Local Riesz Pyramid for Faster Phase-Based Video Magnification

Shoichiro TAKEDA<sup>a</sup>, Megumi ISOGAI<sup>b</sup>, Shinya SHIMIZU<sup>c</sup>, Nonmembers, and Hideaki KIMATA<sup>d</sup>, Senior Member

SUMMARY Phase-based video magnification methods can magnify and reveal subtle motion changes invisible to the naked eye. In these methods, each image frame in a video is decomposed into an image pyramid, and subtle motion changes are then detected as local phase changes with arbitrary orientations at each pixel and each pyramid level. One problem with this process is a long computational time to calculate the local phase changes, which makes high-speed processing of video magnification difficult. Recently, a decomposition technique called the Riesz pyramid has been proposed that detects only local phase changes in the dominant orientation. This technique can remove the arbitrariness of orientations and lower the over-completeness, thus achieving high-speed processing. However, as the resolution of input video increases, a large amount of data must be processed, requiring a long computational time. In this paper, we focus on the correlation of local phase changes between adjacent pyramid levels and present a novel decomposition technique called the local Riesz pyramid that enables faster phase-based video magnification by automatically processing the minimum number of sufficient local image areas at several pyramid levels. Through this minimum pyramid processing, our proposed phase-based video magnification method using the local Riesz pyramid achieves good magnification results within a short computational time.

key words: video magnification, Riesz pyramid, phase, video synthesis, video analysis

1. Introduction

Many attractive physical phenomena are often hidden within a very small world. For example, a baby’s chest slightly moving up and down while he/she is breathing, a car’s engines making periodical revolutions to supply motive power, and a drum membrane vibrating to produce wonderful sounds. However, these attractive subtle motion changes are too small to see with the naked eye.

To visualize such subtle motion changes in a video, phase-based video magnification methods have recently attracted much attention as a way of visualizing these changes using a form of magnification<sup>1</sup>–[6]. These methods are based on the analysis of local phase changes that correspond to local motion changes independent from color changes[7]. They first decompose each image frame in a video into a multi-scale complex-steerable pyramid with various scales and orientations to obtain local phase changes, and these changes are then temporally filtered to detect subtle phase changes to be magnified. The phase-based video magnification achieves good results for magnifying the subtle sway of a bridge, the breathing of a baby, and skin deformations due to blood circulation. Moreover, with recently developed spatio-temporal filtering techniques[2]–[6], current phase-based video magnification methods can magnify and visualize only subtle motion changes related to attractive physical phenomena even if large motions of objects and noise exist in a video. However, the methods require a long computational time to analyze local phase changes, in proportion to the video resolution and the number of time frames.

To counter this problem, Wadhwa et al. [8] have proposed a decomposition technique called the Riesz pyramid as an improvement to the complex-steerable pyramid used in previous works [1]–[6]. The Riesz pyramid removes the arbitrariness of orientations of local phase changes and only detects them with respect to the dominant orientation at every pixel and time frame. This lowers the over-completeness of the complex-steerable pyramid and enables us to decompose each image frame in a video faster and magnify the subtle motion changes quickly. However, since the entire image frame must be processed, decomposition using the Riesz pyramid requires a long computational time in proportion to the video resolution and the number of time frames.

In this paper, we present a novel decomposition technique called the local Riesz pyramid for faster phase-based video magnification. We noticed that we only have to process the minimum number of sufficient local image areas at pyramid level $n$ related to the strongly magnified image areas at the above pyramid level $n + 1$ because local phase changes have a correlation between adjacent pyramid levels as reported by [9], [10]. In our proposed local Riesz pyramid, we calculate and magnify local phase changes at pyramid level $n + 1$ in the same way as [8] and then detect strongly magnified image areas by using Otsu’s thresholding method [11]. After this detection, we propagate the strongly magnified local image areas from pyramid level $n + 1$ to the below pyramid level $n$. At pyramid level $n$, we calculate and magnify local phase changes only in these propagated local image areas. Through this minimum pyramid processing, our proposed method using the local Riesz pyramid achieves good magnification results equivalent to [8] within a short computational time.

The contributions of this paper are as follows. 1) We
propose a novel decomposition technique called local Riesz pyramid that processes only the minimum number of sufficient local image areas at several pyramid levels to perform faster phase-based video magnification. 2) We analyze the algorithmic time complexity of our proposed method against the conventional Riesz pyramid. 3) We show the qualitative and quantitative effects that our method has on video magnification in real videos and a simulation video. 4) We discuss practical insights, limitations, and future work related to our method.

Note that this paper is a previously unpublished research paper but an extended version of the technical report presented in [12] in terms of refining equations, adding a number of figures, and including deep discussions for better understanding of our proposed method.

2. Related Works

2.1 Optical Flow-Based Approach

The first proposed video magnification method is based on optical flow [13]. It captures local motion changes in a video by utilizing optical flow. Through spatial registration of background motions in advance, it can produce a synthesized video output where subtle motion changes are magnified. However, optical flow is computationally expensive and has itself been researched as an unsolved problem [14]–[16].

2.2 Learning-Based Approach

Recently, Oh et al. have been proposed a learning-based video magnification method that uses a deep neural network [17]. The network takes two image frames as input with a magnification factor and outputs a warped frame where subtle motion changes are magnified. However, this approach often misses subtle motion changes due to a strong dependence on a training dataset and the prediction of the network to produce warped frames requires a long computational time.

2.3 Phase-Based Approach

Unlike the above approach, the current video magnification methods are mainly based on a phase-based approach [1]–[6]. Phase-based video magnification methods analyze local phase changes, which correspond to local motion changes independent from color changes [7], at a fixed pixel position over time without object tracking. These methods first decompose each image frame in a video into a multi-scale complex-steerable pyramid with various scales and orientations to obtain local phase changes, and these changes are then temporally filtered to detect subtle phase changes to be magnified. Although the phase changes are often contaminated by the effect of large motions and noise, with recently developed spatio-temporal filtering techniques [2]–[6], current phase-based video magnification methods can detect only subtle phase changes related to attractive physical phenomena even if the large motions of objects and noise exist in a video. After this processing, the subtle phase changes are amplified at each pyramid level, and the amplified complex-steerable pyramid is then reconstructed to output each magnified image frame in a video. However, since the complex-steerable pyramid uses many filter banks to decompose each image frame with various scales and orientations, it has over-complete representations and requires a long computational time, especially in proportion to the video resolution and the number of image frames.

To reduce the over-completeness, Wadhwa et al. [8] have proposed a decomposition technique called the Riesz pyramid as an improvement to the complex-steerable pyramid used in previous works [1]–[6]. The Riesz pyramid removes the arbitrariness of the orientations of local phase changes and only detects them with respect to the dominant orientation at every pixel and time frame. This enables us to decompose each image frame in a video faster because of the lower over-completeness of orientations and to produce a video magnification result quickly. However, since the entire image frame must be processed, the computational time for decomposition using a Riesz pyramid is long in proportion to the video resolution and the number of image frames.

To potentially overcome this problem, we considered the layer-based approach [2], [18], which specifies local image areas in advance by user selection, to be useful. This approach has originally been focused on reducing background noise by only magnifying local image areas, but this also enables us to save computational resources used for magnification. However, since it is difficult to know where subtle changes are in advance, a trial and error approach to searching for the appropriate areas to magnify is required, for which the computational time is long. Thus, layer-based methods are encouraging but insufficient to our goal.

In contrast, with our proposed local Riesz pyramid, we can automatically process the minimum number of sufficient local image areas at several pyramid levels, which are detected by utilizing the correlation between adjacent pyramid levels [9], [10] with Otsu’s thresholding method [11].


Here, we explain the conventional phase-based video magnification method based on the Riesz pyramid proposed by Wadhwa et al. [8].

Given a normalized image frame matrix \( I^0 = \left( \frac{I^0}{y} \right) \in \mathbb{R}^{H 	imes W} \) in one arbitrary color channel (we used the Y color channel in YIQ color space in this paper) at a time frame \( t = 1, \ldots, T \) with a discrete pixel position \((y, x)\), an image height \(H^0\), and an image width \(W^0\) in a video, the image frame is first decomposed into a non-oriented sub-bands pyramid, e.g. a Laplacian pyramid in this paper, as follows.

\[
I^n_i = \text{down} \left( \frac{1}{4} \right) \left[ I^{n-1}_i \right] \quad \text{if} \quad n > 0,
\]
In this paper, and processes respectively, by using two-dimensional Gaussian band image Riesz pyramid consisting of a set of three images: a sub-band image $L_n^x$ and two Riesz-transformed images $R_n^1$, $R_n^2$.

$$L_n^x = \begin{cases} I_n^x & \text{if } n = N \\ I_n^x - \text{down}_n[I_n^{x+1}] & \text{otherwise (}n < N\)). \end{cases}$$

where $n = 0, \ldots, N$ indicates each level of the pyramid, down$_n[.]$ and up$_n[.]$ perform down- and up-sampling processes respectively, by using two-dimensional Gaussian convolution with a pyramid scaling factor $\lambda$ (we used $\lambda = 2$ in this paper), and $L_n^x = (p_{y,xt}) \in \mathbb{R}^{H \times W}$ is a sub-band image with an image height $H$ and an image width $W$ at pyramid level $n$.

Next, the Riesz transform, which generalizes a one-dimensional Hilbert transform into a multi-dimensional one [19], is applied to the sub-band image $L_n^x$ as follows.

$$L_n^x \xrightarrow{\mathcal{FF}} G_n^x = (\pi_{\omega_x,\omega_y,\theta} \cdot i\omega_x / |\omega|) \in \mathbb{C}^{H \times W},$$

$$R_n^1 \xrightarrow{\mathcal{FF}^{-1}} G_n^1 = (\pi_{\omega_x,\omega_y,\theta} \cdot -i\omega_y / |\omega|) \in \mathbb{C}^{H \times W},$$

$$R_n^2 \xrightarrow{\mathcal{FF}^{-1}} G_n^2 = (\pi_{\omega_x,\omega_y,\theta} \cdot -i\omega_x / |\omega|) \in \mathbb{C}^{H \times W},$$

where $R_n^1 = (r_n^1) \in \mathbb{R}^{H \times W}$ and $R_n^2 = (r_n^2) \in \mathbb{R}^{H \times W}$ are the Riesz-transformed images, $\mathcal{FF}: \mathbb{R}^{H \times W} \rightarrow \mathbb{C}^{H \times W}$ is the 2D Fourier transform, $\mathcal{FF}^{-1}: \mathbb{C}^{H \times W} \rightarrow \mathbb{R}^{H \times W}$ is the inverse 2D Fourier transform, and $\omega = (\omega_x, \omega_y)$ represents the $x$ and $y$ axis spatial frequency in the Fourier domain, respectively. Through the Riesz transform, we can build a Riesz pyramid consisting of a set of three images: a sub-band image $L_n^x$ and two Riesz-transformed images $R_n^1$, $R_n^2$ (as shown in Fig. 1).

In the paper reported by Wadhawa et al. [8], a long computational time of the 2D Fourier transform $\mathcal{FF}$ and the inverse 2D Fourier transform $\mathcal{FF}^{-1}$ was pointed out. Furthermore, each sub-band image in a Laplacian pyramid is a critically sampled spatially bandpassed signal with most of the sub-band’s energy concentrated in a frequency band around $|\omega| = \frac{\pi}{2}$. These are why Wadhawa et al. approximated the Riesz transform by convolution in the spatial domain with three tap finite difference filters $d_1^T = [0.5, 0, -0.5]^T$ and $d_2 = [0.5, 0, -0.5]^T$ as follows.

$$R_n^1 \approx d_1^T \otimes L_n^x = \begin{bmatrix} d_1^T \otimes I_n^{x-1}, \ldots, d_1^T \otimes I_n^y, \ldots, d_1^T \otimes I_n^{y-1} \end{bmatrix}^T,$$

$$R_n^2 \approx d_2 \otimes L_n^x = \begin{bmatrix} d_2 \otimes I_n^{x-1}, \ldots, d_2 \otimes I_n^y, \ldots, d_2 \otimes I_n^{y-1} \end{bmatrix}^T,$$

where $\otimes$ is a convolution operator, $I_n^y = [p_{y,xt}]$ is the $y$-th row vector of $L_n^x$, and $I_n^x = [p_{y,xt}]$ is the $x$-th column vector of $L_n^x$. These spatial filters ($d_1^T$ and $d_2$) have frequency response as

$$-i \sin(\omega_x) \approx -i \frac{\omega_x}{|\omega_x|}$$

and

$$-i \sin(\omega_y) \approx -i \frac{\omega_y}{|\omega_y|},$$

respectively, when $\omega_x, \omega_y \approx \frac{\pi}{2}$. Note that this convolution does not have the precise staircase filter response that the original Riesz pyramid has due to the approximation of $|\omega_x| \approx |\omega|$ and $|\omega_y| \approx |\omega|$.

Then, Eqs. (2) and (4), which approximates Eq. (3), have relations at each pixel position $(y, x)$ and the time frame $t$ as

$$p_{y,xt}^n = d_{y,xt}^n \cdot \cos(\theta_{y,xt}^n),$$

$$p_{y,xt}^1 = d_{y,xt}^1 \cdot \sin(\theta_{y,xt}^n) \cdot \cos(\theta_{y,xt}^n),$$

$$p_{y,xt}^2 = d_{y,xt}^2 \cdot \sin(\theta_{y,xt}^n) \cdot \sin(\theta_{y,xt}^n),$$

where $d_{y,xt}^n$ is the amplitude change, $\theta_{y,xt}^n \in (-\pi, \pi)$ is the local phase change, and $\theta_{y,xt}^n$ is the dominant orientation in which the local phase change occurs.

Note that with the complex-steerable pyramid used in previous works [1]–[6], $\theta$ has to be fixed by a user in advance, e.g. eight orientations $\theta = \{0, \frac{\pi}{8}, \ldots, \frac{7\pi}{8}\}$, and $\theta_{y,xt}^n$ is parameterized by $\theta_{y,xt}^n$. Thus, the complex-steerable pyramid is over-complete with respect to $\theta$ and requires that the phase changes are processed in proportion to the number of orientations. In contrast, the Riesz pyramid removes the arbitrariness of the orientations and only detects them with respect to the dominant orientation $\theta_{y,xt}^n$ at each pixel, time frame, and pyramid level. Therefore, the Riesz pyramid lowers the over-completeness and achieves a faster pyramid-decomposition process.

From Eq. (6), the phase change $\theta_{y,xt}^n$ is obtained as

$$\theta_{y,xt}^n = \tan^{-1}\left(\frac{\sqrt{(p_{y,xt}^1)^2 + (p_{y,xt}^2)^2}}{p_{y,xt}^n}\right).$$

This phase change calculation is performed at each position $(y, x)$, each time frame $t$, and each pyramid level $n$.

To detect subtle phase changes $c_{y,xt}^n = [c_{y,xt}^1, \ldots, c_{y,xt}^n]$ with a desired time frequency range $f_1 \leq f \leq f_2$ in the Fourier domain, the temporal filter $h = [h_1, \ldots, h_T]$, e.g., the ideal bandpass filter [1], Butterworth bandpass filter [8], and temporal acceleration filter [4], is convolved to
is obtained with Eq. (6), and the magnified image frame is added to the original one. In this paper, we adopted the ideal bandpass filter in the same way as [1] because it was easy to implement and sufficient for magnifying subtle motion changes in the videos used in this paper. The ideal bandpass filter $h$ can be designed via the Fourier domain as

$$h \leftarrow h = \left[ h_1, \ldots, h_f, \ldots, h_F \right].$$

(8)

where $\mathcal{F}^{-1} : \mathbb{C}^F \rightarrow \mathbb{R}^T$ is the 1D inverse Fourier transform. Then, the ideal bandpass filter $h$ is convolved to $\phi_{yx}$ as

$$c_{yx}^n = h \otimes \phi_{yx}^n. \quad (9)$$

After that, the subtle phase change $c_{yx}^n$ multiplied by the magnification factor $\alpha$ is added to the original one $\phi_{yx}^n$ to obtain the magnified phase change $\hat{\phi}_{yx}^n$:

$$\hat{\phi}_{yx}^n = \phi_{yx}^n + \alpha \cdot c_{yx}^n. \quad (10)$$

Finally, the magnified sub-band image $\hat{I}_n^L = (\hat{I}_{yx}^n) \in \mathbb{R}^{h \times w}$ is obtained with Eq. (6), and the magnified image frame $\hat{I}_n^L = (\hat{I}_{yx}^n) \in \mathbb{R}^{h \times w}$ at pyramid level $n$, where subtle motion changes are only magnified, is then obtained by following the reverse procedure of Eq. (2) as

$$\hat{I}_n^L = \left\{ \begin{array}{ll} I_{yx,n}^L & \text{if } n = N \\ \hat{I}_n^L + u_p[I_{n+1}^L] & \text{otherwise (} n < N \end{array} \right. \quad (11)$$

$$\hat{I}_n^L = \left\{ \begin{array}{ll} I_{yx,n}^L & \text{if } n = N \\ \hat{I}_n^L + u_p[I_{n+1}^L] & \text{otherwise (} n < N \end{array} \right. \quad (12)$$

Through performing Eqs. (11) and (12) sequentially, the original resolution image frame $\tilde{I}_n^L$ is output (black line flow in Fig. 2). For details, see the paper reported by Wadhwa et al. [8].

4. Proposed Method: Local Riesz Pyramid for Faster Phase-Based Video Magnification

We present a novel faster phase-based video magnification method with the proposed local Riesz pyramid that processes the minimum and sufficient local image areas, compared with the original Riesz pyramid [8].

As mentioned above, the Riesz pyramid removes the arbitrariness of orientations $\theta$ and achieves faster phase-based video magnification compared with the complex- steerable pyramid. However, since the entire image frame position $(y, x)$ must be processed, the Riesz pyramid requires a long computational time in proportion to the video resolution and the number of image frames.

To overcome this limitation, we focused on the correlation of local phase changes between the adjacent pyramid levels (Fig. 3) as reported by [9], [10]. Figure 3 shows that, if we observed local phase changes in the same image areas between adjacent pyramid levels, the phase changes behave similarly as $\phi_{yx}^n = \lambda \cdot \phi_{yx}^{n+1}$ with the pyramid scaling parameter $\lambda$. Note that a large error will occur in this correlation if the pyramid level is too far away. From this finding, we noticed that we only have to process local image areas at pyramid level $n$ related to the strongly magnified image...
areas at the above pyramid level \( n + 1 \) for magnifying subtle motion changes. Thus, we propose the local Riesz pyramid, which automatically processes the minimum number of sufficient local image areas to quickly obtain the magnification results (black and red line flow in Fig. 2).

In the proposed local Riesz pyramid, we decomposed \( L_0 \) into a non-oriented sub-band pyramid in the same way as the conventional Riesz method [8]. Then, we processed a set of adjacent pyramid levels where the correlation of phase changes strongly exist: odd- \((n = 2k+1)\) and even-numbered \((n = 2k)\) pyramid levels with \(k = 0, \ldots, \frac{N}{2} - 1\). Note that we assume that \( N \) is an even-number in this paper. In this process, we first applied Eqs. (3)–(11) to the sub-band images \( L_{2k+1} \) at odd-numbered pyramid level \( n = 2k + 1 \). Then, we obtained the magnified sub-band image \( \tilde{L}_{2k+1}^2 = \hat{L}_{2k+1} \).

After that, we calculated a binary image \( B_{2k+1} = (b_{2k+1}^x)^y \in \{0, 1\}^{2^{(2k+1)} \times W^{2k+1}} \) with two steps: (i) summing the difference between the original sub-band images \( L_{2k+1}^2 \) and the magnified sub-band image \( \tilde{L}_{2k+1}^2 \) over all time frames \( t = 1, \ldots, T \) at each position \((y, x)\), and (ii) using Otsu’s thresholding method [11] as follows.

\[
\delta_{y,x}^{2k+1} = \sum_{t=1}^{T} (\hat{L}_{2k+1}^{y,xt} - \tilde{L}_{2k+1}^{y,xt})^2, \tag{13}
\]

\[
b_{2k+1}^x = \begin{cases} 1 & \text{if } \delta_{y,x}^{2k+1} > \epsilon \\ 0 & \text{otherwise} \end{cases}, \tag{14}
\]

where \( \epsilon \) is a threshold calculated by using Otsu’s thresholding method [11] in an implementation of OpenCV [20] to divide the discrete pixel positions \((y, x)\) at pyramid level \(2k + 1\) into one that is strongly magnified (\(b_{2k+1}^x = 1\)) and one that is not (\(b_{2k+1}^x = 0\)). Note that considering all time frames \( t = 1, \ldots, T \) in Eq. (13) is for focusing on cyclic subtle changes from the beginning to the end of an input video, rather than strong changes in a short period of time.

After that, we defined a set of discrete pixel positions \( \mathcal{P}_n \) at pyramid level \( n \) and divided the positions into \( U \times V \) subsets like a grid as

\[
\mathcal{P}_n = \left\{(y, x) \mid 1 \leq y \leq H^n, 1 \leq x \leq W^n \right\} = \{P_{11}^n, \ldots, P_{UV}^n\}, \tag{15}
\]

where \( P_{uv}^n \) is defined as

\[
P_{uv}^n = \left\{(y, x) \mid 1 + \frac{u - 1}{U} H^n \leq y \leq \frac{u}{U} H^n, 1 + \frac{v - 1}{V} W^n \leq x \leq \frac{v}{V} W^n \right\}. \tag{16}
\]

Note that \( P_{uv}^n \) should have \( 3 \times 3 \) or more pixel positions to ensure a minimum grid size.

Then, we collected subsets at pyramid level \( 2k \) (red dot image areas in Fig. 2) that correspond to the strongly magnified pixel positions at pyramid level \( 2k + 1 \) detected by Eq. (14) as

\[
\mathcal{P}_{2k+1}^n = \left\{P_{uv}^n \mid \exists(y, x) \in P_{uv}^n, b_{2k+1}^x = 1, u = 1, \ldots, U, v = 1, \ldots, V \right\}, \tag{17}
\]

which means that \( P_{uv}^n \) is collected if at least one pixel position \( \exists(y, x) \in \mathcal{P}_{2k+1}^n \) meets with \( b_{2k+1}^x = 1 \) at the above pyramid level \( 2k + 1 \).

The local image areas \((y, x) \in \mathcal{P}_{2k+1}^n\) originate from the manipulation of phase changes \( \phi_{y,x}^{2k+1} \) through Eqs. (1)–(14). Therefore, only the local image areas are targets to be magnified because of having the phase correlation between the adjacent pyramid levels [9], [10]. From this consideration, we applied Eqs. (3)–(10) only to the local image areas \((y, x) \in \mathcal{P}_{2k+1}^n\) and then obtained the magnified sub-band image \( \tilde{L}_{2k}^y \) at pyramid level \( 2k \) as follows.

\[
\tilde{L}_{2k}^y = \left\{ \begin{array}{ll}
\phi_{y,x}^{2k+1} \cdot \cos \left( \phi_{y,x}^{2k} \right) & \text{if } (y, x) \in \mathcal{P}_{2k+1}^n \\
0 & \text{otherwise}
\end{array} \right. \tag{18}
\]

This equation means that the magnification process is performed only at local image areas \((y, x) \in \mathcal{P}_{2k+1}^n\) at even-numbered pyramid level \( 2k \) (red dot image area in Fig. 2), and the remaining image areas are copies of the sub-band image \( l_{2k}^y \) (cyan dot image area in Fig. 2), so we achieve a short computational time in constructing an image pyramid for obtaining the magnification results, which we call the local Riesz pyramid (black and red line flow in Fig. 2).

Finally, we sequentially performed Eq. (12) from \( n = N \) to \( n = 0 \) over all time frames \( t = 1, \ldots, T \) and eventually obtained video magnification results where only subtle motion changes are magnified.

### 4.1 Generalized Local Riesz Pyramid

In the section above, we processed a set of two adjacent pyramid levels: odd- \((2k + 1)\) and even-numbered \((2k)\). However, we can generalize the number of pyramid levels in the set as \( M \) because the similarity of local phase changes remains across all pyramid levels as \( \phi_{y,x}^n \approx \lambda^1 \cdot \phi_{y,x}^{n+1} \approx \lambda^2 \cdot \phi_{y,x}^{n+2} \approx \lambda^3 \cdot \phi_{y,x}^{n+3} \approx \cdots \). If we allow large error in the similarity (Fig. 3), Therefore, given a set of \( M \in \{M \mid M|N, M \neq 0\} \)
pyramid levels with \( n = Mk, \ldots, Mk + M - 1 \) and \( k = 0, \ldots, \frac{N}{M} - 1 \), we can generalize Eqs. (13)–(18) as

\[
\delta_{y,x}^{Mk+M-1} = \sum_{t=1}^{T} \left( p_{y,x}^{Mk+M-1} - p_{y,x}^{Mk+M-1} \right)^2 \tag{19}
\]

\[
b_{y,x}^{Mk+M-1} = \begin{cases} 1 & \text{if } \delta_{y,x}^{Mk+M-1} > \epsilon \\ 0 & \text{otherwise} \end{cases} \tag{20}
\]

\[
p_{y,x}^{Mk+m} = \{ p_{y,x}^{Mk+m} \mid E(y,x) \in p_{y,x}^{Mk+M-1}, \]

\[
t = 1, \ldots, U, v = 1, \ldots, V \}, \tag{21}
\]

\[
p_{y,x}^{Mk+m} = \begin{cases} \delta_{y,x}^{Mk+m} \cos \left( \frac{\delta_{y,x}^{Mk+m}}{\epsilon} \right) & \text{if } (y,x) \in p_{y,x}^{Mk+m} \\ \delta_{y,x}^{Mk+M-1} & \text{otherwise} \end{cases} \tag{22}
\]

where \( m = M - 2, \ldots, 0 \) if \( M > 2 \), \( m = 0 \) if \( M = 1 \).

Note that the proposed method is the case of \( M = 2 \), and we can include the conventional Riesz pyramid [8] as the case of \( M = 1 \) by defining \((y, x) \in p_{y,x}^{Mk} = p_y^k\).

We expect a computational time to be shorter when we select a bigger \( M \) because the local image areas can be processed with a few detection procedures defined as Eqs. (19)–(20). However, we should keep in mind that the computational cost would be high when \( M \) is big and the large local image areas are selected in the first pyramid level \( Mk + M - 1 \) in the set of pyramid levels \( M \).

5. Results

5.1 Algorithmic Time Complexity

In this subsection, we analyzed algorithmic time complexity, i.e., how much the proposed local Riesz pyramid reduces the computational time in comparison with the conventional Riesz pyramid [8].

First Laplacian decomposition step, Eqs. (1)–(2), and final reconstruction step, Eq. (12), have the same computational time between the proposed method and the conventional one. However, intermediate step, Eqs. (3)–(11) or Eqs. (3)–(18), is completely different between them, so we first focused on analyzing algorithmic time complexity in the intermediate step.

Table 1 shows algorithmic time complexity for each equation in the intermediate step, Eqs. (3)–(11) or Eqs. (3)–(18). In a method using the Riesz pyramid [8], temporal filtering process, Eq. (9), has the longest computational time. Similarly, in our method, the temporal filtering process has the longest computational time despite the ROI detection process being newly added. Note that we assumed the number of time frame \( T \) is long enough. Therefore, the algorithmic time complexity strongly depends on the number of pixel positions \( H^nW^m \) and time frame \( T \) in the temporal filtering process, Eq. (9). For further analysis, we focused on the number of pixel positions at each time frame \( t \) in the temporal filtering process because it is the reduction target of our proposed method.

The conventional Riesz pyramid [8] processes the entire Laplacian pyramid \( \sum_{n=0}^{N-1} L_n^t \) at a time frame \( t \) in the temporal filtering process Eq. (9). Thus, its algorithmic time complexity can be defined with respect to the number of pixel positions of \( \sum_{n=0}^{N-1} L_n^t \) that are input to Eq. (9) as

\[
g_1 = \sum_{n=0}^{N-1} H^nW^mT \log T
\]

\[
= \sum_{n=0}^{N-1} \left( \frac{1}{A^2} \right) H^nW^mT \log T
\]

\[
= \frac{A^2}{A^2 - 1} \left( 1 - \frac{1}{A^2} \right)^N H^nW^mT \log T, \tag{23}
\]

where \( g_M \) indicates the algorithmic time complexity with respect to the number of pixel positions that are input to Eq. (9). Here, \( M = 1 \) is case of the conventional Riesz pyramid [8].

In contrast, our generalized local Riesz pyramid processes the partial Laplacian pyramid at a time frame \( t \) in the temporal filtering process Eq. (9). If the size of local image areas \( p_{y,x}^{Mk+M-1} \) detected by Eqs. (19)–(21) is the \( \frac{1}{q} \) \((q \in \mathbb{R}_+\) of the original image areas as \( p_{y,x}^{Mk+M-1} \) detected by Eq. (9), the algorithmic time complexity is described as follows.

\[
g_M = \sum_{k=0}^{k=N/M-1} \sum_{m=0}^{m=M-2} \frac{1}{q} H^{Mk+m}W^{Mk+m}T \log T
\]

\[
+ H^{Mk+M-1}W^{Mk+M-1}T \log T
\]

\[
= \sum_{k=0}^{k=N/M-1} \sum_{m=0}^{m=M-2} \frac{1}{q} \left( \frac{1}{A^2} \right)^{Mk+m} H^nW^mT \log T
\]

\[
+ \left( \frac{1}{A^2} \right)^{Mk+M-1} H^nW^mT \log T
\]

Table 1 Algorithmic time complexity for each equation in the intermediate step, Eqs. (3)–(11) or Eqs. (3)–(18). Note that \( |p_{2k+1}^{2k+1}| \) is the size of local image areas detected by ROI detection, Eqs. (13)–(17).

<table>
<thead>
<tr>
<th>Method</th>
<th>Pyr. level</th>
<th>ROI detection Eqs. (13)–(17)</th>
<th>Riesz transform Eq. (3)</th>
<th>Phase calculation Eqs. (6)–(7)</th>
<th>Temporal filtering Eq. (9)</th>
<th>Phase &amp; sub-band mag. Eqs. (10), (11) or Eqs. (10), (18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Riesz pyr [8]</td>
<td>( n )</td>
<td>–</td>
<td>( O(H^nW^m \log H^nW^m) )</td>
<td>( O(H^nW^mT \log T) )</td>
<td>( O(H^nW^mT) )</td>
<td>( O(H^nW^mT) )</td>
</tr>
<tr>
<td>Ours</td>
<td>( 2k + 1 ) max(( \epsilon ), ( O(H^{2k+1}W^{2k+1}) ))</td>
<td>( O(H^{2k+1}W^{2k+1} \log H^{2k+1}W^{2k+1}) )</td>
<td>( O(H^{2k+1}W^{2k+1} \log T) )</td>
<td>( O(H^{2k+1}W^{2k+1}T \log T) )</td>
<td>( O(H^{2k+1}W^{2k+1}T) )</td>
<td>( O(H^{2k+1}W^{2k+1}T) )</td>
</tr>
<tr>
<td>Ours</td>
<td>( 2k )</td>
<td>–</td>
<td>( O(</td>
<td>p_{2k}^{2k+1}</td>
<td>\log</td>
<td>p_{2k}^{2k+1}</td>
</tr>
</tbody>
</table>
conventional Riesz pyramid [8] is calculated as

\[
\frac{g_M}{g_1} = \left[ \frac{\lambda^2}{q(\lambda^2-1)} \left( 1 - \left( \frac{1}{\lambda^2} \right)^{M-1} \right) + \left( \frac{1}{\lambda^2} \right)^{M-1} \right] \\
\cdot \frac{\lambda^{2M}}{\lambda^{2M-1} - 1} \left( 1 - \left( \frac{1}{\lambda^2} \right)^N \right) H^0 W^0 T \log T
\]

(24)

Note that this equation holds for \( M \geq 2 \).

From these equations, the algorithmic time complexity ratio of the generalized Local Riesz pyramid against the conventional Riesz pyramid [8] is calculated as

\[
\frac{g_M}{g_1} = \left[ \frac{\lambda^2}{q(\lambda^2-1)} \left( 1 - \left( \frac{1}{\lambda^2} \right)^{M-1} \right) + \left( \frac{1}{\lambda^2} \right)^{M-1} \right] \\
\cdot \frac{\lambda^{2M}}{\lambda^{2M-1} - 1} \left( 1 - \left( \frac{1}{\lambda^2} \right)^N \right) H^0 W^0 T \log T
\]

(25)

Figure 4 shows the algorithmic time complexity ratio \( \frac{g_M}{g_1} \) as in Eq. (25) with different parameters \( \frac{1}{q} = 0, 0.01, \ldots, 1 \), \( M = 1, 2, 3, 6 \), and \( \lambda = 2, \frac{3}{2} \). In this figure, the computational time simply increases linearly in proportion to \( \frac{1}{q} \) because the image areas of Laplacian pyramid \( \sum_{n=0}^{N-1} L^q_n \) are larger, and the use of half-octave Gaussian pyramid \( \lambda = \frac{3}{2} \) (Fig. 4(b)) simply because the image areas of Laplacian pyramid \( \sum_{n=0}^{N-1} L^q_n \) are larger as well as above. In the case of \( \frac{1}{q} = 0.5 \), where the local image areas are detected as being half the size of the original ones, the algorithmic time complexity ratio \( \frac{g_M}{g_1} \) is near 0.6 at every \( M = 2, 3, N \). Therefore, under this condition, the proposed local Riesz pyramid is expected to be about 2x faster than the conventional Riesz pyramid [8]. Moreover, the algorithmic time complexity decreases in proportion to \( M \) but converges each value that equals the limit of \( \frac{g_M}{g_1} \) as \( \frac{1}{q} \) approaches zero. This convergence can be described as

\[
\lim_{\frac{1}{q} \to 0} \frac{g_M}{g_1} = \frac{\lambda^2 - 1}{\lambda^{2M} - 1}.
\]

(26)

This means the best case in our algorithm but indicates no local image areas to be magnified except for the pyramid level \( Mk + M - 1 \).

From the above analysis of algorithmic time complexity, the worst case of our method is the same computational complexity as the conventional Riesz method [8] (Fig. 4, \( \frac{1}{q} = 1 \)) and the best case is converged with Eq. (26). Thus, our method does not completely guarantee that the computational time will be reduced. However, our method can reduce it in most cases because the local image areas in input videos can be often detected by Otsu’s method [11] thanks to its simplicity and robustness (Fig. 4, \( q > 1 \)). Therefore, our proposed algorithm, local Riesz pyramid, can reduce the computational time in phase-based video magnification efficiently in most cases.

### 5.2 Real Videos

To evaluate the usefulness of our proposed method, which magnifies subtle motion changes within a short computational time, we conducted experiments on real videos for qualitative evaluation and synthetic ones with ground-truth magnification results for quantitative evaluation. We compared our proposed method \( M = 2 \) with a fast phase-based video magnification method using the Riesz pyramid proposed by Wadhwa et al. [8]. We set the parameters for each experiment as listed in Table 2. We performed each method in YIQ color space and divided a set of pixel positions \( P^n \) into \( U \times V = 20 \times 20 \) subsets. In all experiments, we specified the ideal bandpass filter as the temporal filter in Eq. (9), and pyramid level \( N \) as 6. All experiments were implemented by using C++ with OpenCV [20] and ran on a PC with an Intel Core i7-8559U CPU at 2.7 GHz, and 16 GB of RAM.

In Fig. 5, our objective was to magnify and visualize the subtle chest motions caused by the baby’s breathing. Comparing the proposed method and the method of Wadhwa et al. [8], both can magnify the subtle chest motions, but their magnification results are different. Our method has clearer subtlety than the conventional method.

Table 2 Parameters for all videos: amplification factor \( \alpha \), target frequency bands between \( f_1 - f_2 \), sampling rate \( f_s \).

<table>
<thead>
<tr>
<th>Video</th>
<th>( \alpha )</th>
<th>( f_1 )</th>
<th>( f_2 )</th>
<th>( f_s )</th>
</tr>
</thead>
<tbody>
<tr>
<td>baby</td>
<td>25</td>
<td>0.5</td>
<td>1.5</td>
<td>30</td>
</tr>
<tr>
<td>throat</td>
<td>50</td>
<td>100</td>
<td>120</td>
<td>2000</td>
</tr>
<tr>
<td>car</td>
<td>25</td>
<td>0.5</td>
<td>1.5</td>
<td>25</td>
</tr>
<tr>
<td>balance</td>
<td>20</td>
<td>1.5</td>
<td>3.0</td>
<td>30</td>
</tr>
<tr>
<td>drum</td>
<td>20</td>
<td>15</td>
<td>35</td>
<td>200</td>
</tr>
<tr>
<td>simulation</td>
<td>1–10</td>
<td>9–11</td>
<td>60</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 4 Algorithmic time complexity ratio \( \frac{g_M}{g_1} \) of our generalized local Riesz pyramid \( g_M \) against the conventional Riesz pyramid \( g_1 \) [8] with respect to \( \frac{1}{q} \) that determines size of local image areas.

Fig. 5 The breathing of a baby: visualizing subtle chest motions. We show the spatiotemporal slices (right panels) along a single red line in the left panel. Both a method using the Riesz pyramid [8] and our proposed method can magnify and visualize the subtle chest motions in the baby (see the right panels).
Video magnification for visualizing subtle skin vibrations of a stationary man who is speaking. We show the spatiotemporal slices (right panels) along a single red line in the left panel. Both a method using the Riesz pyramid [8] and our proposed method can magnify and visualize the subtle skin vibrations (see the right panels).

A car engine: visualizing subtle cyclic vibrations. We show the spatiotemporal slices (right panels) along a single red line in the left panel. Both the method using a Riesz pyramid [8] and our proposed method can magnify and visualize the subtle cyclic vibrations in the car engine (see the right panels).

A stationary man with a luggage: visualizing subtle tremors of a man in balance. We show the spatiotemporal slices (right panels) along a single red line in the left panel. Both the method using a Riesz pyramid [8] and our proposed method can magnify and visualize the subtle tremors of the man in balance (see the right panels).

In Figs. 10 and 11, we calculated the mean square error (MSE) to check the approximation error of the proposed local Riesz pyramid ($M = 2$) or the generalized one ($M = 3, N$) against the conventional Riesz pyramid [8] as ground-truth. Our proposed method (b) showed lower MSE around areas of baby’s chest compared with higher $M$; thus, it can detect the minimum number of sufficient local image areas for magnifying principal subtle chest motions.

Mean square error (MSE) between the proposed local Riesz pyramid and the conventional Riesz pyramid [8] as ground-truth. Our proposed method (b) shows lower MSE around areas of center of drum’s membrane compared with higher $M$; thus, it can detect the minimum number of sufficient local image areas for magnifying principal subtle skin vibrations.

Baby’s chest and the those of the center of drum’s membrane, respectively; thus, it can detect the minimum number of sufficient local image areas for magnifying principal subtle motions in the input videos. On the other hand, the MSE
increased in proportion to $M$ (c, d), which is the case of the generalized local Riesz pyramid (in particular, $M = N$ is the case for which we chose all pyramid levels except for the top $N$ pyramid level). These results indicate that the large error of the phase change’s similarity (Fig. 3) in proportion to $M$ directly affected the approximation error between the proposed and the conventional methods.

Table 3 shows the computational time and MSE against the conventional Riesz pyramid [8] in all input videos. We produced each magnification video result by using our proposed method ($M = 2$), the generalized one ($M = 3, N$), or that proposed by Wadhwa et al. [8]. This table confirms that our proposed method requires a shorter computational time in processing an input video than that proposed by Wadhwa et al. [8], with the lowest MSE between $M = 2, 3, N$. These results indicate that, for the baby and the simulation videos in $M$, large local image areas were chosen at the first pyramid level. Remarkably, our method often achieved almost half the computational time needed to process a video compared with [8]. In our experiments, it is considered that the local image areas are detected as being about half the size of the original one from the discussion in Sect. 5.1.

![Image](311x510 to 376x575)

**Table 3** Comparison of a computational time and MSE against the conventional Riesz pyramid [8] in an entire input video. In all real videos, our proposed method ($M = 2$) required a shorter computational time than the method proposed by [8] and also had MSE lower than $M = 3, N$.

<table>
<thead>
<tr>
<th>Video</th>
<th>Resolution $H^0 \times W^0 \times T$</th>
<th>Wadhwa et al. [8] computational time (s)</th>
<th>Ours, $M = 2$ comp. time (s)</th>
<th>MSE</th>
<th>Ours, $M = 3$ comp. time (s)</th>
<th>MSE</th>
<th>Ours, $M = N$ comp. time (s)</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>baby</td>
<td>544 × 960 × 301</td>
<td>31.62</td>
<td>12.79</td>
<td>4.94</td>
<td>6.33</td>
<td>9.69</td>
<td>8.95</td>
<td>13.43</td>
</tr>
<tr>
<td>throat</td>
<td>1144 × 720 × 300</td>
<td>40.56</td>
<td>16.89</td>
<td>3.62</td>
<td>12.98</td>
<td>4.27</td>
<td>12.23</td>
<td>6.31</td>
</tr>
<tr>
<td>car engine</td>
<td>452 × 888 × 300</td>
<td>20.99</td>
<td>13.68</td>
<td>5.73</td>
<td>10.63</td>
<td>11.83</td>
<td>8.42</td>
<td>54.48</td>
</tr>
<tr>
<td>balance</td>
<td>384 × 272 × 300</td>
<td>5.37</td>
<td>3.24</td>
<td>12.94</td>
<td>1.89</td>
<td>28.31</td>
<td>1.17</td>
<td>54.22</td>
</tr>
<tr>
<td>drum</td>
<td>360 × 640 × 450</td>
<td>15.15</td>
<td>9.81</td>
<td>6.70</td>
<td>7.26</td>
<td>13.97</td>
<td>5.82</td>
<td>82.58</td>
</tr>
<tr>
<td>simulation</td>
<td>512 × 512 × 240</td>
<td>9.04</td>
<td>3.37</td>
<td>0.77</td>
<td>2.47</td>
<td>2.16</td>
<td>3.31</td>
<td>8.30</td>
</tr>
</tbody>
</table>

### 5.3 Controlled Experiments

To evaluate the effectiveness of our proposed method qualitatively, we conducted controlled experiments to assess the mean square error (MSE) over all image pixel positions and time frames between a magnified synthetic video by each magnification method (Riesz pyramid [8], ours $M = 2$, and $M = 3, N$) and the ground-truth (Fig. 12). In this experiment, we set the pyramid level $N$ to 6. Fig. 12 (top left) shows a 4-second synthetic ball video. The ball had horizontal subtle motions defined as $d = 0.5 \cdot \sin(2\pi t)$. To obtain a ground-truth magnification video for the synthetic one, we created it while changing $d$ to $5 \cdot d$.

Figure 12 (top center) shows the MSE with a different magnification factor $\alpha = 1, \ldots, 10$, and the top right plot is an expanded version of the blue rectangle area in the top center plot. In the top right of this figure, there is almost no difference in MSE between our method ($M = 2$) and the method using a Riesz pyramid [8], so this indicates that our proposed method using the local Riesz pyramid can automatically process the minimum number of sufficient local image areas to perform phase-based video magnification.

In contrast, as $M$ increases, the difference in MSE against Riesz pyramid [8] spreads due to mis-detecting local image areas with the large error of the phase change’s similarity (Fig. 3). Additionally, the effects of the MSE appears in the spatiotemporal slices along the red line (the middle panels) and a green line (the bottom panels) in the input video (the top left plot). All methods can detect subtle ball motions at the left edge of the ball (the bottom panels), but they were ambiguously detected in the cases of $M = 3$ and $N$ at the top edge of the ball (the middle panels). Note that all methods minimize the MSE at the magnification factor $\alpha = 5$, which is consistent with the relationship between the synthetic video and the ground-truth (changing $d$ to $5 \cdot d$). On the other hand, our method outperforms the method using a Riesz pyramid [8] at a high magnification factor $\alpha \geq 6$ because it processes only local image areas and prevents unnecessary magnified output, which is produced by the sum of error in the magnification process.

In Fig. 13, we checked the effect of an input video that has long time frames on conventional method using the Riesz pyramid [8], our proposed method ($M = 2$), and the generalized one ($M = N$) because our method reduces the
number of pixel positions to be processed but adds extra processes, Eqs. (13)–(17); in particular, Eq. (17) is directly affected by the number of time frames due to its summation of times $t = 1, \ldots, T$. Note that we would conclude the effect of summation process in Eq. (17) on a computational time is trivial because of the temporal filtering process in Eq. (9), which exists both in the conventional and proposed method, is dominant for the computational time in a big O notation manner (Table 1). However, in this experiment, we check the effect of an input video that has long time frames on proposed method from an experimental point of view. In this experiment, we set the same conditions as the above control experiment condition except for the resolution $256 \times 256 \times T$, $T = 240, \ldots, 24000$, $N = 5$, and $\alpha = 5$. For checking the effect of an input video that has long time frames on each method, we evaluated a computational time and MSE against ground-truth. Figure 13 left shows that the all magnification methods have the linearity in terms of a computational time in proportion to the number of input time frames, and the linear trends are less likely to intersect between all methods. Note that our proposed method’s MSE is stable for all input time frames (Fig. 13 right). This result indicates that the effect of the number of time frames in an input video on our method is trivial; we can sufficiently ignore the time delay due to the summation process in Eq. (17). Therefore, both the conventional method using Riesz pyramid [8] and our proposed method simply react to the increase or the decrease in the number of input time frames (Fig. 13) and pixel positions (Fig. 4).

6. Discussions and Limitations

We focused on the correlation of phase changes between adjacent pyramid levels and proposed a novel decomposition technique called the local Riesz pyramid that automatically processes the minimum number of sufficient local image areas for faster phase-based video magnification. This enables us to output good video magnification results with a shorter computational time than the previous work [8]. It is expected that our method will spark the application of video magnification to real environments where high-speed processing is needed, but there are a number of limitations as follows.

Our proposed method achieves good magnification results equivalent to the previous work [8] despite the fact that our method processes only local image areas at several pyramid levels. We consider this to be because our method identifies no-motion local image areas and magnifies only local image areas where subtle motions exist. However, the boundaries of the local images are considered to have the negative effect of producing discontinuous video magnification results. Fortunately, the boundary effect is hardly seen because our local processing is applied only to even-numbered pyramid levels (Figs. 10, 11 (b)). In contrast, as $M$ increases, the boundary effect clearly appears and leads to high MSE against the conventional Riesz pyramid [8] (Figs. 10, 11 (c, d)). One possible approach to further reducing this boundary effect is weighting the magnification factor $\alpha$ near the boundaries using a Gaussian distribution, but we need to propose a radical way of overcoming this problem in the future.

In our proposed method, the plausible local image areas to be processed are estimated by using all image frames (Eq. (13)) because we focused on cyclic subtle changes from the beginning to the end of an input video, rather than strong changes over a short period of time. This suggests that we implicitly assume that subtle motion changes to be magnified will stay in the same local image areas over all time frames, in other words, objects do not move largely. Therefore, to magnify subtle motion changes under the presence of large motions within a short computational time, we need to develop a method where the local image area to process is adaptively determined in each image frame without misdetection of subtle changes.

7. Conclusion

We proposed the local Riesz pyramid for faster phase-based video magnification to produce magnification results within a short computational time. With the local Riesz pyramid, we focused on the correlation of phase changes between adjacent pyramid levels and automatically processed the minimum number of sufficient local image areas at several pyramid levels to perform phase-based video magnification. The results of several experiments demonstrate that our method achieves good magnification results equivalent to previous work [8] within a short computational time.

Acknowledgements

The authors are grateful to N. Wadhwa, M. Rubinstein, F. Durand, and W.T. Freeman for sharing their dataset publicly via URLs in [1], [8].

References


