

# Semi-Automated Magnification of Small Motions in Videos

Sushma M<sup>1</sup>, Anubha Gupta<sup>2</sup>, and Jayanthi Sivaswamy<sup>1</sup>

<sup>1</sup> Center for Visual Information Technology, IIT Hyderabad, India.

<sup>2</sup> Signal Processing and Communications Research Center, IIT Hyderabad, India.  
{sushma.m@research.iit.ac.in, agupta@iit.ac.in, jsivaswamy@iit.ac.in}

**Abstract.** In this paper, we present a semi-automated method to magnify small motions in videos. This method amplifies invisible or hidden motions in videos. To achieve motion magnification, we process the spatial and temporal information obtained from the video itself. Advantage of this work is that it is application independent. Proposed technique estimates required parameters to get desirable results. We demonstrate performance on 9 different videos. Motion magnification performance is equivalent to existing manual methods.

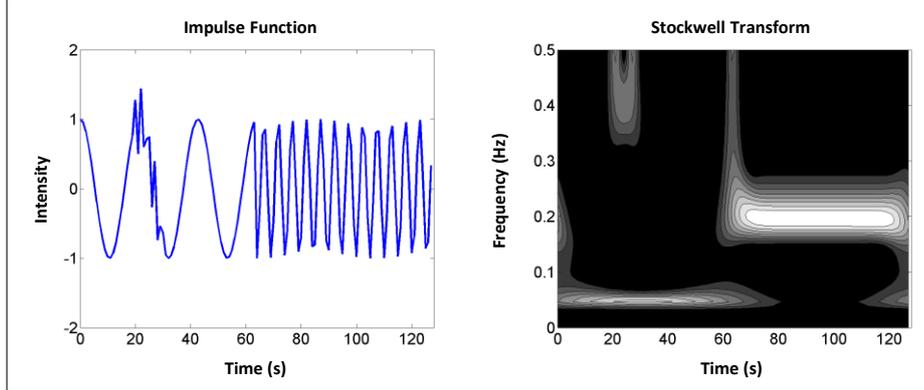
**Keywords:** Motion Magnification, Time Frequency Analysis, Stockwell Transform, Automated Process.

## 1 Introduction

Human eye and its brain interface can visualize or detect the motion within a certain range of spatial and temporal frequencies. But in most of the cases, it might be possible that frequencies which are below this range also can have useful information. We can simplify this by saying that there can be small motions which are not visible to the naked eye. Even though these small motions are difficult to detect, they may contain useful information.

There are several applications for motion magnification, for example, visualization, physical diagnosis, pre-measurement planning for precise physical measurements, and surveillance etc. [7]. Therefore, there is a large scope for research to develop a system to amplify or magnify small motions.

There are previous works in this direction. In [1,2], human motions are generated by reusing the captured motion to create new motions. In this paper, we propose to process the data obtained from the video and reconstruct the video from the modified data such that new video shows magnified motion. In [3], small motion is magnified without tracking motion. However, this approach needs users to provide a set of parameters as input for every video in order to magnify motion. For unknown video, this consumes time to find desirable parameters. While all these techniques require user interaction in some or the other way, we are not aware of any previous work addressing automated approach to magnify small motions. This motivated us to develop a mechanism to estimate these parameters automatically from the given data. We use a time frequency representation



**Fig. 1.** An example of Stockwell transform. Time signal shown in left is  $h[0 : 63] = \cos(2\pi t * 6/128)$ ,  $h[64 : 127] = \cos(2\pi t * 25/128)$ ,  $h[20 : 30] = h[20 : 30] + 0.5 * \cos(2\pi t * 52/128)$ . It contains a low-frequency signal for the first half, a middle-frequency signal for the second half and a high-frequency burst at  $t = 20$ . All these frequencies are clearly visible along with time location in Stockwell transform as shown in right (bright pixels indicate high strength of transform).

called Stockwell transform for this purpose. We introduce computationally inexpensive techniques to estimate parameters. We illustrate the utility of the proposed method on examples in which small motions were made visible.

## 2 Background

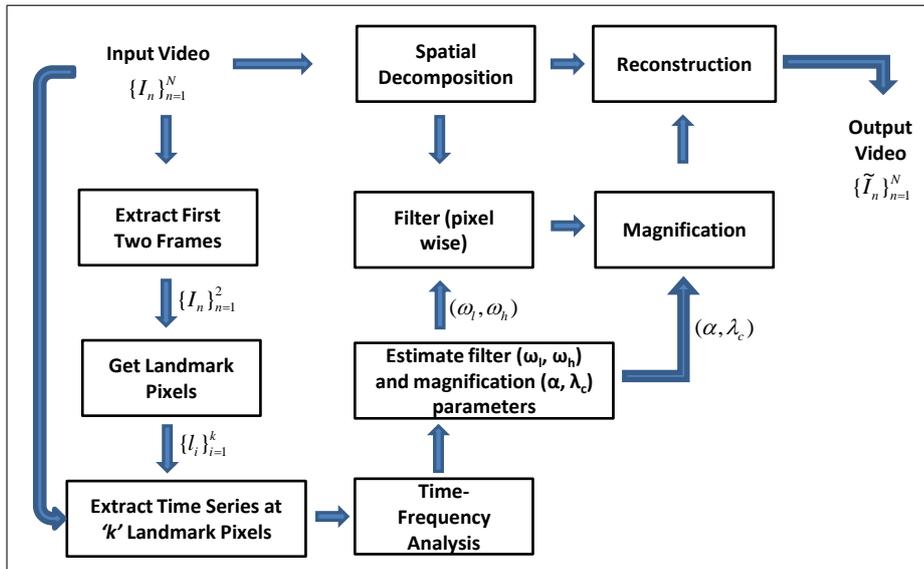
### 2.1 Stockwell Transform

Even though Fourier transform gives the information about the spectral components in a signal, it fails to locate where those frequencies occur in that signal. So, it is preferable to consider time frequency representation (TFR). Different techniques for time frequency representation have been proposed. A few of them are short time Fourier transform (STFT), Gabor transform, continuous wavelet transform (CWT) and Wigner ville distribution etc. In [6], it was proven that Stockwell transform outperforms all these TFR techniques in localizing time and frequency because it has frequency dependent resolution whereas other transforms have windows of fixed width.

For a given time signal  $h(t)$ , its Stockwell transform is defined as,

$$S(\tau, f) = \int_{-\infty}^{\infty} h(t) \frac{|f|}{\sqrt{2\pi}} e^{-(\tau-t)^2 f^2 / 2} e^{-i2\pi f t} dt \quad (1)$$

where  $h(t)$  is the time signal,  $f$  denotes the frequency and  $\tau$  denotes time shift parameter. An example is shown for a synthetic signal in Fig. 1.



**Fig. 2.** Overview of the proposed framework. Initially, parameters  $(\omega_l, \omega_h, \alpha, \lambda_c)$  are estimated from the landmark pixels  $(\{l_i\}_{i=1}^k)$  determined from the first two frames of input video  $(\{I_n\}_{n=1}^2)$ . Then input video  $(\{I_n\}_{n=1}^N)$  is divided into spatial bands, which are pixel-wise filtered with same temporal filter and amplified by magnification factor,  $\alpha$ . Reconstruction  $(\{\tilde{I}_n\}_{n=1}^N)$  is done by adding these amplified bands along with original spatial bands.

## 2.2 Semi-Automated Video Magnification

Video magnification can be defined as magnifying or amplifying small motions in videos. This process helps in visualizing invisible or hidden deformations. It can be done by (i) detecting or tracking motion in a video and then by amplifying that motion [7,8]; (ii) directly magnifying motion without detecting or tracking [3]. We follow the second approach because tracking motion is computationally expensive. We magnify motion automatically unlike [3]. In [3], every video requires a set of parameters as input whereas in our method, we determine these parameters automatically by using the information from the video.

## 3 Parameter Estimation

To magnify small motions, in [3], given video is decomposed spatially. Then user has to set parameters for filtering and magnification. In our work, we estimate these parameters automatically. Main contribution of this work is to magnify the small motions in video automatically. It consists of the following two contributions: (1) Estimate bandwidth for temporal filter and (2) Estimate the magnification parameters, automatically.

Overview of the proposed framework is given in Fig. 2. Given an input video of  $N$  frames,  $\{I_n\}_{n=1}^N$ , the proposed method consists of the following steps: **(1) Estimation of parameters:** (i) Extract first two frames,  $\{I_n\}_{n=1}^2$ ; (ii) Estimate landmark pixels,  $\{l_i\}_{i=1}^k$ ; (iii) Extract time series at these ‘ $k$ ’ pixels from  $N$  frames; (iv) Apply time frequency representation; (v) Estimate parameters for bandpass filter  $(\omega_l, \omega_h)$  and magnification  $(\alpha, \lambda_c)$ ; **(2) Magnification:** (i) Decompose the given video spatially into different spatial bands; (ii) Apply pixel-wise bandpass filter with  $\omega_l$  and  $\omega_h$  as lower and higher cut-off frequencies; (iii) Multiply with magnification factor,  $\alpha$  for wavelengths less than  $\lambda_c$ ; **(3) Reconstruction:** (i) Reconstruct the video,  $\{\tilde{I}_n\}_{n=1}^N$  by adding magnified signal to original signal.

To get the landmark pixels, it was observed that considering all pixels for determining bandwidth and magnification factor is not efficient due to the fact that (1) the whole process will be time consuming and (2) videos, considered for this work, have small motions in few regions and hence, all pixel locations do not necessarily undergo motion. Therefore, we adapted a mechanism to find landmark pixels. These pixels are obtained from the difference of edge maps of first two frames of a given video because the pixels at edges definitely experience motion from one frame to another frame. Time signals at these landmark pixels are used to determine the parameters.

Previous works, for example [5] has shown that temporal processing is done generally to extract invisible information from the signal. In general, spectral band is determined either empirically or assumed depending on application but such approaches fail to find the dominant frequency components automatically. We eliminate such dependency. In this work, we consider time signals from a video sequence, only at the locations of landmark pixels. We find bandwidth for bandpass filtering the time signals automatically. These limits of frequency are determined using Stockwell transform. It is observed that the minimum and maximum frequencies of mean of frequencies obtained from each time series of a pixel can give useful information for temporal processing. We carried out this processing on time series data in Y space of YIQ colour space.

From Stockwell transform, we estimate the magnification factor  $\alpha$ . It is determined as follows,

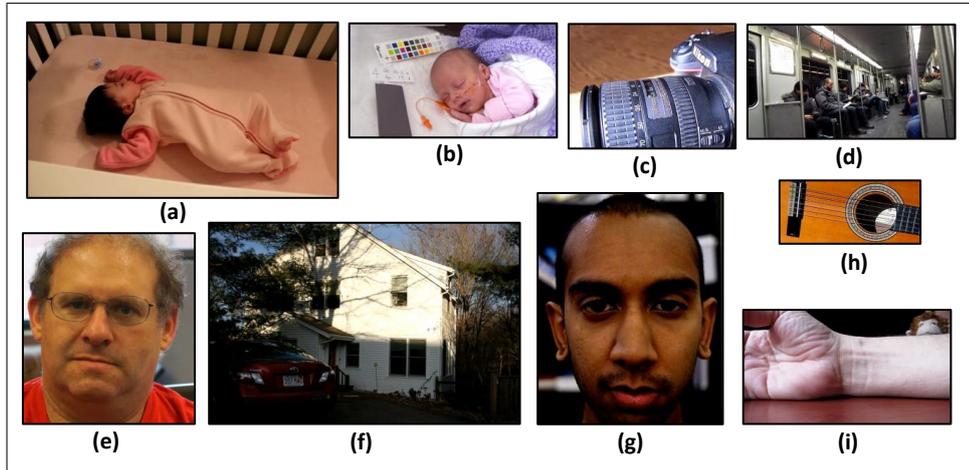
$$(1 + \alpha)h(t) < \frac{\lambda_c}{8} \quad (2)$$

where  $h(t)$  is time signal and  $\lambda_c$  is cut-off wavelength beyond which magnification factor,  $\alpha$  is zero. Eq. (2) provides the largest magnification factor. As explained in Section 2.1, Stockwell transform gives time frequency representation. We utilize this time frequency representation to find desirable magnification factor  $\alpha$ . From Stockwell transform, we obtain information about which wavelengths (or frequencies) are occurring at what times. Since videos involve small motion, we need to consider landmark pixels at which time signals can be used to determine magnification factor. A few of these time signals may contain noise and hence, may give incorrect  $\alpha$  and  $\lambda_c$  values. To overcome this problem, we have considered median of all magnification factors and maximum of all wavelengths obtained from Eq. (2).

Spatial decomposition of videos into different spatial frequency bands is done to increase the signal-to-noise ratio using Laplacian pyramid [9]. These frequency bands are filtered and magnified pixel-wise differently according to the level in pyramid [3] using estimated parameters. Reconstruction is done by adding original signal from spatial decomposition and magnified signal.

## 4 Experiments and Results

We use data given in [4] for our experiments. This data consists of 9 videos, namely *baby*, *baby2*, *camera*, *face*, *face2*, *guitar*, *shadow*, *subway* and *wrist* as mentioned in Table 1. Sample frames from these videos are shown in Fig. 3. All experiments are implemented using MATLAB on a system with 4GB RAM and Intel<sup>®</sup> core i5 CPU with 2.5 GHz processor. Every video takes time in the order of a few minutes to compute.



**Fig. 3.** Sample frames from videos used in experimentation (a)*baby* (b)*baby2* (c)*camera* (d)*subway* (e)*face* (f)*shadow* (g)*face2* (h)*guitar* (i)*wrist*. All these are shown in proportion with the size of the corresponding video.

Filter parameters, magnification factor and cut-off wavelength are determined from Stockwell transform by considering only landmark pixels. Obtained magnification factors  $\alpha$  and cut-off wavelengths  $\lambda_c$  for each video are listed in Table 1. Even though for some videos, magnification factors are not close to the values mentioned in [3], reconstructed videos are comparable to those of [3]. These results are available on <http://researchweb.iiit.ac.in/~sushma.m/premiResults>.

Video	$\alpha$		$\lambda_c$	$\omega_l$ (Hz)	$\omega_h$ (Hz)
	Proposed	[3]			
<i>baby</i>	32	10	30	0.5	3
<i>baby2</i>	30	150	30	0.6	2.8
<i>camera</i>	7	120	29	1.5	16.9
<i>face</i>	55	100	30	0.6	5.6
<i>face2</i>	17	20	30	1.02	10.56
<i>guitar</i>	101	100	29	0.77	8.16
<i>shadow</i>	13	5	30	0.92	5.8
<i>subway</i>	42	60	29	0.52	4.05
<i>wrist</i>	11	10	30	0.5	4

**Table 1.** Estimated parameters for magnification factor  $\alpha$ , cut-off wavelength  $\lambda_c$ , lower cut-off frequency for temporal filter  $\omega_l$ , higher cut-off frequency for temporal filter  $\omega_h$ .

## 5 Conclusions and Future Work

This paper presents a semi-automated method for magnifying small motions in videos. We have demonstrated that the parameters required to magnify the motion can be generated in automated fashion by using the time frequency representation called Stockwell transform. However, it is observed that noise is introduced in some reconstructed videos. In future, we would modify the proposed method to reduce this noise and attempt to make this process fully automatic. We believe that this method can have potential applications in medical imaging.

## References

1. Arikan O. and Forsyth D. A.: Synthesizing constrained motions from examples, ACM Trans. on Graph. (2002)
2. Lee J., Chai J., Reitsma P. S. A., Hodgins J. K. and Pollard N. S.: Interactive control of avatars animated with human motion data, ACM Trans. on Graph. (2002)
3. Hao-Yu Wu, M. Rubinstein and E. Shih, J. V. Guttag, F. Durand and W. T. Freeman: Eulerian video magnification for revealing subtle changes in the world, SIGGRAPH (2012)
4. Eulerian Video Magnification for Revealing Subtle Changes in the World, <http://people.csail.mit.edu/mrub/vidmag/>
5. Poh M. Z., McDuff D. J. and Picard R. W.: Non-contact, automated cardiac pulse measurements using video imaging and blind source separation, Optical Express (2010)
6. R. G. Stockwell, L. Manisha and R. P. Lowe: Localization of the Complex Spectrum: The  $S$  Transform, IEEE Trans. on Signal Processing (1996)
7. Liu C., Torralba A., Freeman W. T., Durand F. and Adelson E. H.: Motion magnification, ACM Trans. Graph. (2005)
8. Wang J., Drucker S. M., Agrawala M. and Cohen M. F.: The cartoon animation filter. ACM Trans. Graph. (2006)
9. Burt P. J. and Adelson E. H.: The Laplacian Pyramid as a Compact Image Code. IEEE Trans. on Comm. (1983)