure, better is the fused output.

- Zhao et al. ($Q^P$) [49]
  Zhao et al. have proposed a phase congruency based measure, which computes the amount of corner and edge information transferred from each of the input images to the fused image. Higher the value of the measure, better is the fused output.

**TABLE 1.** Performance comparison of multi-focus image fusion techniques for VSN based on no-reference quality measures (best result in bold). Values are the average measures on 18 pairs of multi-focus images (shown in Figure 3) from the database given by [35].

| Techniques | Measures | | | | | |
|---|---|---|---|---|---|---|
| | $Q^P$ | $Q^Y$ | $Q^{CB}$ | $Q^W$ | $Q^{AB/F}$ | $Q^{FMI}$ |
| DCTav | 0.684626 | 0.846461 | 0.645878 | 0.858908 | 0.576328 | 0.52968 |
| DCTcm | 0.62666 | 0.891506 | 0.677677 | 0.887738 | 0.624925 | 0.500087 |
| DCTma | 0.629277 | 0.893086 | 0.677055 | 0.888339 | 0.62747 | 0.5018 |
| DCTe | 0.643645 | 0.900435 | 0.680853 | 0.888393 | 0.646795 | 0.511754 |
| DCTah | 0.632845 | 0.871415 | 0.634714 | 0.870826 | 0.587355 | 0.474087 |
| DCTv | 0.737727 | 0.945009 | 0.732916 | 0.887071 | 0.72176 | 0.603651 |
| DCTvc | 0.802398 | 0.979223 | 0.774756 | 0.891871 | 0.734682 | 0.630196 |
| DCTs | 0.771352 | 0.948223 | 0.741154 | 0.891563 | 0.727816 | 0.592686 |
| DCTsc | 0.803894 | 0.97052 | 0.767397 | 0.892696 | 0.734246 | 0.620918 |
| DCT+SVD | 0.755091 | 0.957192 | 0.743148 | 0.886968 | 0.723834 | 0.610683 |
| DCT+SVD+CV | 0.795008 | 0.979018 | 0.770677 | 0.888801 | 0.731912 | 0.628412 |
| DCT+EOL | 0.754103 | 0.9551 | 0.737507 | 0.889988 | 0.724085 | 0.609271 |
| DCT+EOL+CV | 0.783496 | 0.974198 | 0.763163 | 0.8903 | 0.727746 | 0.624949 |
| DCT+VOL | 0.754104 | 0.9551 | 0.737515 | 0.889988 | 0.724085 | 0.609276 |
| DCT+VOL+CV | 0.783496 | 0.974198 | 0.763163 | 0.8903 | 0.727746 | 0.624949 |
| DCT+Corr_Eng | 0.770531 | 0.961196 | 0.746972 | 0.891891 | 0.730833 | 0.614217 |
| DCT+Corr_Eng+CV | 0.797672 | 0.978964 | 0.771888 | 0.89236 | 0.734526 | 0.62906 |
| Proposed | 0.77092 | 0.95776 | 0.74546 | 0.891762 | 0.73067 | 0.612878 |
| Proposed+CV | **0.806552** | **0.981288** | **0.777909** | **0.892931** | **0.736382** | **0.631641** |

Table 1 presents the average results obtained for the 18 pairs of images from the database provided in [35]. One image

**FIGURE 3.** Images of database shared by [35] used for evaluation.



**FIGURE 4.** Images of Lytro [36] database used for evaluation.

**TABLE 2.** Performance comparison of multi-focus image fusion techniques for VSN based on no-reference quality measures (best result in bold). Values are the average measures on 20 pairs of multi-focus images (Figure 4) from Lytro database [36].

| Techniques | Measures | | | | | |
|---|---|---|---|---|---|---|
| | $Q^P$ | $Q^Y$ | $Q^{CB}$ | $Q^W$ | $Q^{AB/F}$ | $Q^{FMI}$ |
| DCTav | 0.754537 | 0.847164 | 0.625876 | 0.883818 | 0.585472 | 0.500611 |
| DCTcm | 0.706037 | 0.919389 | 0.680054 | 0.921788 | 0.671816 | 0.498642 |
| DCTma | 0.706838 | 0.919687 | 0.679726 | 0.921351 | 0.671588 | 0.499311 |
| DCTe | 0.725846 | 0.926015 | 0.683203 | 0.922236 | 0.690306 | 0.509151 |
| DCTah | 0.693889 | 0.867218 | 0.622349 | 0.891647 | 0.599697 | 0.443263 |
| DCTv | 0.789731 | 0.960006 | 0.763938 | 0.922956 | 0.745344 | 0.585728 |
| DCTvc | 0.8361 | 0.984403 | 0.80495 | 0.92501 | 0.75792 | 0.608952 |
| DCTs | 0.817352 | 0.960613 | 0.766419 | 0.927597 | 0.751385 | 0.569688 |
| DCTsc | 0.83822 | 0.980825 | 0.801335 | 0.927388 | 0.758523 | 0.602682 |
| DCT+SVD | 0.81663 | 0.96925 | 0.781002 | 0.928122 | 0.754129 | 0.593757 |
| DCT+SVD+CV | 0.840151 | 0.984038 | 0.803881 | 0.929008 | 0.759748 | 0.608566 |
| DCT+EOL | 0.824876 | 0.972539 | 0.78559 | 0.92837 | 0.756902 | 0.59697 |
| DCT+EOL+CV | 0.839949 | 0.985669 | 0.804738 | 0.928686 | 0.759892 | 0.608867 |
| DCT+VOL | 0.824806 | 0.972549 | 0.785599 | 0.928379 | 0.756902 | 0.59697 |
| DCT+VOL+CV | 0.839949 | 0.985669 | 0.804738 | 0.928686 | 0.759892 | 0.608867 |
| DCT+Corr_Eng | 0.82584 | 0.973052 | 0.785572 | 0.928715 | 0.757102 | 0.597011 |
| DCT+Corr_Eng+CV | 0.839889 | 0.98552 | 0.804337 | 0.928686 | 0.759886 | 0.608741 |
| Proposed | 0.82299 | 0.97251 | 0.78434 | 0.92848 | 0.75655 | 0.59596 |
| Proposed+CV | **0.84254** | **0.986366** | **0.808567** | **0.929021** | **0.76064** | **0.610169** |

**TABLE 3.** Performance comparison of multi-focus image fusion techniques on the image Girl (shown in Figure 9) from [35] based on no-reference quality measures (best result in bold).

| Techniques | Measures | | | | | |
|---|---|---|---|---|---|---|
| | $Q^P$ | $Q^Y$ | $Q^{CB}$ | $Q^W$ | $Q^{AB/F}$ | $Q^{FMI}$ |
| DCTav | 0.682355 | 0.855686 | 0.692504 | 0.86621 | 0.520467 | 0.518554 |
| DCTcm | 0.625948 | 0.918899 | 0.701264 | 0.863736 | 0.559376 | 0.482116 |
| DCTma | 0.625353 | 0.919333 | 0.70164 | 0.863789 | 0.559492 | 0.482403 |
| DCTe | 0.642693 | 0.930239 | 0.69999 | 0.86182 | 0.585107 | 0.504453 |
| DCTah | 0.621108 | 0.895582 | 0.684978 | 0.868071 | 0.534656 | 0.454823 |
| DCTv | 0.767697 | 0.970559 | 0.752792 | 0.856582 | 0.675896 | 0.583657 |
| DCTvc | 0.83523 | 0.989939 | 0.787419 | 0.859741 | 0.691663 | 0.603898 |
| DCTs | 0.821593 | 0.984197 | 0.770944 | 0.861091 | 0.690657 | 0.597222 |
| DCTsc | 0.838083 | 0.991052 | 0.789143 | 0.860806 | 0.693014 | 0.604636 |
| DCT+SVD | 0.820133 | 0.984916 | 0.782144 | 0.861347 | 0.69026 | 0.598684 |
| DCT+SVD+CV | 0.83766 | 0.990857 | 0.788978 | 0.860705 | 0.693079 | 0.604711 |
| DCT+EOL | 0.787653 | 0.978166 | 0.755094 | 0.856971 | 0.678908 | 0.592902 |
| DCT+EOL+CV | 0.798767 | 0.983026 | 0.769396 | 0.856933 | 0.68055 | 0.597555 |
| DCT+VOL | 0.787653 | 0.978166 | 0.755094 | 0.856933 | 0.678908 | 0.592902 |
| DCT+VOL+CV | 0.798767 | 0.983026 | 0.769396 | 0.856933 | 0.68055 | 0.597555 |
| DCT+Corr_Eng | 0.812843 | 0.983981 | 0.770296 | 0.860373 | 0.689354 | 0.598726 |
| DCT+Corr_Eng+CV | 0.837704 | 0.991032 | 0.789501 | 0.860812 | 0.692594 | 0.604778 |
| Proposed | 0.827807 | 0.986497 | 0.773212 | **0.860898** | 0.691468 | 0.600304 |
| Proposed+CV | **0.839439** | **0.991638** | **0.789812** | 0.860821 | **0.693287** | **0.60536** |

**TABLE 4.** Performance comparison of multi-focus image fusion techniques on the image Grass (shown in Figure 7) from [35] based on no-reference quality measures (best result in bold).

| Techniques | Measures | | | | | |
|---|---|---|---|---|---|---|
| | $Q^P$ | $Q^Y$ | $Q^{CB}$ | $Q^W$ | $Q^{AB/F}$ | $Q^{FMI}$ |
| DCTav | 0.567551 | 0.818494 | 0.631225 | 0.845391 | 0.489766 | 0.475491 |
| DCTcm | 0.493284 | 0.881646 | 0.656181 | 0.842863 | 0.505234 | 0.43705 |
| DCTma | 0.49317 | 0.881018 | 0.658097 | 0.842241 | 0.504717 | 0.436747 |
| DCTe | 0.501225 | 0.891449 | 0.661488 | 0.84047 | 0.526939 | 0.452474 |
| DCTah | 0.516295 | 0.882993 | 0.620043 | 0.852172 | 0.504135 | 0.428867 |
| DCTv | 0.625604 | 0.943898 | 0.715074 | 0.821239 | 0.637941 | 0.54536 |
| DCTvc | 0.708542 | 0.981507 | 0.757271 | 0.83533 | 0.666159 | 0.574559 |
| DCTs | 0.672967 | 0.96938 | 0.738794 | 0.832833 | 0.660997 | 0.56287 |
| DCTsc | 0.696524 | 0.982532 | 0.757885 | 0.835231 | 0.666563 | 0.576623 |
| DCT+SVD | 0.685992 | 0.970795 | 0.744877 | 0.835231 | 0.661556 | 0.564857 |
| DCT+SVD+CV | 0.708611 | 0.986011 | 0.761117 | 0.838436 | 0.669812 | 0.577972 |
| DCT+EOL | 0.633206 | 0.962026 | 0.722839 | 0.830499 | 0.649736 | 0.556317 |
| DCT+EOL+CV | 0.652272 | 0.973428 | 0.736205 | 0.832689 | 0.653402 | 0.566682 |
| DCT+VOL | 0.633206 | 0.962026 | 0.722839 | 0.830499 | 0.649736 | 0.556317 |
| DCT+VOL+CV | 0.652272 | 0.973428 | 0.736205 | 0.832689 | 0.653402 | 0.566682 |
| DCT+Corr_Eng | 0.677444 | 0.971185 | 0.744589 | 0.66137 | 0.66137 | 0.565958 |
| DCT+Corr_Eng+CV | 0.701892 | 0.983276 | 0.759226 | 0.837671 | 0.667649 | 0.576521 |
| Proposed | 0.697877 | 0.976819 | 0.749446 | 0.83726 | 0.66846 | 0.56940 |
| Proposed+CV | **0.70965** | **0.986166** | **0.761974** | **0.83866** | **0.670560** | **0.578252** |

**TABLE 5.** Performance comparison of multi-focus image fusion techniques on the image Temple (shown in Figure 8) from [35] based on no-reference quality measures (best result in bold).

| Techniques | Measures | | | | | |
|---|---|---|---|---|---|---|
| | $Q^P$ | $Q^Y$ | $Q^{CB}$ | $Q^W$ | $Q^{AB/F}$ | $Q^{FMI}$ |
| DCTav | 0.535549 | 0.769406 | 0.482256 | 0.764342 | 0.406198 | 0.489361 |
| DCTcm | 0.474515 | 0.868115 | 0.595027 | 0.862927 | 0.603607 | 0.448845 |
| DCTma | 0.477549 | 0.870364 | 0.596019 | 0.86479 | 0.60758 | 0.452048 |
| DCTe | 0.481468 | 0.873465 | 0.60097 | 0.865227 | 0.630088 | 0.468045 |
| DCTah | 0.470461 | 0.812117 | 0.482781 | 0.796312 | 0.472882 | 0.403019 |
| DCTv | 0.622575 | 0.931432 | 0.726309 | 0.874378 | 0.75157 | 0.551869 |
| DCTvc | 0.747824 | 0.980881 | 0.793125 | 0.886276 | 0.762931 | 0.574517 |
| DCTs | 0.705404 | 0.960762 | 0.76896 | 0.885908 | **0.764755** | 0.56351 |
| DCTsc | 0.76388 | 0.986295 | 0.801436 | 0.887538 | 0.764006 | 0.576079 |
| DCT+SVD | 0.758771 | 0.97945 | 0.792728 | 0.884087 | 0.76196 | 0.572637 |
| DCT+SVD+CV | 0.789268 | 0.993057 | 0.807982 | 0.883066 | 0.761132 | 0.578546 |
| DCT+EOL | 0.723929 | 0.97382 | 0.775705 | 0.884084 | 0.753079 | 0.567918 |
| DCT+EOL+CV | 0.761453 | 0.988605 | 0.795365 | 0.883376 | 0.752531 | 0.574566 |
| DCT+VOL | 0.723929 | 0.97382 | 0.775705 | 0.884084 | 0.753079 | 0.567918 |
| DCT+VOL+CV | 0.761453 | 0.988605 | 0.795365 | 0.883376 | 0.752531 | 0.574566 |
| DCT+Corr_Eng | 0.747601 | 0.980181 | 0.793567 | **0.887596** | 0.762531 | 0.572795 |
| DCT+Corr_Eng+CV | 0.785412 | 0.992951 | 0.808772 | 0.886647 | 0.762627 | 0.578484 |
| Proposed | 0.761616 | 0.98338 | 0.794689 | 0.886797 | 0.763482 | 0.57407 |
| Proposed+CV | **0.791191** | **0.993736** | **0.809501** | 0.884092 | 0.762805 | **0.579051** |

from each pair is shown in Figure 3. We also present the average results obtained for the 20 pairs of images from the Lytro database [36]. One image from each pair is shown in Figure 4. We consider four individual images namely, Girl, Grass, Temple and Lytro-16 to present their quantitative performance in Tables 3, 4, 5 and 6. The results show that
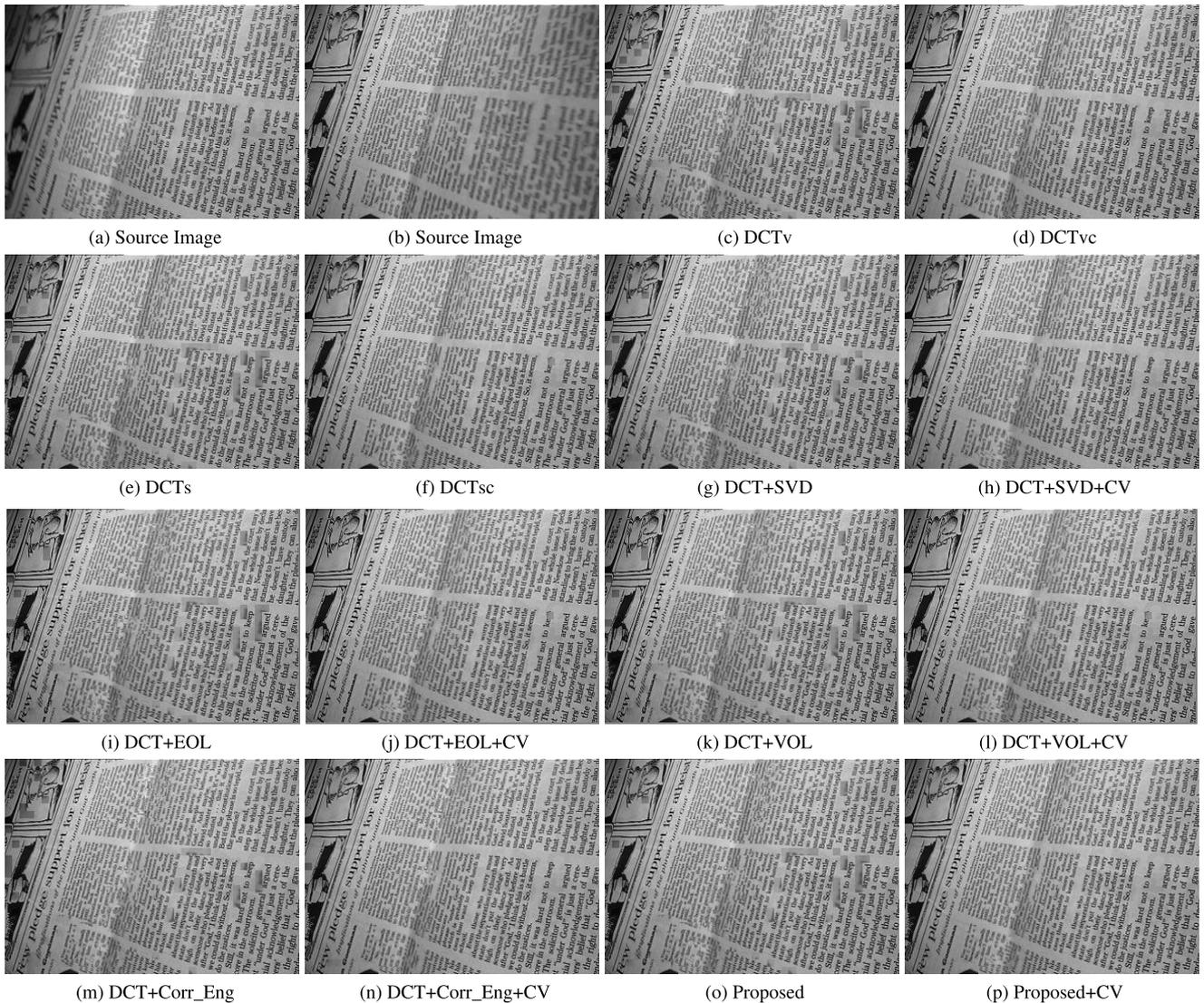
**FIGURE 5.** Source images: The 'Newspaper' pair of images from the database given by [35] and multi-focus image fusion with different techniques.

**TABLE 6.** Performance comparison of multi-focus image fusion techniques on the image Lytro-16 from [36] based on no-reference quality measures (best result in bold).

| Techniques | Measures | | | | | |
|---|---|---|---|---|---|---|
| | $Q^P$ | $Q^Y$ | $Q^{CB}$ | $Q^W$ | $Q^{AB/F}$ | $Q^{FMI}$ |
| DCTav | 0.731407 | 0.86192 | 0.633311 | 0.893698 | 0.63663 | 0.477702 |
| DCTcm | 0.694482 | 0.920158 | 0.670675 | 0.939085 | 0.702928 | 0.453413 |
| DCTma | 0.694992 | 0.920322 | 0.668083 | 0.939116 | 0.703045 | 0.453763 |
| DCTe | 0.723718 | 0.931818 | 0.670566 | 0.940128 | 0.72216 | 0.466781 |
| DCTah | 0.682332 | 0.885901 | 0.629209 | 0.903698 | 0.649151 | 0.432088 |
| DCTch | 0.80757 | 0.967954 | 0.732578 | 0.943942 | 0.774618 | 0.569389 |
| DCTv | 0.855092 | 0.984818 | 0.759399 | 0.946149 | 0.785893 | 0.600425 |
| DCTvc | 0.837064 | 0.961974 | 0.732047 | 0.946301 | 0.778619 | 0.551777 |
| DCTs | 0.85551 | 0.981949 | 0.757173 | 0.945997 | 0.785566 | 0.596166 |
| DCTsc | 0.831655 | 0.97158 | 0.739777 | 0.946224 | 0.781649 | 0.578819 |
| DCT+SVD | 0.83564 | 0.976407 | 0.749321 | 0.94648 | 0.781868 | 0.583944 |
| DCT+SVD | 0.855075 | 0.985679 | 0.760728 | 0.946648 | 0.786543 | 0.60218 |
| DCT+EOL | 0.831665 | 0.97160 | 0.739781 | 0.946228 | 0.781649 | 0.578819 |
| DCT+EOL+CV | 0.853867 | 0.98481 | 0.759277 | 0.945819 | 0.78610 | 0.601759 |
| DCT+VOL | 0.831655 | 0.97158 | 0.739777 | 0.946224 | 0.781643 | 0.578815 |
| DCT+VOL+CV | 0.853862 | 0.984809 | 0.759272 | 0.945815 | 0.786105 | 0.601762 |
| DCT_Corr_Eng | 0.830904 | 0.971562 | 0.739664 | 0.946313 | 0.781359 | 0.578098 |
| DCT_Corr_Eng+CV | 0.855213 | 0.984311 | 0.758577 | **0.946677** | 0.786523 | 0.60121 |
| Proposed | 0.838565 | 0.975401 | 0.746185 | 0.946264 | 0.782799 | 0.5832 |
| Proposed+CV | **0.855857** | **0.985964** | **0.760527** | 0.946314 | **0.78672** | **0.603663** |

**TABLE 7.** Comparison of CPU processing time (in seconds) of few recent state-of-the-art multi-focus image fusion techniques (Best in bold and second best in italic).

| Techniques | Image Size | | |
|---|---|---|---|
| | $256 \times 256$ | $512 \times 512$ | $1024 \times 1024$ |
| DCT+SVD | 0.020 sec | 0.112 sec | 0.421 sec |
| DCT+SVD+CV | 0.028 sec | 0.119 sec | 0.432sec |
| DCT+EOL | 0.025 sec | 0.118 sec | 0.390 sec |
| DCT+EOL+CV | 0.031 sec | 0.126 sec | 0.403 sec |
| DCT+VOL | 0.020 sec | 0.108 sec | 0.382 sec |
| DCT+VOL+CV | 0.026 sec | 0.117 sec | 0.395 sec |
| DCT+Corr_Eng | 0.063 sec | 0.262 sec | 0.965 sec |
| DCT+Corr_Eng+CV | 0.070 sec | 0.271 sec | 1.00 sec |
| Proposed | **0.012** sec | **0.052** sec | **0.235** sec |
| Proposed+CV | *0.016 sec* | *0.060 sec* | *0.245 sec* |

and in terms of $Q^W$ in Table 6 where our 'Proposed+CV' performs very close to the best.

The above quantitative evaluation on two databases and four individual multi-focus image pairs show the superiority

our 'Proposed+CV' outperforms the rest in terms of all the six measures except in terms of $Q^W$ and $Q^{AB/F}$ in Table 5

**FIGURE 6.** Source images: The 'Book' pair of images from database given by [35] and multi-focus image fusion with different techniques.

of our approach. All the techniques are run on Matlab® platform in a system with Intel® Core(TM) i5-4590 CPU @ 3.30 GHz having 16 GB RAM. Now, we present the computation time comparison on three image sizes with recent state-of-the-art techniques in Table 7. The result shows that our approach with the initial decision map (Proposed) is faster than the rest of the recent state-of-the-art techniques. It also shows that Proposed+CV is better than all the other techniques shown here except for our approach with the initial decision map (Proposed). It further shows that the superiority of our approach is more evident with the increase in image size.

### B. SUBJECTIVE EVALUATION
In Figure 5, 6, 7, 8, 9 and 10, we perform subjective evaluation of six standard images, namely, Newspaper, Grass, Temple, Girl, Book, and Pepsi, respectively, from the

standard multi-focus image fusion database [35]. We compare our approach with best performing and recent/state-of-the-art techniques like DCTv [18], DCTvc [18], DCTs [19] and DCTsc [19], DCT+SVD [22], DCT+SVD+CV [22], DCT+EOL [20], DCT+EOL+CV [20], DCT+VOL [20], DCT+VOL+CV [20], DCT+Corr_Eng [20] and DCT+Corr_Eng+CV [20]. 'Newspaper' is one of the critical image pair where our 'Proposed' performs as good as or better than other techniques without consistency verification (CV). Further, the 'Proposed+CV' performs as good as or better than other techniques with or without consistency verification (CV). 'Book' is another critical image pair, where we can see that our approach performs the fusion properly. We can see that our approach with consistency verification (CV) is better than a few approaches like DCTvc, DCTsc, etc., and produces less artifacts (See the text 'terms' in the image). A similar performance can be seen from the results of other

**FIGURE 7.** Source images: The 'Grass' pair of images from the database given by [35] and multi-focus image fusion with different techniques.

test image pairs, where our approach performs as good as or better than the rest of the techniques.

We present comparative subjective results of the initial and final maps generated by recent/state-of-the-art techniques including ours in Figure 11. Note that the final decision map of our approach and DCT+SVD+CV are binary, as both perform fusion selecting a block from either of the two images, whereas DCT+EOL+CV, DCT+VOL+CV and DCT+Corr_Eng+CV have gray final maps as the selection is either a block in one of the images or an average of the two. The figure shows that our decision maps determine the focused area with limited noise and initial and final maps are as good as or better than state-of-the-art techniques.

### C. HARDWARE-FRIENDLY SOLUTION

Quantitative and subjective evaluation show that our approach is as good as or better than the other techniques. But as discussed in Sections I and II, VSN demands a computationally simple, hardware-friendly solution, which is our main motivation.

As shown in Section IV, our focus measure computation requires $N$ fixed point multiplications over $N^3$ floating point multiplications as for DCTs, DCTsc on $N \times N$ block. Moreover, our focus measure does not require $N^2$ floating point squaring operations to calculate $F^{D^2}$ (See expression (21) where $F^D$ represents the DCT transformed version of spatial domain signal $F$) as required in DCTs or DCTv. On the other hand, techniques like DCT+EOL, DCT+VOL and DCT+Corr_Eng require six non-diagonal matrix multiplication operations. Further, DCT+VOL and DCT+EOL require floating point element wise squaring operations and DCT+Corr_Eng requires floating point element-wise multiplications, divisions, square root operations. DCT+SVD requires singular value decomposition whose computational complexity is in the order of $N^2$. Thus, in comparison to the existing state-of-the-art techniques, our approach requires $N$ fixed point multiplications. Such reduction is a huge benefit in terms of energy consumption, and therefore, suits resource-constrained VSN.

**FIGURE 8.** Source images: The 'Temple' pair of images from the database given by [35] and multi-focus image fusion with different techniques.

Apart from the focus measure, the state-of-the-art fusion techniques involve consistency verification through a decision map. The refinement of the decision map is performed through a filtering operation, and the final decision map is a non-binary map for techniques like DCTsc, DCT+EOL+CV, DCT+VOL+CV and DCT+Corr_Eng+CV. On the other hand, our modified decision map (See Section V-B) computation is filtering operation, which requires integer multiplication and addition on the initial decision map.

The refined decision map computation requires a similar filtering operation on the sign information of the modified decision map. The last part of the process is fusion, where techniques like DCTsc, DCT+EOL+CV, DCT+VOL+CV, and DCT+Corr_Eng+CV either select a block from two images or average two blocks to generate the fused image. But, our Proposed+CV (See Section V-D), only requires to perform move operation based on the sign bit of the decision map elements. Therefore,

|  |  |  |  |
|---|---|---|---|
| (a) Source Image | (b) Source Image | (c) DCTv | (d) DCTvc |
| (e) DCTs | (f) DCTsc | (g) DCT+SVD | (h) DCT+SVD+CV |
| (i) DCT+EOL | (j) DCT+EOL+CV | (k) DCT+VOL | (l) DCT+VOL+CV |
| (m) DCT+Corr_Eng | (n) DCT+Corr_Eng+CV | (o) Proposed | (p) Proposed+CV |
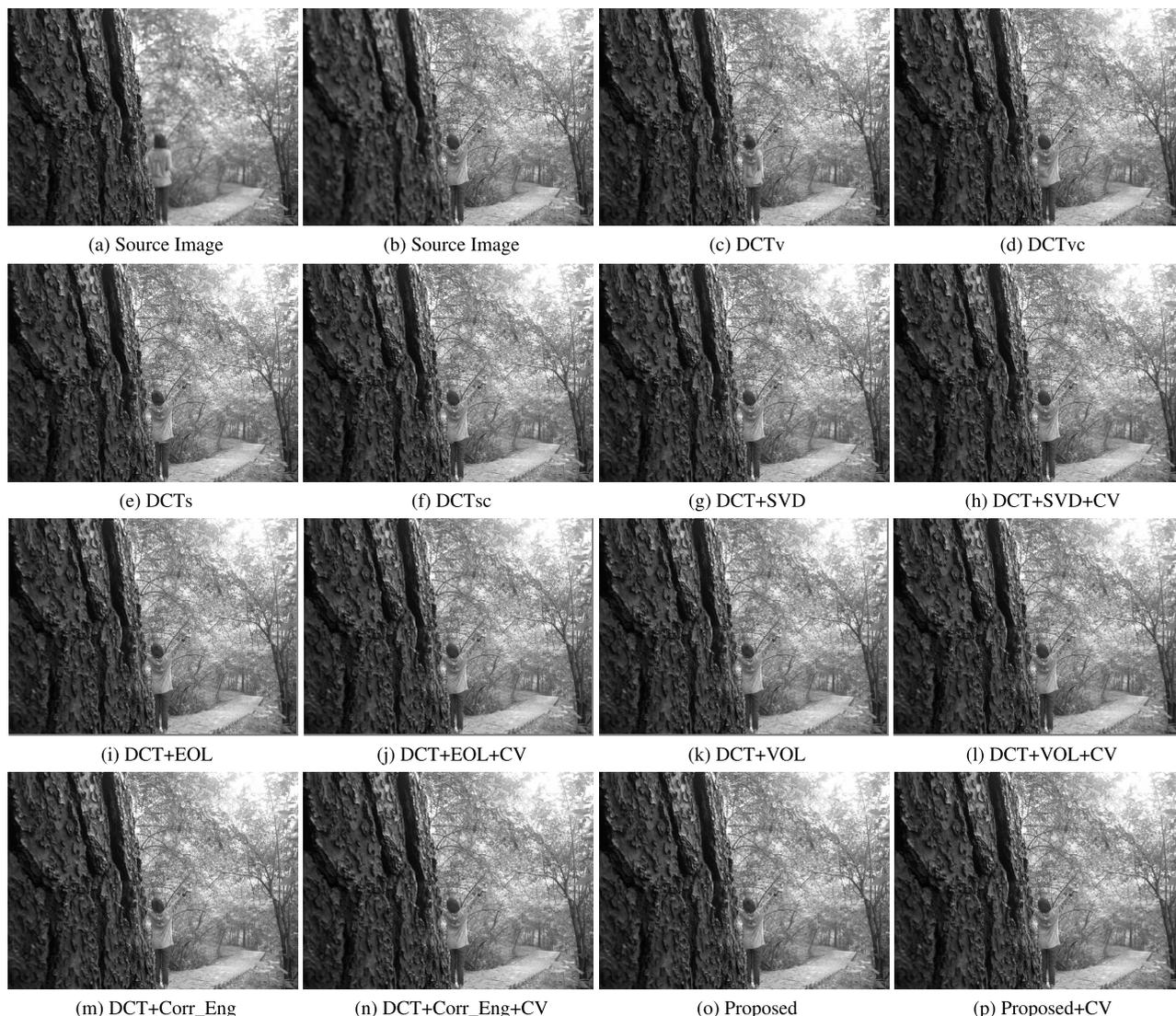
**FIGURE 9.** Source images: The 'Girl' pair of images from the database given by [35] and multi-focus image fusion with different techniques.

each module of our proposed approach is computationally simple and is very much hardware-friendly especially when compared with techniques like DCTs, DCTsc, DCTv, DCTvc, DCT+SVD, DCT+SVD+CV, DCT+EOL, DCT+EOL+CV, DCT+VOL, DCT+VOL+CV, DCT+ Corr_Eng, and DCT+Corr_Eng+CV.

To summarize our observations, quantitative and subjective evaluations show that our approach is as good as or better than state-of-the-art techniques. Computation time comparison shows the superiority of our approach. The hardware-friendly solution in DCT domain indicates that our proposed technique is best suited for resource-constrained VSN compared to other state-of-the-art techniques.

### D. COMPARISON WITH GENERIC MULTI-FOCUS IMAGE FUSION

VSN demands real-time performance with hardware-friendly implementation in the DCT domain. There is a plethora

of techniques [37]–[40], [54]–[56] proposed for generic multi-focus image fusion, which are not DCT based. These techniques are generally computationally intensive and require considerably more time to process. Few of the techniques also require graphical processing unit (GPU). Thus, most of these techniques are not suitable for VSN.

We perform no-reference based quantitative evaluation of a few latest generic multi-focus image fusion techniques that represent the the state-of-the-art such as MFCNN [37], MFGFDF [38], MFRW [39], MADCNN [40] and ECNN [41], which are presented in Tables 8 and 9, analogous to that in Tables 1 and 2, respectively. Consider the evaluations shown in Tables 8 and 9. From the tables we see that barring one existing technique for each database, no single technique performs the best with respect to more than one no-reference quality measure. Our approach gives the best $Q^{AB/F}$ performance for the database provided in

(a) Source Image     (b) Source Image     (c) DCTv     (d) DCTvc

(e) DCTs     (f) DCTsc     (g) DCT+SVD     (h) DCT+SVD+CV

(i) DCT+EOL     (j) DCT+EOL+CV     (k) DCT+VOL     (l) DCT+VOL+CV

(m) DCT+Corr_Eng     (n) DCT+Corr_Eng+CV     (o) Proposed     (p) Proposed+CV
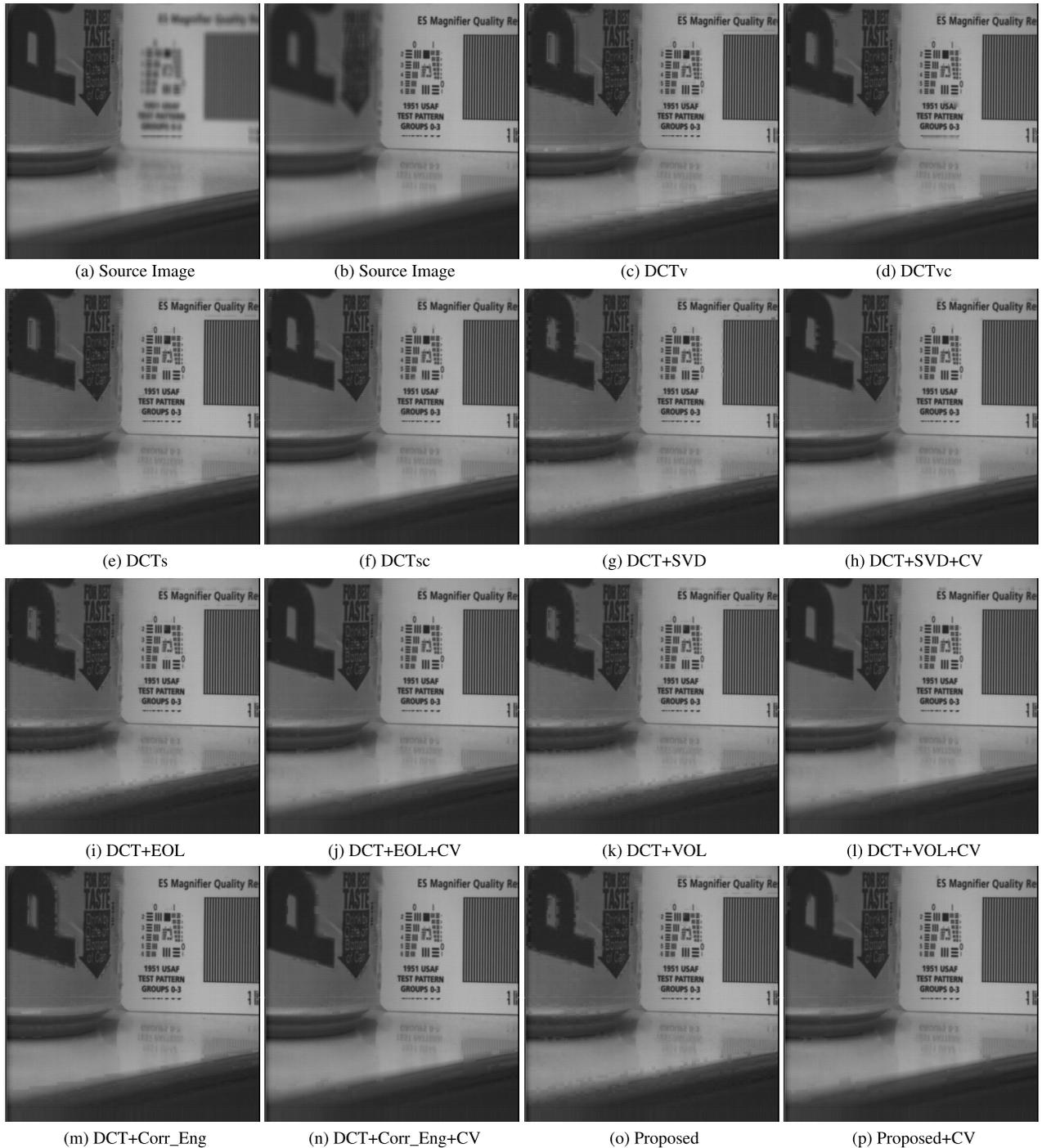
**FIGURE 10.** Source images: The 'Pepsi' pair of images from the database given by [35] and multi-focus image fusion with different techniques.

[35] and $Q^{FMI}$ performance for Lytro database. We also show quantitative evaluation on an individual image from Lytro in Table 10. We see similar performance (except MFRF performs better in two of two measures). Hence, we can infer that performance of our approach is comparable to the state-of-the-art in generic multi-focus image fusion, in spite of being fast, hardware-friendly and computationally simple,

and most importantly, suitable for VSN. We also perform subjective comparison with the generic techniques and present the results in Figs. 12, 13 and 14. The results shows that the all the generic multi-focus image fusion techniques perform similar. The results also show that our 'Proposed+CV' performs similar to the generic multi-focus image fusion techniques.
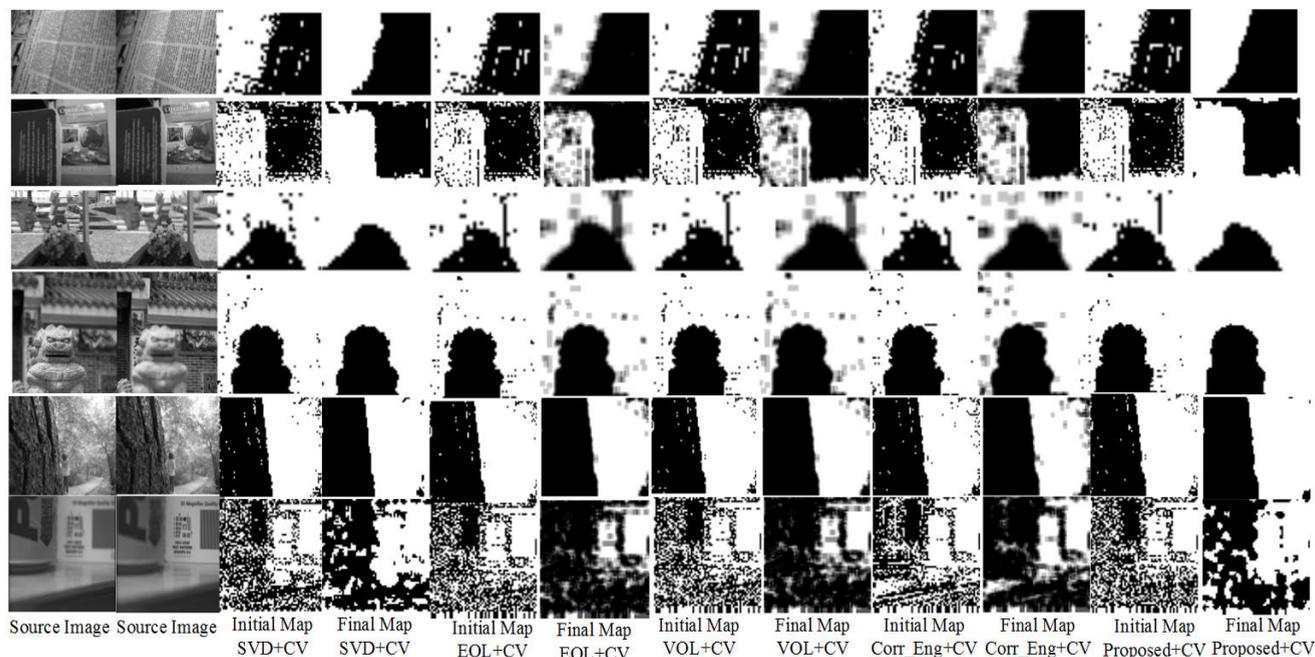
**FIGURE 11.** Initial and refined decision maps generated by different approaches.
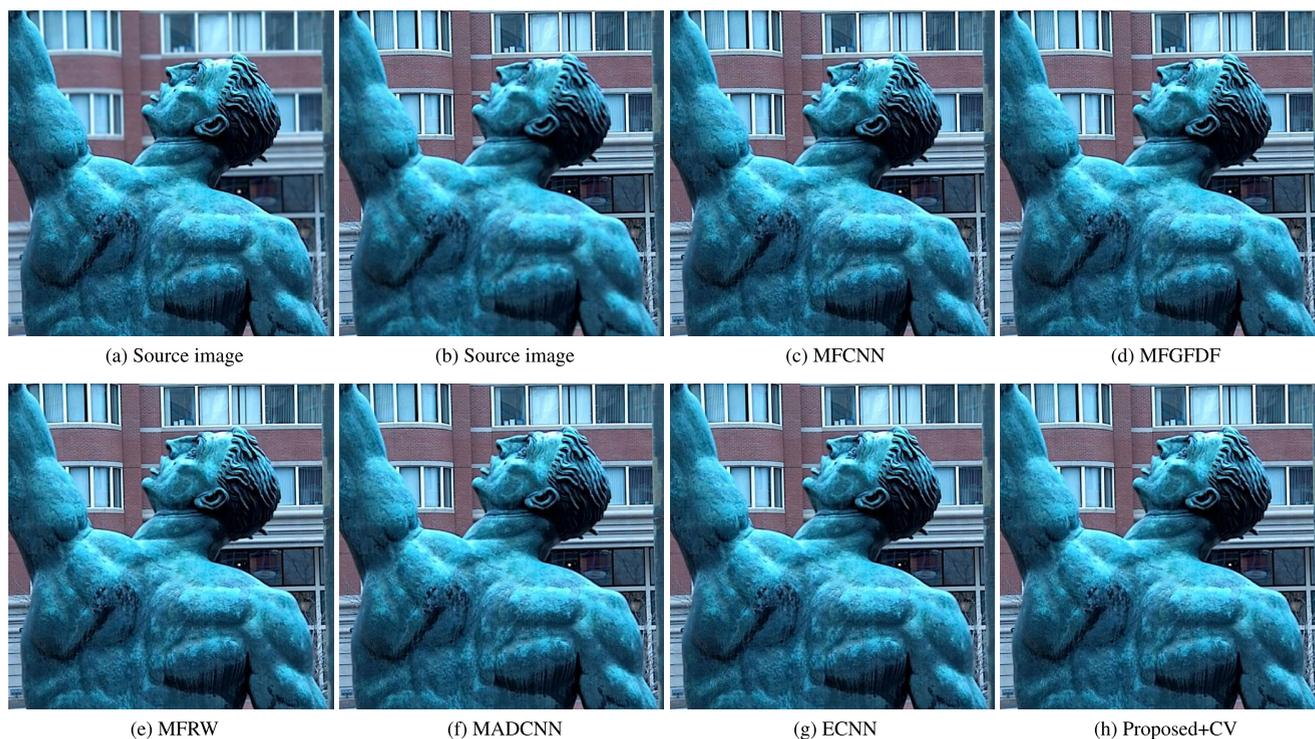


**FIGURE 12.** Source images: Lytro-08 and multi-focus image fusion with different techniques.

We compare computation time (in terms of CPU time) of our approach with that of generic multi-focus image fusion techniques such as MFCNN [37], MFGFDF [38] and MFRW [39], which are presented in Table 11. The techniques including ours are run on Matlab® platform in a system of Intel® Core(TM) i5-4590 CPU @ 3.30 GHz having 16 GB RAM. MADCNN and ECNN are not considered as their codes are not available for the said platform. The results show that time taken by the generic techniques is considerably high. It is also evident that

| (a) Source image | (b) Source image | (c) MFCNN | (d) MFGFDF |
| (e) MFRW | (f) MADCNN | (g) ECNN | (h) Proposed+CV |
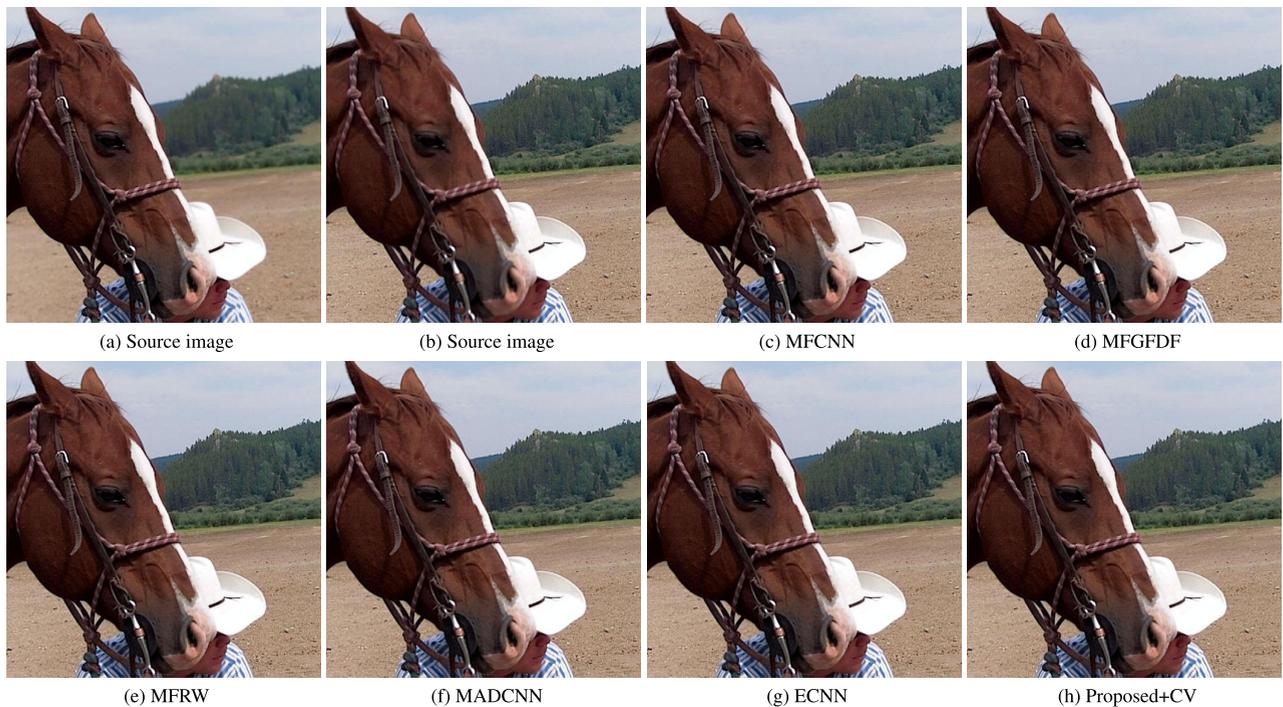
**FIGURE 13.** Source images: Lytro-15 and multi-focus image fusion with different techniques.



| (a) Source image | (b) Source image | (c) MFCNN | (d) MFGFDF |
| (e) MFRW | (f) MADCNN | (g) ECNN | (h) Proposed+CV |
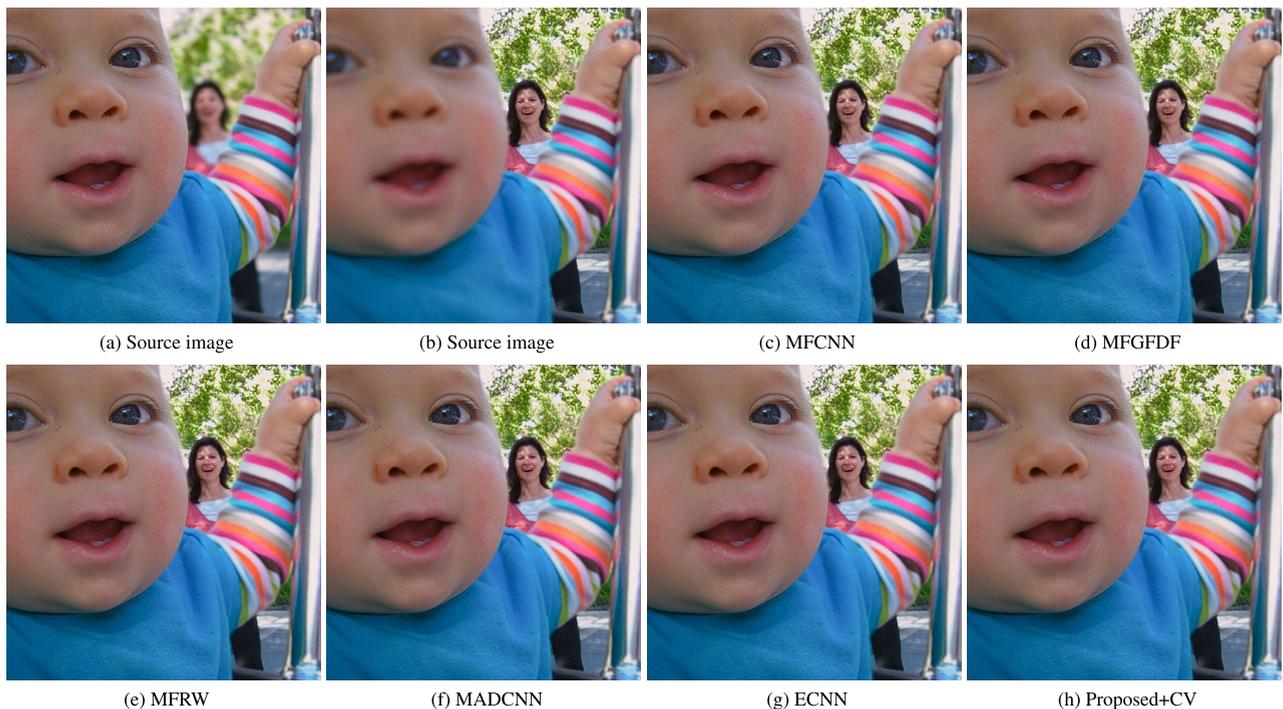
**FIGURE 14.** Source images: Lytro-16 and multi-focus image fusion with different techniques.

in the Matlab platform with the said specification our approach can fuse video frames with 256 × 256, 512 × 512 and 1024 × 1024 size at 62 fps, 16 fps and 4 fps, respectively.

It is clear that our approach takes very much less time with hardware-friendly solution in DCT domain for VSN and it produces results comparable to the state-of-the-art multi-focus fusion approaches like MFCNN [37],

**TABLE 8.** Performance comparison of multi-focus image fusion techniques for VSN based on no-reference quality measures (best result in bold). Values are the average measures on 18 pairs of multi-focus images (shown in Figure 3) from the database given by [35].

| Techniques | Measures | | | | | |
|---|---|---|---|---|---|---|
| | $Q^P$ | $Q^Y$ | $Q^{CB}$ | $Q^W$ | $Q^{AB/F}$ | $Q^{FMI}$ |
| MFCNN | 0.819679 | 0.984334 | 0.781784 | 0.897055 | 0.735391 | 0.625869 |
| MFGFDF | 0.820257 | 0.983249 | **0.781951** | 0.896023 | 0.736328 | 0.626295 |
| MFRW | 0.819737 | 0.982512 | 0.776225 | 0.896685 | 0.734327 | 0.613717 |
| MADCNN | 0.788231 | 0.964901 | 0.757605 | **0.900344** | 0.718517 | 0.585375 |
| ECNN | **0.808047** | **0.984951** | 0.779576 | 0.890699 | 0.73452 | **0.632515** |
| Proposed | 0.77092 | 0.95776 | 0.74546 | 0.891762 | 0.73067 | 0.612878 |
| Proposed+CV | 0.806552 | 0.981288 | 0.777909 | 0.892931 | **0.736382** | 0.631641 |

**TABLE 9.** Performance comparison of multi-focus image fusion techniques for VSN based on no-reference quality measures (best result in bold). Values are the average measures on 20 pairs of multi-focus images (Figure 4) from Lytro database [36].

| Techniques | Measures | | | | | |
|---|---|---|---|---|---|---|
| | $Q^P$ | $Q^Y$ | $Q^{CB}$ | $Q^W$ | $Q^{AB/F}$ | $Q^{FMI}$ |
| MFCNN | **0.850717** | 0.986451 | 0.808438 | **0.932093** | 0.761953 | 0.606032 |
| MFGFDF | 0.850672 | 0.987534 | **0.811361** | 0.93143 | 0.762126 | 0.607372 |
| MFRW | 0.850654 | **0.987709** | 0.808297 | 0.931609 | 0.761322 | 0.603095 |
| MADCNN | 0.847171 | 0.984951 | 0.804119 | 0.931083 | 0.761359 | 0.59662 |
| ECNN | 0.840507 | 0.986982 | 0.807342 | 0.928258 | 0.758841 | 0.606731 |
| Proposed | 0.82299 | 0.97251 | 0.78434 | 0.92848 | 0.75655 | 0.59596 |
| Proposed+CV | 0.84254 | 0.986366 | 0.808567 | 0.929021 | 0.76064 | **0.610169** |

**TABLE 10.** Performance comparison of multi-focus image fusion techniques on the image Lytro-16 based on no-reference quality measures (best result in bold).

| Techniques | Measures | | | | | |
|---|---|---|---|---|---|---|
| | $Q^P$ | $Q^Y$ | $Q^{CB}$ | $Q^W$ | $Q^{AB/F}$ | $Q^{FMI}$ |
| MFCNN | 0.861193 | 0.986132 | 0.759932 | 0.948236 | **0.789704** | 0.600379 |
| MFGFDF | 0.86108 | **0.98633** | 0.760812 | 0.947798 | 0.78928 | 0.601234 |
| MFRF | **0.861957** | 0.98677 | 0.760462 | **0.947901** | 0.789363 | 0.601507 |
| MADCNN | 0.855365 | 0.984685 | 0.758516 | 0.947581 | 0.788619 | 0.584367 |
| ECNN | 0.858248 | 0.986134 | **0.760969** | 0.946577 | 0.787357 | 0.602034 |
| Proposed | 0.838565 | 0.975401 | 0.746185 | 0.946264 | 0.782799 | 0.5832 |
| Proposed+CV | 0.855857 | 0.985964 | 0.760527 | 0.946314 | 0.78672 | **0.603663** |

**TABLE 11.** CPU time (in seconds) comparison with generic multi-focus image fusion.

| Techniques | Image Size | | |
|---|---|---|---|
| | $256 \times 256$ | $512 \times 512$ | $1024 \times 1024$ |
| MFCNN | 38.51 sec | 129.94 sec | 525.40 sec |
| MFGFDF | 0.47 sec | 1.26 sec | 4.45 sec |
| MFRW | 0.25 sec | 1.15 sec | 6.29 sec |
| Proposed+CV (for VSN) | 0.016 sec | 0.060 sec | 0.245 sec |

MFGFDF [38], MFRW [39], MADCNN [40] and ECNN [41].

## VII. CONCLUSION

The paper has shown that the DCT basis function and the block Laplacian operator are related in such a way that DCT operation on block Laplacian operator with proper boundary condition generates a diagonal matrix. The property is exploited to propose a novel focus measure, which can directly operate on the DCT coefficients to detect the focused region of an image. Such a solution is well suited for visual sensor network (VSN). Therefore, we propose an approach for multi-focus images or video frames in DCT based image or video coding framework for VSN.

The quantitative and qualitative evaluations show that our proposed approach outperforms all the techniques designed for VSN. Moreover, a simple and hardware-friendly fast approach of our approach is suitable for resource-constrained VSN. In addition, our approach performs similar to the state-of-the-art generic techniques for multi-focus image fusion.

## APPENDIX A

*To Prove*: If $A = XZ$, where $X$ and $Z$ are two anti-diagonal matrices of size $N \times N$, then $A$ is a diagonal matrix whose diagonal element $a_{i,i} = x_{i,N+1-i} \times z_{N+1-i,i}$

*Proof:* For anti-diagonal matrices $X$ and $Z$, an element of $x_{i,j}$ or $z_{i,j}$ can be defined as follows

$$x_{i,j} \ or \ z_{i,j} = \begin{cases} g_{i,j}, & j = N+1-i, \quad \forall g_{i,j} \in \{\mathbb{R}, \mathbb{C}\} \\ 0, & elsewhere \end{cases} \quad (27)$$

To avoid confusion, let us assume that $g_{i,j}$ is a nonzero entry. So, the above matrices have values at the $(i, N+1-i)$ index point. Now, for $A = XZ$, $a_{i,j} = \sum_{n=1}^{N} x_{i,n} \times z_{n,j}$. We know that $x_{i,j} = 0$ except at the index $(i, N+1-i)$. Therefore, $a_{i,j}$ can have a nonzero value $x_{i,N+1-i} \times z_{N+1-i,j}$. But $z_{i,j} = 0$ except at the index $(i, N+1-i)$. This implies that we only can have nonzero values *iff* $i = j$. Therefore, $a_{i,j} = 0$ except for $i = j$, which implies $A$ is a diagonal matrix having $a_{i,i} = x_{i,N+1-i} \times z_{N+1-i,i}$. It is implied that the proof is true for all values of $g_{i,j}$.

## APPENDIX B

*To Prove*: If $B = YZ$, where $Y$ is diagonal matrix and $Z$ is an anti-diagonal matrix of size $N \times N$, then $B$ is an anti-diagonal matrix, whose anti-diagonal element $b_{i,N+1-i} = y_{i,i} \times z_{i,N+1-i}$.

*Proof:* To avoid confusion, let us assume that diagonal term of the diagonal matrix $Y$ is nonzero. Thus, $y_{i,j} = 0$ except for $i = j$. Now, for $B = YZ$, any element $b_{i,j} = \sum_{n=1}^{N} y_{i,n} \times z_{n,j}$. As $y_{i,j} = 0$ except $i = j$, $b_{i,j}$ can have a nonzero value when $i = j$. On the other hand, $z_{i,j} = 0$ except at the index $(i, N+1-i)$. Therefore, $b_{i,j}$ will have a nonzero value only at index $(i, N+1-i)$. This implies $B$ is an anti-diagonal matrix having $b_{i,N+1-i} = y_{i,i} \times z_{i,N+1-i}$. It is implied that the proof is true for any values of diagonal terms.

## APPENDIX C

*To Prove*: If $C = XYZ$, where $X$ and $Z$ are two anti-diagonal matrices and $Y$ is a diagonal matrix of size $N \times N$, then $C$ is a diagonal matrix such that the element $c_{i,i} = x_{i,N+1-i} \times y_{N+1-i,N+1-i} \times z_{N+1-i,i}$

*Proof:* From the proof in Appendix B, $B = YZ$ is an anti-diagonal matrix with $b_{i,N+1-i} = y_{i,i} \times z_{i,N+1-i}$. Again, $X$ is an anti-diagonal matrix and from Appendix A, we can say that $C$ is a diagonal matrix, where $C = XB$ and $c_{i,i} = x_{i,N+1-i} \times b_{N+1-i,i}$. Plugging $b_{i,N+1-i} = y_{i,i} \times z_{i,N+1-i}$ in $c_{i,i} = x_{i,N+1-i} \times b_{N+1-i,i}$, we have, $c_{i,i} = x_{i,N+1-i} \times y_{N+1-i,N+1-i} \times z_{N+1-i,i}$.

## REFERENCES

[1] S. Soro and W. Heinzelman, "A survey of visual sensor networks," *Adv. Multimedia*, vol. 2009, pp. 1–21, May 2009.

[2] R. C. Luo and M. G. Kay, "Multisensor integration and fusion in intelligent systems," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, no. 5, pp. 901–931, Oct. 1989.

[3] B. Dieber, C. Micheloni, and B. Rinner, "Resource-aware coverage and task assignment in visual sensor networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 10, pp. 1424–1437, Oct. 2011.

[4] S. Motiian, F. Siyahjani, R. Almohsen, and G. Doretto, "Online human interaction detection and recognition with multiple cameras," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 3, pp. 649–663, Mar. 2017.

[5] B. Wang, Y. Hu, J. Gao, Y. Sun, and B. Yin, "Laplacian LRR on product Grassmann manifolds for human activity clustering in multicamera video surveillance," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 3, pp. 554–566, Mar. 2017.

[6] D. Drajic and N. Cvejic, "Adaptive fusion of multimodal surveillance image sequences in visual sensor networks," *IEEE Trans. Consum. Electron.*, vol. 53, no. 4, pp. 1456–1462, Nov. 2007.

[7] V. Maik, D. Cho, J. Shin, and J. Paik, "Regularized restoration using image fusion for digital auto-focusing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 10, pp. 1360–1369, Oct. 2007.

[8] A. Redondi, D. Buranapanichkit, M. Cesana, M. Tagliasacchi, and Y. Andreopoulos, "Energy consumption of visual sensor networks: Impact of spatio-temporal coverage," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 12, pp. 2117–2131, Dec. 2014.

[9] D. Costa, L. Guedes, F. Vasques, and P. Portugal, "Research trends in wireless visual sensor networks when exploiting prioritization," *Sensors*, vol. 15, no. 1, pp. 1760–1784, Jan. 2015.

[10] Y. A. V. Phamila and R. Amutha, "Discrete cosine transform based fusion of multi-focus images for visual sensor networks," *Signal Process.*, vol. 95, pp. 161–170, Feb. 2014.

[11] C. Li, J. Zou, H. Xiong, and C. W. Chen, "Joint coding/routing optimization for distributed video sources in wireless visual sensor networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 2, pp. 141–155, Feb. 2011.

[12] K. Pandremmenou, L. P. Kondi, and K. E. Parsopoulos, "Geometric bargaining approach for optimizing resource allocation in wireless visual sensor networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 8, pp. 1388–1401, Aug. 2013.

[13] G. K. Wallace, "The JPEG still picture compression standard," *Commun. ACM*, vol. 34, no. 4, pp. 30–44, Apr. 1991.

[14] I. E. Richardson, *Video Codec Design: Developing Image and Video Compression Systems*. Hoboken, NJ, USA: Wiley, 2002.

[15] P. K. Meher, S. Y. Park, B. K. Mohanty, K. S. Lim, and C. Yeo, "Efficient integer DCT architectures for HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 1, pp. 168–178, Jan. 2014.

[16] J. Tang, "A contrast based image fusion technique in the DCT domain," *Digit. Signal Process.*, vol. 14, no. 3, pp. 218–226, 2004.

[17] V. Naidu, "Discrete cosine transform based image fusion techniques," *J. Commun., Navigat. Signal Process.*, vol. 1, no. 1, pp. 35–45, 2012.

[18] M. B. A. Haghighat, A. Aghagolzadeh, and H. Seyedarabi, "Multi-focus image fusion for visual sensor networks in DCT domain," *Comput. Electr. Eng.*, vol. 37, no. 5, pp. 789–797, Sep. 2011.

[19] L. Cao, L. Jin, H. Tao, G. Li, Z. Zhuang, and Y. Zhang, "Multi-focus image fusion based on spatial frequency in discrete cosine transform domain," *IEEE Signal Process. Lett.*, vol. 22, no. 2, pp. 220–224, Feb. 2015.

[20] M. Amin-Naji and A. Aghagolzadeh, "Multi-focus image fusion in DCT domain using variance and energy of Laplacian and correlation coefficient for visual sensor networks," *J. AI Data Mining*, vol. 6, no. 2, pp. 233–250, 2018.

[21] E. Vakaimalar, K. Mala, and S. R. Babu, "Multifocus image fusion scheme based on discrete cosine transform and spatial frequency," *Multimedia Tools Appl.*, vol. 78, no. 13, pp. 17573–17587, Jul. 2019.

[22] M. Amin-Naji, P. Ranjbar-Noiey, and A. Aghagolzadeh, "Multi-focus image fusion using singular value decomposition in DCT domain," in *Proc. 10th Iranian Conf. Mach. Vis. Image Process. (MVIP)*, Nov. 2017, pp. 45–51.

[23] M. A. Naji and A. Aghagolzadeh, "Multi-focus image fusion in DCT domain based on correlation coefficient," in *Proc. 2nd Int. Conf. Knowl.-Based Eng. Innov. (KBEI)*, Nov. 2015, pp. 632–639.

[24] X. Wang, "Laplacian operator-based edge detectors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 5, pp. 886–890, May 2007.

[25] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 8, pp. 824–831, Aug. 1994.

[26] M. Subbarao and J.-K. Tyan, "The optimal focus measure for passive autofocusing and depth-from-focus [2598-10]," in *Proc. SPIE, Int. Soc. Opt. Eng.*, 1995, p. 89.

[27] M. Subbarao and J.-K. Tyan, "Selecting the optimal focus measure for autofocusing and depth-from-focus," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 8, pp. 864–870, Aug. 1998.

[28] W. Huang and Z. Jing, "Evaluation of focus measures in multi-focus image fusion," *Pattern Recognit. Lett.*, vol. 28, no. 4, pp. 493–500, Mar. 2007.

[29] G. Strang, "The discrete cosine transform," *SIAM Rev.*, vol. 41, no. 1, pp. 135–147, 1999.

[30] M. B. A. Haghighat, A. Aghagolzadeh, and H. Seyedarabi, "Real-time fusion of multi-focus images for visual sensor networks," in *Proc. 6th Iranian Conf. Mach. Vis. Image Process.*, Oct. 2010, pp. 1–6.

[31] V. Lecuire, L. Makkaoui, and J.-M. Moureaux, "Fast zonal DCT for energy conservation in wireless image sensor networks," *Electron. Lett.*, vol. 48, no. 2, pp. 125–127, Jan. 2012.

[32] D.-U. Lee, H. Kim, M. Rahimi, D. Estrin, and J. D. Villasenor, "Energy-efficient image compression for resource-constrained platforms," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 2100–2113, Sep. 2009.

[33] Z. He and M. Bystrom, "Improved conversion from DCT blocks to integer cosine transform blocks in H. 264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 6, p. 851, May 2008.

[34] H. Li, B. S. Manjunath, and S. K. Mitra, "Multisensor image fusion using the wavelet transform," *Graph. Models Image Process.*, vol. 57, no. 3, pp. 235–245, May 1995.

[35] Y. Zhang, X. Bai, and T. Wang, "Boundary finding based multi-focus image fusion through multi-scale morphological focus-measure," *Inf. Fusion*, vol. 35, pp. 81–101, May 2017.

[36] M. Nejati, S. Samavi, and S. Shirani, "Multi-focus image fusion using dictionary-based sparse representation," *Inf. Fusion*, vol. 25, pp. 72–84, Sep. 2015.

[37] Y. Liu, X. Chen, H. Peng, and Z. Wang, "Multi-focus image fusion with a deep convolutional neural network," *Inf. Fusion*, vol. 36, pp. 191–207, Jul. 2017.

[38] X. Qiu, M. Li, L. Zhang, and X. Yuan, "Guided filter-based multi-focus image fusion through focus region detection," *Signal Process., Image Commun.*, vol. 72, pp. 35–46, Mar. 2019.

[39] J. Ma, Z. Zhou, B. Wang, L. Miao, and H. Zong, "Multi-focus image fusion using boosted random walks-based algorithm with two-scale focus maps," *Neurocomputing*, vol. 335, pp. 9–20, Mar. 2019.

[40] R. Lai, Y. Li, J. Guan, and A. Xiong, "Multi-scale visual attention deep convolutional neural network for multi-focus image fusion," *IEEE Access*, vol. 7, pp. 114385–114399, 2019.

[41] M. Amin-Naji, A. Aghagolzadeh, and M. Ezoji, "Ensemble of CNN for multi-focus image fusion," *Inf. Fusion*, vol. 51, pp. 201–214, Nov. 2019.

[42] S. Pistonesi, J. Martinez, S. M. Ojeda, and R. Vallejos, "Structural similarity metrics for quality image fusion assessment: Algorithms," *Image Process. Line*, vol. 8, pp. 345–368, Oct. 2018.

[43] M. B. A. Haghighat, A. Aghagolzadeh, and H. Seyedarabi, "A non-reference image fusion metric based on mutual information of image features," *Comput. Electr. Eng.*, vol. 37, no. 5, pp. 744–756, Sep. 2011.

[44] M. Haghighat and M. A. Razian, "Fast-FMI: Non-reference image fusion metric," in *Proc. IEEE 8th Int. Conf. Appl. Inf. Commun. Technol. (AICT)*, Oct. 2014, pp. 1–3.

[45] C. S. Xydeas and V. Petrović, "Objective image fusion performance measure," *Electron. Lett.*, vol. 36, no. 4, pp. 308–309, 2000.

[46] G. Piella and H. Heijmans, "A new quality metric for image fusion," in *Proc. Int. Conf. Image Process.*, vol. 3, Sep. 2003, p. III-173.

[47] C. Yang, J.-Q. Zhang, X.-R. Wang, and X. Liu, "A novel similarity based quality metric for image fusion," *Inf. Fusion*, vol. 9, no. 2, pp. 156–160, Apr. 2008.

[48] Y. Chen and R. S. Blum, "A new automated quality assessment algorithm for image fusion," *Image Vis. Comput.*, vol. 27, no. 10, pp. 1421–1432, Sep. 2009.

[49] J. Zhao, R. Laganiere, and Z. Liu, "Performance assessment of combinative pixel-level image fusion based on an absolute feature measurement," *Int. J. Innov. Comput. Inf. Control*, vol. 3, no. 6, pp. 1433–1447, 2007.

[50] G. Qu, D. Zhang, and P. Yan, "Information measure for performance of image fusion," *Electron. Lett.*, vol. 38, no. 7, pp. 313–315, Mar. 2002.

[51] N. Cvejic, C. N. Canagarajah, and D. R. Bull, "Image fusion metric based on mutual information and Tsallis entropy," *Electron. Lett.*, vol. 42, no. 11, pp. 626–627, May 2006.

[52] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

[53] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[54] M. Amin-Naji, A. Aghagolzadeh, and M. Ezoji, "Fully convolutional networks for multi-focus image fusion," in *Proc. 9th Int. Symp. Telecommun. (IST)*, Dec. 2018, pp. 553–558.

[55] X. Guo, R. Nie, J. Cao, D. Zhou, and W. Qian, "Fully convolutional network-based multifocus image fusion," *Neural Comput.*, vol. 30, no. 7, pp. 1775–1800, Jul. 2018.

[56] M. Amin-Naji, A. Aghagolzadeh, and M. Ezoji, "CNNs hard voting for multi-focus image fusion," *J. Ambient Intell. Hum. Comput.*, vol. 11, no. 4, pp. 1749–17691, 2020.

National University of Singapore and the Center for Soft Computing Research, Indian Statistical Institute. He currently heads the Vision, Image, and Perception Research Group and the ArtEye Laboratory in his department, which are funded by multiple agencies of Government of India and prominent industries in India. His current research interests include vision, image, and video processing, uncertainty handling, eye movement analysis, machine vision, and deep learning. He has authored/coauthored more than 50 research articles in high-impact journals and conferences including those published/sponsored by IEEE. He is on the editorial board of *IET Image Processing*, *Circuits, Systems, and Signal Processing* (Springer), and the IEEE *MMTC Communications-Review*. He has received the Young Scientist Award from the Institution of Engineers, India, the Qualcomm Innovation Fellowship, the ERCIM Alain Bensoussan Fellowship, the Ministry of Manpower (Singapore) Research Fellowship, and a couple of best paper awards from IET.

**SOBHAN KANTI DHARA** (Graduate Student Member, IEEE) received the B.Tech. degree in electronics and communication engineering from the Institute of Engineering and Management, West Bengal University of Technology, India, in 2010, and the M.Tech. degree in signal and image processing from the National Institute of Technology Rourkela, India, in 2015. He is currently pursuing the Ph.D. degree in electronics and electrical communication engineering from IIT Kharagpur, India. He was a Visiting Researcher with Concordia University, Canada, in 2018. His research interests include multi-image fusion, image enhancement, super resolution, uncertainty handling, quantum logic, and so on. He is a Reviewer of the reputed IEEE, IET, and Springer journals and conferences. He was a recipient of the IEEE Richard E. Merwin Award cum Scholarship, from 2019 to 2020, the Shastri Research Student Fellowship (Doctoral), from 2017 to 2018, the Institute Gold Medal for Best M.Tech. student, in 2015, the Institute Highest DGPA in 2015, the Academic Excellence Award in 2014, and a couple of travel grants and scholarships. Recently, he has been selected as the IEEE Student Ambassador. He has served in a number of administrative positions in his department and IEEE IIT Kharagpur student branch. Earlier, he was a member of the Institute Senate at National Institute of Technology Rourkela, from 2014 to 2015.

**DEBASHIS SEN** (Senior Member, IEEE) received the Ph.D. degree in image processing from Jadavpur University, Kolkata, India, and the M.A.Sc. degree in electrical engineering from Concordia University, Montreal, Canada. He is currently an Assistant Professor with the Department of Electronics and Electrical Communication Engineering and the Faculty Member with the Centre of Excellence in Advanced Manufacturing Technology, IIT Kharagpur. He was a Postdoctoral Researcher with the Multimedia Analysis and Synthesis Laboratory,

**M. N. S. SWAMY** (Life Fellow, IEEE) received the B.Sc. degree (Hons.) in mathematics from Mysore University, India, the Diploma degree in electrical communication engineering from the Indian Institute of Science, Bengaluru, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Saskatchewan, Saskatoon, Canada. He was conferred the title of Honorary Professor at National Chiao Tong University, Taiwan, in 2009. He is currently a Research Professor with the Department of Electrical and Computer Engineering, Concordia University, Montreal, QC, Canada, where he served as the Founding Chair of the Department of Electrical Engineering, from 1970 to 1977, the Dean of Engineering and Computer Science, Concordia University, from 1977 to 1993. During that time, he developed the Faculty into a research-oriented one from what was primarily an undergraduate Faculty. Since July 2001, he holds the Concordia Chair (Tier I) in signal processing. He has also taught in the Electrical Engineering Department, Technical University of Nova Scotia, Halifax, and the University of Calgary, Calgary, and the Department of Mathematics, University of Saskatchewan. He has published extensively in the areas of number theory, circuits, and systems and signal processing. He holds five patents. He is the coauthor of nine books and several book chapters. He was a Founding Member of Micronet from its inception, in 1990 as a Canadian Network of the Centers of Excellence until its expiration, in 2004, and also its coordinator for Concordia University. He is a Fellow of the Institute of Electrical Engineers, U.K., the Engineering Institute of Canada, the Institution of Engineers, India, and the Institution of Electronic and Telecommunication Engineers, India. He was inducted, in 2009 to the Provosts Circle of Distinction for career achievements. He was a recipient of many IEEE-CAS Society awards, including the Education Award, in 2000, Golden Jubilee Medal, in 2000, and the 1986 Guillemin-Cauer Best Paper Award. He has served the IEEE in various capacities, such as the President-Elect, in 2003, the President, in 2004, the Past-President in 2005, the Vice President (Publications), from 2001 to 2002, the Vice-President, in 1976, the Editor-in-Chief of the IEEE Transactions on Circuits and Systems—I, from June 1999 to December 2001, an Associate Editor of the IEEE Transactions on Circuits and Systems, from June 1985 to May 1987, the Program Chair for the 1973 IEEE CAS Symposium, a General Chair for the 1984 IEEE CAS Symposium, the Vice-Chair for the 1999 IEEE Circuits and Systems (CAS) Symposium, and a member of the Board of Governors of the CAS Society. He has been the Editor-in-Chief of the journal *Circuits, Systems, and Signal Processing* (CSSP), since 1999. Recently CSSP has instituted a Best Paper Award in his name.

● ● ●