




# Saliency Driven Video Motion Magnification

Manisha Verma<sup>1</sup>, Ramyani Ghosh<sup>2</sup>, and Shanmuganathan Raman<sup>1</sup>

<sup>1</sup> Electrical Engineering, IIT Gandhinagar, Gandhinagar, Gujarat, India  
manisha.verma89@gmail.com, shanmuga@iitgn.ac.in

<sup>2</sup> Computer Science Engineering, PES University, Bengaluru, Karnataka, India  
ramyani.ghosh@gmail.com

**Abstract.** The main goal of the proposed work is to detect certain spatial and temporal changes in videos that are not visible to the human eye and magnify them in order to make them perceptible while making sure that the background noise is not amplified. We apply Eulerian motion magnification on only the salient area of each frame of the video. The salient object is processed independent of the rest of the image using alpha matting aided by scribbles. We demonstrate the need to isolate the salient object from background motions and propose a simple and efficient way to do so. The proposed algorithm is tested on videos with imperceptible motion along with background motion to illustrate the significance of the proposed method. We compare the proposed method with linear and phase based Eulerian motion magnification techniques.

## 1 Introduction

Our environment constantly undergoes changes in color and motion. A lot of these changes are so subtle that they are imperceptible to the human eye. Needless to say, revealing these invisible changes via video processing can find numerous applications and provide interesting study in many research problems. The analysis of these changes gives significantly valuable information that is not visible naturally. Important phenomena such as human pulse from the wrist can be studied through these methods of video magnification. Efficiency, accuracy, and computational simplicity are necessary in order to use these concepts and obtain easily functional results.

Previously, Lagrangian and Eulerian magnification methods have been applied in order to make these changes visible in videos [1, 2]. Lagrangian method works on the optical flow of each pixel in the video and magnifies the motion. On the other hand, Eulerian motion magnification evaluates temporal changes in two consecutive frames and magnifies accordingly. Most of the existing video magnification techniques, such as Eulerian Video Magnification (EVM), involve magnification of all the motions in the video uniformly. This process tends to unnecessarily magnify large motions in the background and may lead to noise. The unwanted magnification of all motions in the video disrupts the study of the phenomenon of amplification of imperceptible motion.

In this paper, we discuss the use of saliency detection in order to isolate the salient region in the video [3]. Our approach is to detect the region of interest (ROI) so that video magnification can be selectively applied while leaving other motions or variations unchanged. In doing so, we have used saliency detection to get a primary idea of ROI and then apply image matting in order to get a near perfect ROI with fuzzy boundary. For motion magnification, we use Eulerian motion magnification technique which includes both linear and phase based methods. Eulerian approach is faster than Lagrangian as it uses temporal change of full frame at a time and not the optical flow of each pixel. We analyze the results of both linear and phase based Eulerian methods on the ROI.

The main contributions of the proposed work are as follows:

- We propose a fully automatic motion magnification method in order to magnify only the ROI using saliency and EVM.
- While magnifying only the ROI, the proposed approach is able to reduce noise because of the presence of the background motion.

The rest of the paper is organized as follows. Section 2 presents a brief literature review related to the problem. Saliency, image matting, and motion magnification are briefly illustrated in Sect. 3. Section 3.4 presents the proposed framework. Section 4 shows the experiments and results. Finally, the paper has been concluded in Sect. 5.

## 2 Related Work

Magnification of motion in videos is a field that has been explored by researchers for many years now. Numerous approaches have been made and there are multiple techniques for efficient video magnification. A Lagrangian approach was introduced by Liu *et al.* where pixel trajectories were magnified in order to magnify the motion [1]. Wang *et al.* proposed “cartoon animation filter”, where an input signal was modified to make it look more animated [4]. Wu *et al.* proposed a motion magnification technique for subtle changes in the video by the method of EVM [2]. It uses Laplacian pyramids for spatial decomposition and temporal filters to calculate the temporal difference and magnification is performed on the temporal difference using an amplification factor. Further, it was improved by using complex steerable pyramids. Wadhwa *et al.* utilized the concept of phase in optical flow to develop the phase based video motion magnification [5] which was further improvised using Riesz pyramid in order to perform real-time phase based motion amplification [6]. In the Riesz pyramid, the image features are phase shifted only along their dominant directions as opposed to complex steerable pyramids where it is done for every orientation.

These methods have found numerous applications over the years. Balakrishnan *et al.* tracked subtle head motions by performing principal component analysis (PCA) and used it to extract heart rate and beat lengths from videos [7]. Eulerian magnification and empirical mode decomposition (EM-EMD)

have been used to differentiate between emotional and physical stress on performing frequency division processing [8]. Face spoofing detection and facial expression recognition has been performed using Eulerian motion magnification [9, 10]. Detection of pulse transit time and finger vein liveness detection have been performed using the Eulerian color amplification [11, 12].

Elgharib *et al.* used a layer based magnification that enables the magnification of small motions within large ones [13]. They used manually marked scribbles and image matting to single out the ROI and magnify subtle motions in that region. Kooji *et al.* selectively magnified motion based on the depth of layers using bilateral filters that require additional depth map information along with input video [14]. Zhang *et al.* proposed motion magnification using acceleration [15]. Instead of magnifying linear changes, they chose to magnify acceleration to avoid magnification of large motions. The proposed approach also aims to magnify the subtle motions and avoid large background motions. It does not require any additional information to locate the object of interest as it does in [13, 14].

### 3 Saliency Driven Motion Magnification

#### 3.1 Saliency Detection

Visual saliency is the property that makes a region or a group of regions in an image distinctively stand out from its surroundings. In the proposed work, we have used saliency detection using structure matrix decomposition (SMD) [3]. It calculates the saliency of any part of an image by calculating its color and pattern distances from its immediate surroundings. It is performed by segmenting the image into super pixels and comparing the color distances between adjacent super pixels. Simple Linear Iterative Clustering (SLIC) is used to divide the image into superpixels which makes comparison and detection easier. Superpixels simplify the extraction of features (color, edge, and texture) and help to form a feature matrix. Model based on structure matrix decomposition to detect salient object is as follows.

$$\min_{B, F} \Psi(B) + a\Omega(F) + b\Theta(B, F) \quad \text{s.t.} \quad M = B + F, \quad (1)$$

where  $M$ ,  $B$ , and  $F$  are the feature matrix, background, and salient foreground respectively.  $\Psi(\cdot)$  is a low-rank constraint,  $\Omega(\cdot)$  is a structured sparsity regularization,  $\Theta(\cdot, \cdot)$  is an interactive regularization term to enlarge the distance between the subspaces drawn from  $B$  and  $F$ , and  $a, b$  are positive parameters.

Distance between feature subspaces of salient objects ( $F$ ) and background ( $B$ ) is calculated using following distance measure.

$$d(B, F) = \|BB^T - FF^T\|_M^2 \quad (2)$$

SMD works by generating an index tree. Index tree is constructed from a high-level prior map obtained from location, color, and background priors. This guides the matrix decomposition, enabling better detection. On incorporating

the feature matrix and the high-level prior map, it defines the low rank part and the structured sparse part. Saliency map example for three video frames are shown in Fig. 1.



**Fig. 1.** Saliency examples. First and second columns represent first frames of videos and their saliency maps respectively.

### 3.2 Image Matting

In order to find the ROI with accuracy, image matting is performed. To do image matting, all the color and the pattern variations of the foreground need to be detected and isolated from the background while making sure that the boundary of the foreground is maintained. This is done with the help of scribbles. In order to draw appropriate scribbles, we have used super pixel segmentation and Bezier curves [16]. Initially, the salient map of the first frame is obtained using SMD based saliency detection as described in Sect. 3.1. A gray scale saliency map is converted into black and white image as an approximation of background and foreground regions using a threshold parameter. To get the region in which scribbles are to be drawn, we must make sure that no scribbles are drawn on the boundary of the object as this may cause ambiguity in the detection of the edges of the ROI in image matting. To avoid that, we invert the black and white regions and erode the boundaries, thus, effectively shrinking our detected region away from its borders. Once the above steps are done, we draw scribbles in the foreground and background regions followed by the process of alpha matting.

**Scribbles:** Scribbles are randomly generated using Bezier curves that are drawn on the image to distinguish between regions [16]. The scribbles are generated such that their shape and curvature ensure that the maximum variations in color and pattern of the ROI are covered and thus included in the final matte. First, the superpixels are computed to divide the foreground and background regions. The Bezier curves are generated by first selecting a random pixel within the superpixel. The next pixel is obtained by finding the mirror image of the first pixel with respect to the centroid of the current superpixel. The remaining points are placed within the superpixel by rotating and translating the same point by a random angle. The curve is then drawn by connecting six points inside the superpixel. Thus, random scribbles are generated in each superpixel.

**Matting:** Matting is a technique to find a gray scale approximation of foreground and background image regions with a fuzzy boundary. Matting is performed using a trimap (a pre-segmented image with absolute background, absolute foreground and unknown region) or scribbled image. An image intensity can be written as the combination of background and foreground as follows.

$$I(x, y) = \alpha(x, y)F(x, y) + (1 - \alpha(x, y))B(x, y) \quad (3)$$

where  $I(x, y)$ ,  $F(x, y)$ ,  $B(x, y)$ , and  $\alpha(x, y)$  are pixel intensity, foreground intensity, background intensity, and foreground opacity at the position  $(x, y)$  respectively. We have utilized closed form image matting method using a scribbled image. It assumes smoothness in foreground and background in the local region of the neighborhood of each pixel and minimizes a cost function in  $\alpha$  in order to calculate the matte.

### 3.3 Motion Magnification

Motion magnification methods magnify the motion in order to make it more visual. Motion magnification for subtle changes makes it possible to convert imperceptible into perceptible motion. In the proposed work, we have analysed linear and phase based Eulerian motion magnification techniques in order to magnify salient region in the video. Brief descriptions of both the methods are given below.

**Linear Eulerian Video Magnification:** Linear EVM decomposes the video frames into a Laplacian pyramid [2]. Temporal filters are applied on spatial bands to detect the variations in specific frequencies. Temporal processing is uniformly performed within each band of every two consecutive frames and their temporal differences are obtained. The extracted temporal differences are multiplied by the magnification factor. The magnified temporal differences are reconstructed and added back to the original frame in order to produce desired results.

A simple translational motion is considered to connect temporal processing and motion magnification [2]. If the linear displacement function is  $d(t)$ , image intensity as a function of the linear motion is given by:

$$A(x, t) = f(x + d(t)) \quad (4)$$

The magnified intensity can then be given by the expression:

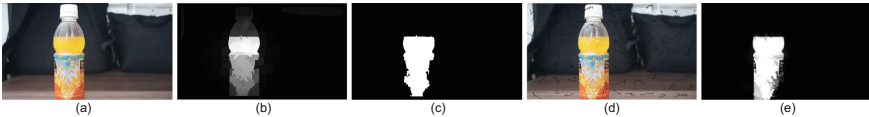
$$\bar{A}(x, t) = f(x + (1 + \beta)d(t)) \quad (5)$$

where  $\beta$  is the magnification factor. This is a simple method with very low computational cost involved, however it works for small magnification factors and magnifies background noise with an increment in the magnification factor.

**Phase-Based Motion Magnification:** Phase based motion magnification uses the concept of the phase and applies magnification to phase differences in order to magnify the motion [5]. This method of video magnification uses complex steerable pyramids. The motion is amplified in selected frequency bands by performing temporal processing. Small motions are amplified by modifying the local phase variations in a complex steerable pyramid for every orientation and scale. Once magnification is applied, the original video is reconstructed from the pyramid.

### 3.4 Framework

We have used both linear and phase based methods to magnify the video for salient region. Since the motion of desired object is very less, scribbles are drawn only on the first frame of the video and it is assumed that the scribbles will be on the right place throughout the video as the object has no large motion. The method has two phases, the first is to obtain foreground object and the second is to apply magnification to the foreground object only.



**Fig. 2.** (a) Video frame, (b) Saliency map, (c) Threshold, (d) Scribbles and (e) Alpha matte.

Pipeline of the proposed work with linear and phase based methods is as follows.

### 3.5 Phase 1

Initially, the video is uploaded and saliency map of the first frame is calculated using structured matrix decomposition (SMD) saliency detection [3] as shown in Fig. 2(b). The saliency map has gray scale intensities where light to dark represents higher saliency to lower saliency of object. To determine the foreground

object, gray scale saliency map is thresholded and a black & white image is obtained for a possible foreground and background approximation (Fig. 2(b)). The foreground and background object boundaries are eroded before drawing scribbles in order to maintain the scribbles only inside the object and not on the boundaries. After erosion, the foreground and the background images are divided into super pixels and at the centroid of each super pixel, Bezier curves are drawn [16, 17] (Fig. 2(d)). Image matting is performed with the help of closed form alpha matting as shown in Fig. 2(e).

### 3.6 Phase 2

In phase two, we have used both linear and phase based methods to magnify the salient region. Work flows of both the methods are as follows.

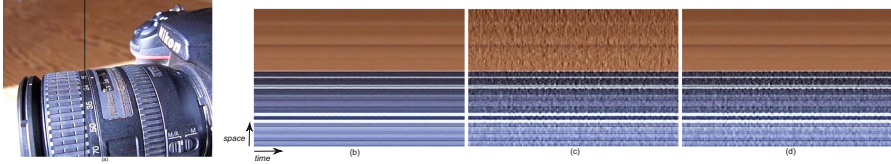
**Linear:** Input video frames are decomposed into sub-band images using the Laplacian pyramid. Every pixel of each sub-band image is passed through a temporal filter and temporal difference of two consecutive frames is obtained. Temporal difference is magnified using a magnification factor and multiplied with the alpha matte (calculated in Phase 1) in order to remove the magnification of the background. The magnified temporal difference of foreground region is reconstructed using the Laplacian pyramid and added back to the original frame. The process is applied to all the video frames in order to obtain the resultant video.

**Phase Based:** Alpha matte is multiplied with all the video frames and complex steerable pyramids are calculated for each frame. Phase difference is calculated between the frames and magnified in order to magnify motion. After magnification, pyramid is reconstructed in the original form and summed up with the background of each frame (inversion of alpha matte), and hence the motion magnified video is produced.

## 4 Experiments and Results

Experiments have been performed on various videos which were previously used in [2, 5] for motion magnification and new videos with large background motions. The proposed method uses both the linear and the phase based motion magnification. Hence, experiments have been performed to test both the techniques and results have been compared with both the previous methods [2, 5]. Threshold parameter which is used to convert saliency map to approximate background and foreground (black and white), is kept empirical (approximately .45 and can be adjusted as required) as motion magnification of the foreground is significant and it should not be missed and small intervention of background as false foreground is acceptable. Moreover, this parameter can be set automatically according to intensities of saliency map. Motion magnification parameters depend on the frequency which needs to be magnified. A few experimental results are presented

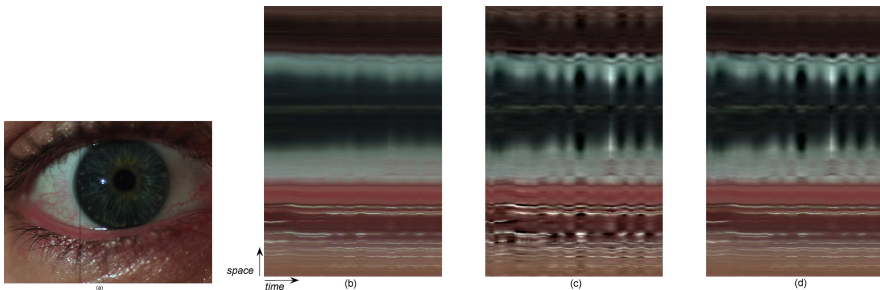
here to show the absence of noise in the proposed method as compared to the previous methods. Space-time graphs for a particular pixel range are plotted in order to measure the difference in motion from the original video and the motion magnified video.



**Fig. 3.** (a) First frame of camera sequence with highlighted black line, (b) space-time plot of original video of highlighted black line, (c) motion magnified plot using [2], and (d) motion magnified plot using the proposed method.

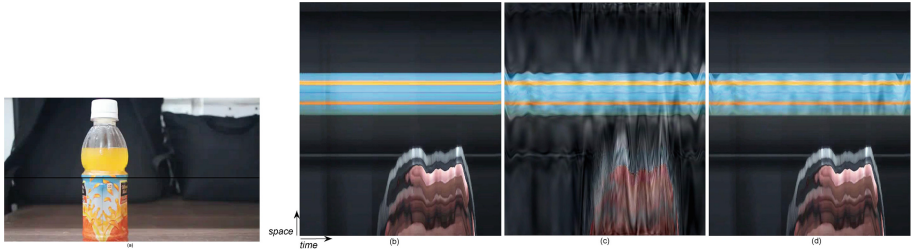
In all the experiment figures, first original frame of video is shown with some highlighted pixels with black line, which are shown in space-time intensity plot. In Fig. 3, variations of camera video sequence are shown. Figure 3(a) is first image of camera video sequence and horizontal black line is shown for original video, linear Eulerian motion magnification and saliency driven linear motion magnification in Fig. 3(b), (c) and (d). Camera motion is nearly invisible to human eyes as shown very straight in Fig. 3(b). Motion magnified video using [2] is shown to be producing noise along with motion magnification for camera.

In Fig. 4, variations of eye video sequence are shown. Figure 4(a) is first image of eye video sequence and vertical black line is shown for original video, linear Eulerian motion magnification and saliency driven linear motion magnification in Fig. 4(b), (c), and (d). Eye ball movement is nearly invisible as it is seen to be straight in Fig. 4(b). Motion magnified video using [2] is shown to be producing noise due to magnification surrounding motions.

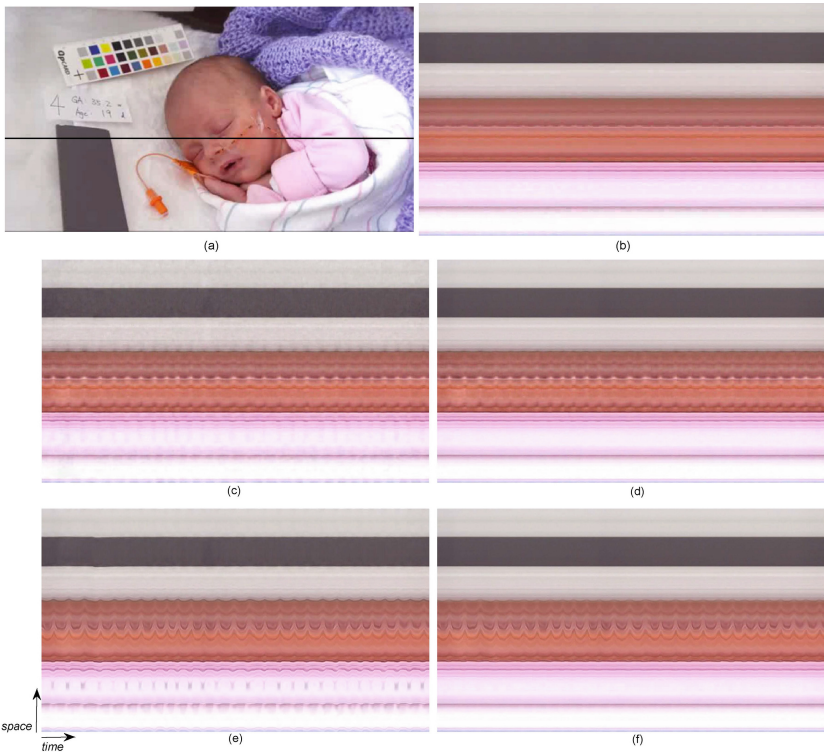


**Fig. 4.** (a) First frame of eye sequence with highlighted black line, (b) space-time plot of original video of highlighted black line, (c) motion magnified plot using [2], and (d) motion magnified plot using the proposed method.



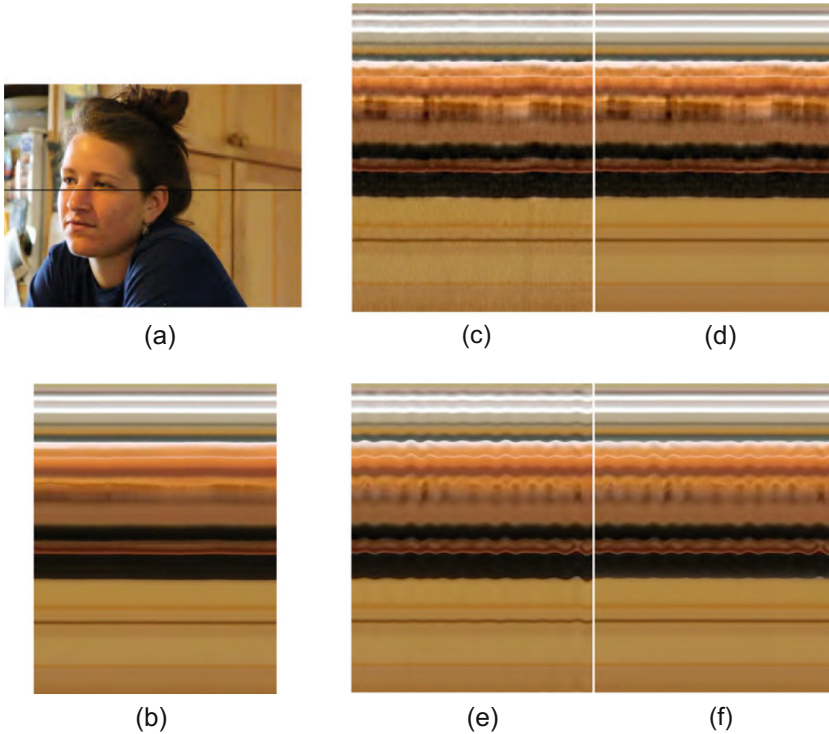


**Fig. 5.** (a) First frame of bottle sequence with highlighted black line, (b) space-time plot of original video of highlighted black line, (c) motion magnified plot using [5], and (d) motion magnified plot using the proposed method. (Color figure online)



**Fig. 6.** (a) First frame of baby video, (b) space-time plot of original video, (c) motion magnified plot using [2], (d) motion magnified plot using the proposed method with linear motion magnification, (e) motion magnification using [5], and (f) motion magnified plot using the proposed method with phase based motion magnification.

In Fig. 5, the results of the phase based method are shown on a video sequence of a juice bottle, where a bottle of juice is still and the background has large motions over different times. As shown in Fig. 5(a), the bottle is still all over the period (yellow stripe) and after a few frames there is motion in the video. In the



**Fig. 7.** (a) First frame of woman video, (b) space-time plot of original video, (c) motion magnified plot using [2], (d) motion magnified plot using the proposed method with linear motion magnification, (e) motion magnification using [5], and (f) motion magnified plot using the proposed method with phase based motion magnification.

magnified video using [5], the large motion is magnified along with the bottle's invisible motion and also, background noise is increased due to unnecessary magnification. On the contrary, the proposed method detected the salient object (bottle) and magnified the desired motion.

In Fig. 6, comparisons of both the linear and the phased based methods with the proposed methods of linear and phase based framework are shown. In Fig. 6(a), the first frame of the baby sequence is shown with highlighted black line to show the variation over time. In Fig. 6(b), the space-time plot of the original frame is shown that represents almost no change over time. In Figs. 6(c) and 6(d), plots of motion magnified videos are shown of [2] and the proposed framework with linear motion magnification. It is clearly visible that the proposed method has overcome the noise and magnifies only the breath of the baby. On the other hand, phase based method produces more significant amplification and the proposed phase based method removes the other outliers except the breath as shown in Figs. 6(e) and (f).

Again in Fig. 7, space-time plot of for both linear and phase-based motion magnification are shown. Similarly as Fig. 6, in this example also first frame of

woman video is shown with a black line (Fig. 7(a)). In Fig. 7(b), variation of spatial intensities are shown over time, where movement is negligible. In Fig. 7(c–f), motion magnified space-time plots are illustrated. Linear motion magnification using [2] and the proposed saliency method are shown in Figs. 7(c) and 7(d) respectively. As compare to [2], the proposed method avoid unnecessary motion magnification in background of woman. Similarly, motion magnification using [5] and the proposed method (using phase based) are shown in Figs. 7(e) and 7(f). Phase-based method produce less noise as compare to linear motion magnification. However, the proposed saliency method helps in removing noise because of outliers.

## 5 Conclusion

In this paper, we have proposed an efficient method to automatically detect the ROI and apply magnification selectively. We combine object detection techniques with EVM and presented a model that effectively amplifies only the motions of the salient object, thus eliminating background noise and making the result more significant to video magnification applications. We would like to improve this work by processing foregrounds such as transparent objects for which matting is not directly applicable.

**Acknowledgement.** The authors would like to thank SERB-DST for support through Young Scientists Startup Research Grant.

## References

1. Liu, C., Torralba, A., Freeman, W.T., Durand, F., Adelson, E.H.: Motion magnification. *ACM Trans. Graph.* **24**(3), 519–526 (2005)
2. Wu, H.Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., Freeman, W.: Eulerian video magnification for revealing subtle changes in the world. *ACM Trans. Graph.* **31**(4), 1–8 (2012)
3. Peng, H., Li, B., Ling, H., Hu, W., Xiong, W., Maybank, S.J.: Salient object detection via structured matrix decomposition. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(4), 818–832 (2017)
4. Wang, J., Drucker, S.M., Agrawala, M., Cohen, M.F.: The cartoon animation filter. *ACM Trans. Graph.* **25**(3), 1169–1173 (2006)
5. Wadhwa, N., Rubinstein, M., Durand, F., Freeman, W.T.: Phase-based video motion processing. *ACM Trans. Graph.* **32**(4), 80 (2013)
6. Wadhwa, N., Rubinstein, M., Durand, F., Freeman, W.T.: Riesz pyramids for fast phase-based video magnification. In: *Proceedings of IEEE International Conference on Computational Photography*, pp. 1–10 (2014)
7. Balakrishnan, G., Durand, F., Guttag, J.: Detecting pulse from head motions in video. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3430–3437 (2013)
8. Hong, K.: Classification of emotional stress and physical stress using facial imaging features. *J. Opt. Technol.* **83**(8), 508–512 (2016)

9. Bharadwaj, S., Dhamecha, T.I., Vatsa, M., Singh, R.: Computationally efficient face spoofing detection with motion magnification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 105–110 (2013)
10. Park, S.Y., Lee, S.H., Ro, Y.M.: Subtle facial expression recognition using adaptive magnification of discriminative facial motion. In: Proceedings of the 23rd ACM International Conference on Multimedia, pp. 911–914. ACM (2015)
11. He, X., Goubran, R.A., Liu, X.P.: Using Eulerian video magnification framework to measure pulse transit time. In: 2014 IEEE International Symposium on Medical Measurements and Applications (MeMeA), pp. 1–4. IEEE (2014)
12. Raghavendra, R., Avinash, M., Marcel, S., Busch, C.: Finger vein liveness detection using motion magnification. In: 2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS), pp. 1–7. IEEE (2015)
13. Elgharib, M., Hefeeda, M., Durand, F., Freeman, W.T.: Video magnification in presence of large motions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4119–4127 (2015)
14. Kooij, J.F.P., van Gemert, J.C.: Depth-aware motion magnification. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9912, pp. 467–482. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46484-8\\_28](https://doi.org/10.1007/978-3-319-46484-8_28)
15. Zhang, Y., Pintea, S.L., van Gemert, J.C.: Video acceleration magnification. arXiv preprint [arXiv:1704.04186](https://arxiv.org/abs/1704.04186) (2017)
16. Sonane, B., Ramakrishnan, S., Raman, S.: Automatic video matting through scribble propagation. In: Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing, vol. 87, no. (1–87), p. 8 (2016)
17. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: SLIC superpixels. Technical report (2010)