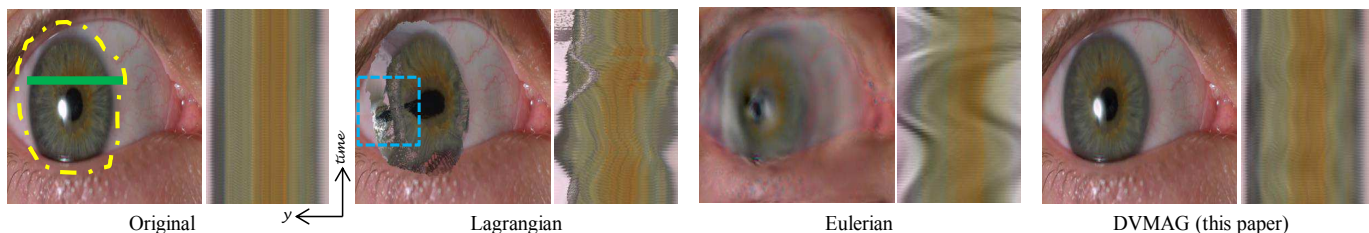


# Video Magnification in Presence of Large Motions\*

Mohamed A. Elgharib<sup>1</sup>, Mohamed Hefeeda<sup>1</sup>, Frédo Durand<sup>2</sup>, and William T. Freeman<sup>2</sup>

<sup>1</sup>Qatar Computing Research Institute    <sup>2</sup>MIT CSAIL



**Figure 1:** The Eye sequence and its magnification with (from left): the Lagrangian approach [7], the Eulerian approach [14] and DVMAG (ours). The Region of Interest is the dashed yellow region (left). For each magnification we show the spatio-temporal slice for the green line (left). For easier comparison all slices are temporally stabilized. This sequence shows an eye moving along the horizontal direction. Processing the sequence with DVMAG shows that the Iris wobbles as the eye moves (see DVMAG, spatio-temporal slice). Such wobbling is too small to be observed in the original sequence (left). The global motion of the eye causes significant blurring artifacts when processed with the Eulerian approach. The Lagrangian approach sensitivity to motion errors generates noisy magnification (see dashed blue).

## Abstract

*Video magnification reveals subtle variations that would be otherwise invisible to the naked eye. Current techniques require all motion in the video to be very small, which is unfortunately not always the case. Tiny yet meaningful motions are often combined with larger motions, such as the small vibrations of a gate as it rotates, or the microsaccades in a moving eye. We present a layer-based video magnification approach that can amplify small motions within large ones. An examined region/layer is temporally aligned and subtle variations are magnified. Matting is used to magnify only region of interest while maintaining integrity of nearby sites. Results show handling larger motions, larger amplification factors and significant reduction in artifacts over state of the art.*

## 1. Introduction

The world is full of small temporal variations that are hard to see with naked eyes. Variations in skin color occur as blood circulates [11], structures sway imperceptibly in the wind [11], and human heads wobble with each heart beat [2]. While usually too small to notice, such variations can

be magnified computationally to reveal a fascinating and meaningful world of small motions [7, 2, 17, 11, 14, 15]. Current video magnification approaches assume that the objects of interest have very small motion. However, many interesting deformations occur within or because of larger motion. For example, our skin deforms subtly when we make large body motion. A toll gate that closes exhibits tiny vibrations in addition to the large rotational motion. And microsaccades are often combined with large-scale eye movements (Fig. 1). Furthermore, videos or objects might be shot by handheld cameras and may not be perfectly still, and a standard video magnification technique will amplify handshake in addition to the motion of interest. When applied to videos that contain large motions, current magnification techniques result in large artifacts such as haloes or ripples, and the small motion remains hard to see because it is overshadowed by the then magnified large motion and its artifacts (see Fig. 1).

In the special case of camera motion, it might be possible to apply video stabilization as a preprocess to remove the undesirable large handshake, e.g. [8, 3, 9, 10], before magnification. However, this approach does not work for general object motion, and even in the case of camera shake, one has to be careful because any error in video stabilization will be amplified by the video magnification step. This problem is especially challenging at the boundary between

\*This research was supported in part by Qatar Computing Research Institute (QCRI).

a moving object, such as an arm, and its background, where multiple motions are present: the large motion, the subtle deformation to be amplified, and the background motion. Current video magnification such as the linear Eulerian [17], phase-based [14], and Riesz [15] algorithms assume that there is a locally single motion. This generates a background dragging effect around object boundaries. In addition the Lagrangian approach [7] is sensitive to motion errors. This generates noisy magnifications.

This paper presents a video magnification technique capable of handling small motions within large ones. Our technique is called DVMAG short for Dynamic Video Motion Magnification. Users select a region of interest (ROI) to magnify. We discount the global motion of the ROI through regularized parametric motion models. We handle boundaries between the moving object and the background using layer decomposition and matting. We show that, by applying Eulerian video magnification to both the foreground layer and its matte, we can dramatically reduce artifacts. Further, we handle potential dis-occlusion due to the amplified motion using texture synthesis [5, 7].

Aspects of novelty in this paper include: 1) Handling large motions in magnification through stabilization, 2) The use of mattes in magnification to maintain integrity of nearby sites, and 3) The use of texture synthesis in Eulerian magnification to fill-in revealed holes.

## 2. State of the Art

Video magnification is the task of amplifying and visualizing subtle variations in image sequences. Current techniques are classified into two main categories: Lagrangian [7, 2] and Eulerian [17, 14, 11]. In Lagrangian approaches motions are estimated explicitly. Here motions are the subtle variations to be magnified. The Eulerian approaches, on the other hand, do not estimate motions explicitly. Instead, they estimate subtle variations by calculating non-motion compensated frame differences. Lagrangian approaches can only magnify motion changes [7, 2], while Eulerian can magnify motion [14, 11] as well as color changes [17].

Liu et al. [7] presented one of the first video magnification techniques. Feature point trajectories are extracted and segmented into two sets, stationary and moving. An affine motion model is fitted on the stationary points which registers the examined sequence on a reference frame. Motions are re-estimated, scaled and added back to the registered sequence. This generates the magnified output.

Few techniques for Eulerian video magnification have been presented in [17, 14, 15]. In [17], an input sequence is first decomposed into a multiscale stack (Laplacian or Gaussian). Subtle variations are temporally filtered. When scaled and added back to the input sequence, a magnified output is rendered. Impressive results are generated [17], however only small motions and amplification factors can

be handled. Larger values generate artifacts in form of clipping which can kill useful information in the magnified video (see Fig. 6, Eulerian). The second Eulerian magnification technique is based on the observation that phase variations over time correspond to motion [14, 15]. An input video is decomposed into a multi-scale, multi-orientation stack [13]. The amplitude and phase of each band are separated, and the phase is temporally filtered at each location, orientation and scale. This estimates the subtle temporal changes in an image sequence. The corresponding phase changes are magnified and added back to the input video. Reconstructing the space-time representation renders the magnified video. The phase-based technique [14, 15] has better noise handling characteristics than the linear technique [17]. However its main drawback is still the inability to handle large motions. Such motions when processed generate significant blurring artifacts (see Fig. 1, Eulerian).

Our work is related to Bai et al. selective deanimation [1]. A user selects a region to be stabilized. However unnatural motions could be introduced to previously motion-less regions. A graph-cut based optimization is used to composite the immobilized region with still frames. The stabilization stage is related to ours. However we address a completely different problem, video magnification, with different form of artifacts that require different treatment.

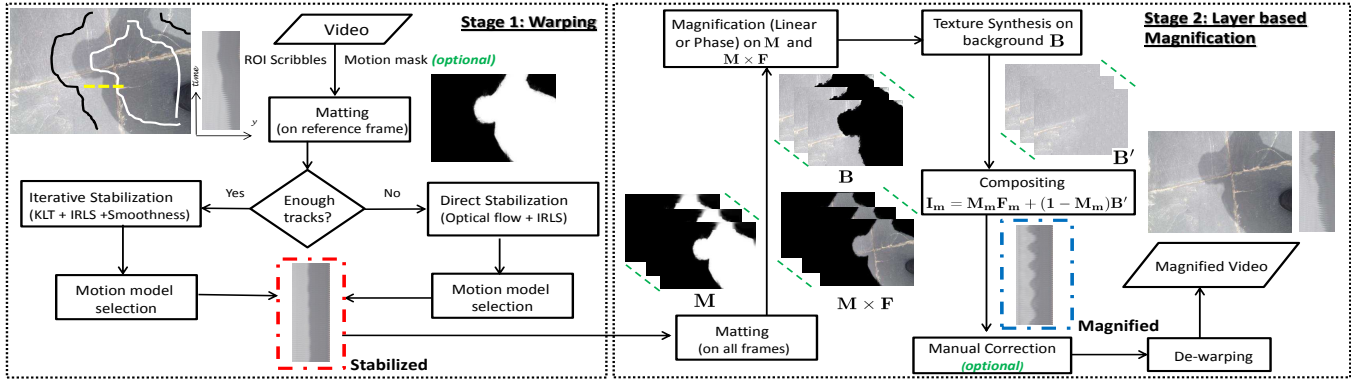
## 3. DVMAG: Dynamic Video Magnification

We present a video magnification technique to amplify small motions within large ones. Our technique (Fig. 2) has two main components: 1. Warping to discount large motion and 2. Layer-based Magnification. The Warping stage seeks to remove large motion while preserving small ones, and without introducing artifact that could be magnified. For this, we use either KLT tracking [12] or Optical Flow [6] as well as regularized low-order parametric models for the large-scale motion. Our layer-based magnification is based on decomposing an image into a foreground, background through an alpha matte. We magnify each layer and generate a magnified sequence through matte inversion. We use texture synthesis to fill in image holes revealed by the magnified motion. Finally, we de-warp the magnified sequence back to the original space-time co-ordinates. Users specify the region to be magnified using scribbles on a reference frame (Fig. 2, top-left).

### 3.1. Warping

Given an input sequence  $I$ , we want to estimate a stabilized sequence  $I^S$  by temporally registering it over a reference frame  $r$ . We model the large scale motion in the ROI with low-order models  $\Phi$  (either affine or translation-only) to preserve the small-scale motion to be magnified:

$$I^S(\mathbf{x}, t) = I(\Phi_{r,t}(\mathbf{x}), t), \quad (1)$$



**Figure 2:** Components of DVMAG. White and black strokes denote positive and negative samples for Region of Interest (top-left). We show the spatio-temporal slice of the dashed yellow line (top-left). In this sequence the shadow of the camera-man appears on the ground. The camera-man has his left hand on his shoulder and he is periodically moving it up and down. This motion however is too small to be observed in the original sequence (dashed-red, bottom left). Our technique has 2 main stages. The first temporally aligns the ROI while the second does the actual magnification (dashed-blue, bottom right). We remove potential magnification artifacts through texture synthesis and manual intervention (if required). Finally we bring the video back to the original space-time co-ordinates. In this sequence the motion mask is the entire frame. Note that the motion mask and the matting strokes are always assumed to be static in the stabilized sequence.

where  $\mathbf{x}$  denotes 2D pixel co-ordinates. Given a set of points  $X_t$  in frame  $t$  and their corresponding  $X_r$ , we find  $\Phi_{r,t}$  by minimizing  $\|\Phi_{r,t}(X_t) - X_r\|^2$ . We propose two methods for generating  $X_t$  and its corresponding  $X_r$ : one using KLT tracks [12] and the other using optical flow [6].

**Estimating  $\Phi_{r,t}$  for KLT Tracks:** For a set of points in frame  $t$  and their correspondence in the next frame, we attempt to match them by fitting either an affine or translation model. We impose temporal smoothness on  $\Phi$  using a Moving Average Filter. We use a local window of 5 frames centered on the examined frame  $t$  and weighted by  $W \sim \mathcal{N}(0, 4)$ . To reduce fitting errors we use Iterative Reweighted Least Squares (IRLS) for solving  $\Phi$ . Weights here are set inversely proportional to  $\|\Phi_{r,t}(\mathbf{x}) - \mathbf{x}\|^2$ . We use a temporally iterative scheme to estimate  $\Phi$  between an examined frame  $t$  and the reference  $r$ . First we generate an estimate of  $\Phi$  for each pair of consecutive frames. For instance if  $r > t$  we estimate  $\Phi_{t+1,t}, \Phi_{t+2,t+1}, \Phi_{t+3,t+2}, \dots, \Phi_{r,r-1}$ . Hence the direct transformation from  $t$  to  $r$  becomes  $\Phi_{r,t} = \prod_{u=t}^{u=r-1} \Phi_{u+1,u}$ . Given  $\Phi_{r,t}$  we stabilize frame  $t$  by applying Eq. 1. For  $r < t$ , we do the same process but in the opposite time direction. This stabilizes the entire sequence  $I$  over the reference frame  $r$ .

**Estimating  $\Phi_{r,t}$  for Optical Flow:** As optical flow is more sensitive to motion errors than feature point trajectories, one main adjustment is necessary: We do not use a temporally iterative scheme for estimating  $\Phi_{r,t}$  as errors could pile up. Instead we directly estimate  $\Phi$  between the examined frame  $t$  and the reference frame  $r$ , i.e., in one shot. Here optical flow is estimated between the reference  $r$  and all frames of the examined sequence using [6]. With

this in consideration we proceed with estimating the model parameters as for feature point trajectories.

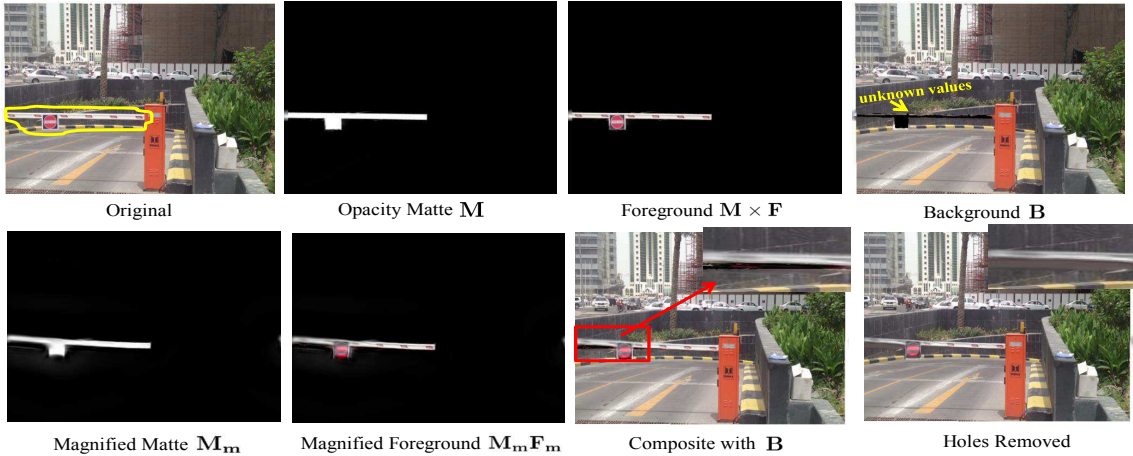
**Optical Flow vs. Feature Point Trajectories:** It is important to base the motion modeling of the warping stage on good motion candidates. This will reduce the risk of magnifying stabilization errors later on. Hence, we choose between two motion generation methods, KLT tracks [12] and optical flow [6]. We only consider motion candidates inside a motion mask. We compute KLT tracks, and if the number of tracks as a percentage of the number of motion mask pixels is greater than some threshold we use KLT tracks. Else we use optical flow estimates. We set that threshold to 5%.

**Affine vs. Translation Modeling:** We estimate affine and translational models for  $\Phi$  and we pick the model with least stabilization error. That is the one minimizing  $\sum_{t=1}^T \sum_{\mathbf{x}=1}^P |I^S(\mathbf{x}, t) - I(\mathbf{x}, r)|$ . Here  $I(\mathbf{x}, r)$  is the reference frame,  $T$  is the number of frames in the stabilized sequence  $I^S$  and  $P$  is the number of pixels in one frame. We carry this calculation over the ROI only.

### 3.2. Layer-Based Eulerian Magnification

We present a layer-based approach for video magnification. Given a region of interest, we decompose an image into three layers: 1. Opacity matte, 2. Foreground and 3. Background. We use Levin et al. matting [4]. We magnify the opacity and foreground using the Eulerian approach [17, 14]. If we are interested in magnifying temporal color changes we use the Linear technique [17]. Otherwise we use the phase-based technique [14]. We place the magnified foreground over the original background to reconstruct the remaining unmagnified sites. We use texture synthesis to fill in image holes revealed by the amplified foreground motion.





**Figure 3:** Illustrating our layer-based magnification. Here we seek to magnify the subtle motions in the parking gate (see yellow). We estimate opacity, foreground and background of the gate (top row). Eulerian magnification [14] amplifies the matte and foreground (bottom row). Direct compositing over  $B$  generates image holes revealed by the magnified motion (bottom row, red region). Hence we inpaint the unknown sites of  $B$  (top row, yellow text) and we composite over the inpainted  $B$  instead. This removes image holes (bottom row, right).

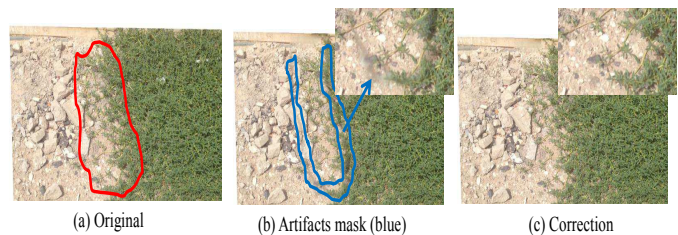
We remove remaining compositing artifacts through manual correction, if necessary. Finally, we de-warp the magnified sequence back to the original space-time co-ordinates. All steps are performed on every video frame.

Fig. 3 illustrates our approach in more detail where we magnify the motion of the parking gate (see yellow). Note that matting techniques [4] generate temporally inconsistent foreground values outside the ROI. Directly magnifying such values generates strong artifacts (see Fig. 4, left). Hence, we always magnify the truncated foreground  $M \times F$  instead of just  $F$  (see Fig. 4, right). For simplicity, we denote the magnified  $M \times F$  as  $M_m \times F_m$ , where  $M_m$  is the magnified matte. Given  $M_m$  and  $M_m \times F_m$ , we generate a composite sequence by placing the magnified foreground  $M_m \times F_m$  over the original background  $B$ . Nevertheless, directly using  $B$  would generate image holes in sites revealed by the magnified motion (see Fig. 3, bottom, red region). Hence, prior to image compositing we fill in the unknown background values (see Fig. 3, top, yellow text) through texture synthesis [5]. Given the new inpainted background  $B'$ , we calculate the magnified sequence  $I_m(\mathbf{x}) = M_m(\mathbf{x})F_m(\mathbf{x}) + (1 - M_m(\mathbf{x}))B'(\mathbf{x})$ , where  $\mathbf{x}$  denote all image pixels. We apply this process for each frame of the examined sequence. This generates a stabilized magnified sequence (see Fig. 3, last row, right).

Compositing artifacts can be generated in sites where foreground and background estimates are similar (see Fig. 5 (b), inset). To fully remove such artifacts, we give the user the option to selectively inpaint specific regions. The user selects the corrupted sites in only the reference frame (Fig. 5 (b) blue mask). The entire video is then corrected by filling the corrupted sites with original sequence values (Fig. 5 (c)). Out of nine examined sequences this manual correc-



**Figure 4:**  $F$  vs.  $M \times F$  magnification. Matting [4] generate temporally inconsistent foreground outside the examined object. This generates strong blurring artifacts when magnified (see left). Magnifying  $M \times F$  instead remove such artifacts (right).



**Figure 5:** Example of compositing artifacts (in blue). Region of Interest is shown in red. To remove such artifacts the user selects the corrupted sites in only the reference frame (using the blue mask). The entire video is then corrected by filling the corrupted sites with original values.

tion was required in only one sequence.

The last step of our algorithm de-warps the magnified composite sequence  $I_m$  back to the original space-time co-ordinates. In our implementation we use the previously saved motion parameters ( $\Phi_{r,t}$  in Sec. 3.1) to interpolate the de-warped sequence.

## 4. Results

We have performed experiments on real sequences as well as on synthetically-generated inputs with ground truth

Video	$\alpha$	$\omega_l$ (Hz)	$\omega_h$ (Hz)	$f_s$ (fps)
Eye	30	20	70	1000
Bulb	15	20	70	1000
Gun	20	8	33	480
Water	10	5	6	29
Shadow	50	2.8	3.2	29
Parking	40	4	5	29
Leaves	20	0.5	2	29
Sim1	0-120	4.9	5.1	24
Sim2	50	72	92	600

**Table 1:** Examined Sequences and values for: amplification factor  $\alpha$ , examined frequency spectrum  $\omega_l - \omega_h$  Hz and sampling frequency  $f_s$ . **Eye** and **Bulb** are from [www.youtube.com/user/theslowmoGuys](http://www.youtube.com/user/theslowmoGuys). **Gun** is from [www.guntalk.com](http://www.guntalk.com).

available. For real sequences we assess performance qualitatively. For controlled experiments we assess performance quantitatively against ground-truth. Results show that state of the art methods optimized for small motion generate magnification artifacts when handling large displacements. Our technique, DVMAG, significantly reduces artifacts and increases the domain of applicability. Table 1 lists the examined sequences with the corresponding parameters.

#### 4.1. State of the Art Techniques

We compare our technique (DVMAG) against two video magnification approaches, Lagrangian and Eulerian [17, 14]. For the Eulerian approach we use the Linear technique [17] for the **Bulb** sequence, while remaining sequences are processed with the phase technique [14]. We use the authors' implementation of both the phase and linear techniques. For the Lagrangian approach, the original implementation is not available. We first temporally stabilize the examined sequence using our stabilization (Sec. 3.1). We then estimate a dense motion field with Liu et al.'s optical flow [6]. Finally magnification is achieved by scaling the motion field.

To assess our stabilization (Sec. 3.1) we compare DVMAG against two further magnification approaches. Here we use two off-the-shelf stabilization techniques to stabilize the examined sequence. We then magnify the stabilized sequence using the Eulerian approach [17, 14]. We examine Youtube [3] and Adobe stabilization [9]. The former is available through Youtube video manager while in the latter we use After Effects Warp Stabilizer VFX. After Effects allows the user to define a motion mask. Hence for fair comparison After Effects results use the same motion mask of DVMAG. We also compare against one motion compensation technique. We use Liu optical flow [6] to generate motion estimates between each frame and the reference. For each frame we move its pixels using the dense optical flow estimates. This generates a prediction of the examined

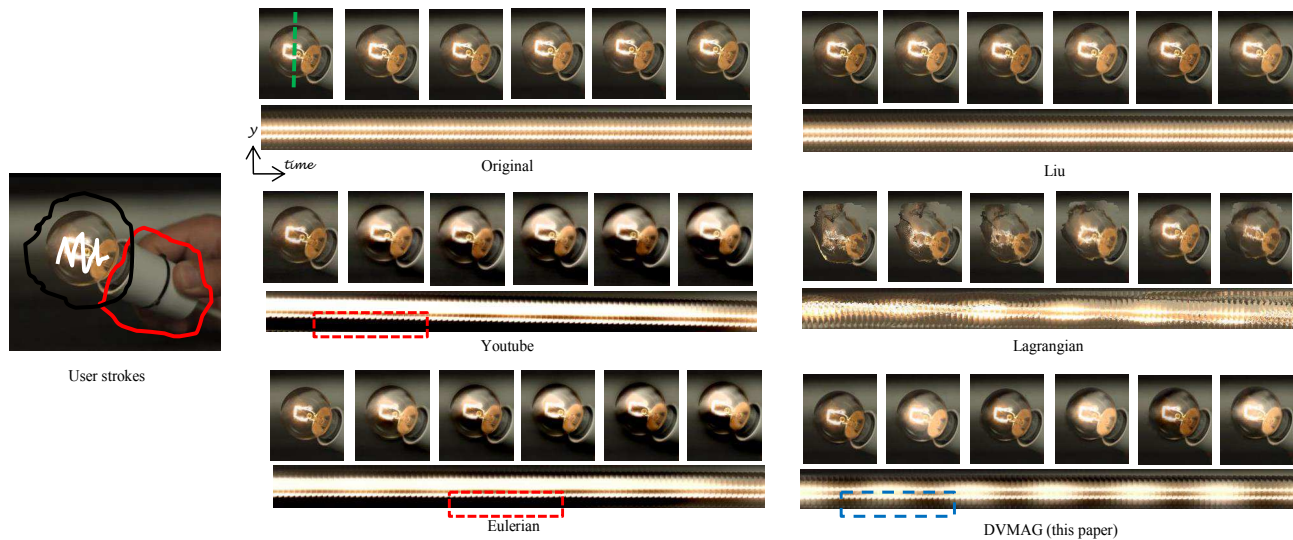
frame as seen by the reference. We finally proceed with Eulerian magnification as in Youtube and Adobe.

For Lagrangian, After Effects and Liu [6] we only magnify the Region of Interest. For the remaining techniques we magnify the entire frame as in many cases the ROI is moving, even after stabilization. For easier assessment all results are temporally stabilized over the reference frame of DVMAG. Eulerian magnifications are temporally stabilized using our technique (Sec. 3.1). The remaining techniques by definition should be stabilized prior to magnification.

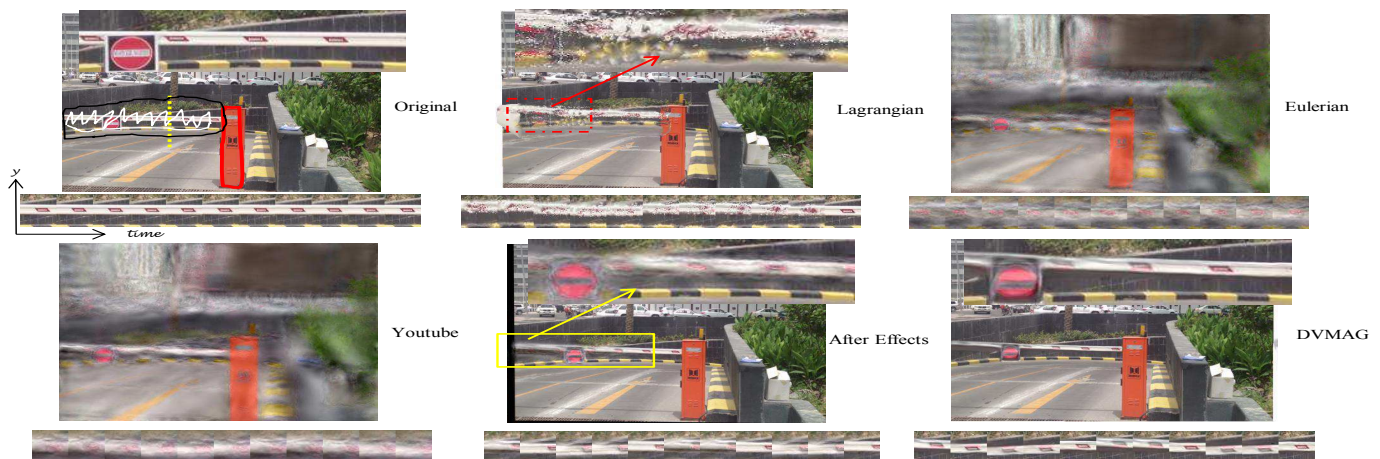
#### 4.2. Real Sequences

Fig. 6-8 show DVMAG magnification for some sequences and their comparison against different techniques. For each technique we show one frame and a spatio-temporal slice from the magnified sequence. In Fig. 6 we examine the **Bulb** sequence. In this sequence a person is holding a bulb and moves it up in the vertical direction. Processing this sequence with DVMAG reveals a temporal variation in the light strength. This variation is caused by the alternating electrical current and is hardly noticeable with no magnification (see Fig. 6 original). Processing **Bulb** with Liu [6] does not reveal any temporal changes. This is because optical flow is estimated in a way to minimize temporal variations. The remaining techniques also do not reveal any useful temporal variations. The Eulerian approach generates color clipping artifacts (see dashed red). Such artifacts are due to filtering the temporal misalignments of the input frames. Similarly clipping artifacts are generated by Youtube stabilization errors (see dashed red). Finally the Lagrangian approach generates noisy results.

Fig. 7 shows the results generated for the **Parking** sequence. This video shows the entrance of an underground car parking. The opening and closing of the parking white gate causes the gate to vibrate. Such vibration is too small to be observed from the original sequence (see Fig. 7 Original). Processing **Parking** with DVMAG magnifies the gate vibration. Our layered-based magnification maintains the integrity of the rest of the sequence (see inset). After Effects does magnify the vibration however it generates more blurred results than DVMAG (see yellow region). In addition it corrupts sites around the gate boundaries. However DVMAG maintains the integrity of such sites through matting and texture synthesis. The parking vibration is not magnified by any other technique. Eulerian and Youtube generate blurring artifacts while Lagrangian generates noisy results (see dashed red). Fig. 8 processes the **Gun** sequence with different magnification techniques. This sequence shows a person firing a gun. We show the spatio-temporal slice of the green dashed line as the gun is fired (see top, left). Examining the original sequence shows that the shooter hand is static while taking the shot. Magnifying the sequence with DVMAG shows that the arm moves as the



**Figure 6:** Original sequence (top) and magnification using (in clock-wise direction): Liu [7], Lagrangian, DVMAG, Eulerian and Youtube. In this sequence a person is holding a bulb and moves it up in the vertical direction. The ROI strokes are shown in white and black while the motion mask is shown in solid red (left-most column). For each magnification we show the spatio-temporal slice for the dashed green line (original, top-left). Our approach DVMAG reveals the temporal light changes in the bulb caused by the alternating electrical current. The remaining techniques do not reveal any temporal variations. Eulerian and Youtube generate clipping artifacts that manifest as sharp transition in color (see dashed red, compare with dashed blue). The Lagrangian approach generates noisy magnification.



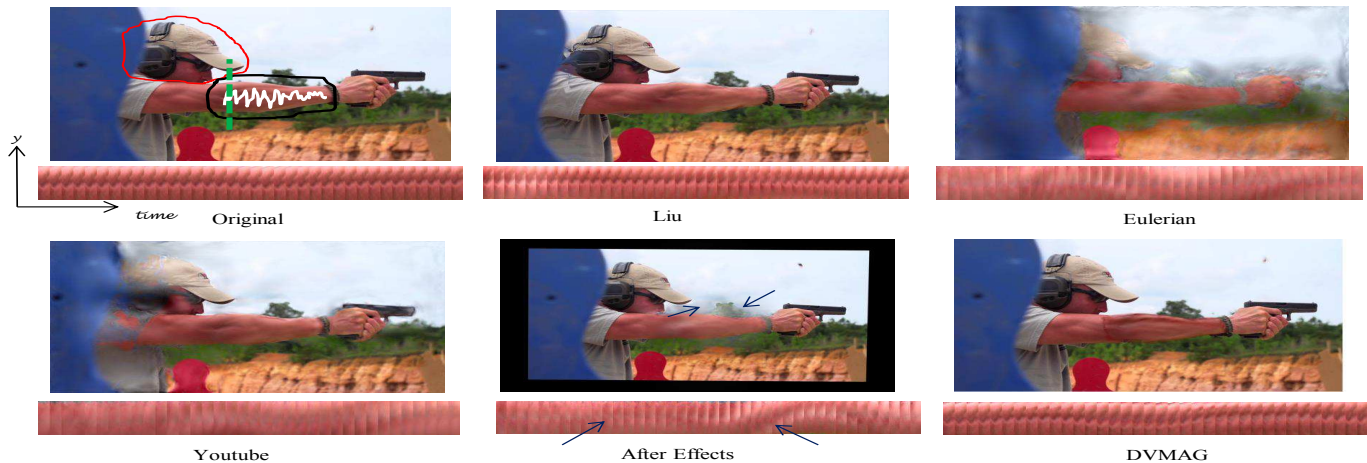
**Figure 7:** Original sequence (top) and magnification using (in clock-wise direction): Lagrangian, Eulerian, DVMAG, After Effects and Youtube. The ROI strokes are shown in white and black while the motion mask is shown in solid red (top-left). For each processing we show the spatio-temporal slice for the dashed yellow line (top-left). DVMAG reveals the vibrations in the white parking gate and maintains integrity of nearby data. Eulerian and Youtube generate significant blurred results and Lagrangian generates noisy magnification (see dashed red). After-Effects reveals the gate vibration however it generates more blurred results than DVMAG (see yellow region). In addition it corrupts sites nearby the gate boundaries. Thanks to Matting and texture synthesis DVMAG maintains integrity of such sites.

shot is taken. Eulerian, Youtube and After-Effects generate blurred results. Liu [6] does not reveal the arm movement.

In summary, our approach magnifies regions of interest, maintains integrity of nearby sites and outperforms all other techniques. The Eulerian approach generates blurry magnifications. Youtube stabilization can not remove large motions and hence also generates blurry results. Multiple mov-

ing objects generate stabilization errors in After Effects. These errors are magnified. In addition After Effects usually corrupts sites around the examined object boundaries. Lagrangian is sensitive to motion errors and hence generates noisy results. Finally direct motion compensation (Liu [6]) hardly amplifies temporal variations. This is because optical flow is estimated in way to minimize temporal changes.





**Figure 8:** Original sequence (top) and magnification using (in clock-wise direction): Liu et al. [6], Eulerian, DVMAG, After Effects and Youtube. The ROI strokes are shown in white and black while the motion mask is shown in solid red (top-left). For each processing we show the spatio-temporal slice for the dashed green line (top-left). In this sequence a person is firing a gun which causes his arm to slightly move up and down. This arm movement however is not observed in the original sequence (see top-left). DVMAG magnifies the arm movement (see the spatio-temporal slice) while the remaining techniques generate blurring artifacts (see blue arrows for After Effects).

### 4.3. Controlled Experiments

**Sim1:** We create a reference frame containing a white circle and a red rectangle (see Fig. 9, left). The white circle is the region of interest (ROI) to be magnified, while the red rectangle is used to generate motion candidates. We define a local motion as  $d_j = A \sin(2\pi \frac{f}{f_s} j)$  where  $A = 0.25$  pixels,  $f = 5$  cycle/frame and  $f_s = 24$  frame/second. We generate frame  $j$  of **Sim1** by shifting the white circle with  $d_j$  along the horizontal direction. Doing that for 200 frames generates a sequence with the white circle vibrating. Here  $(A, f)$  are the amplitude and frequency of vibration respectively. We then add a large global motion to the vibrating sequence by shifting each frame by  $\Delta_j = A \sin(2\pi \frac{f}{f_s} j)$ .  $\Delta_j$  is the global motion at frame  $j$  where  $A = 40$  and  $f = 0.1$ . The global motion only occurs along the horizontal direction. The final generated sequence is **Sim1**.

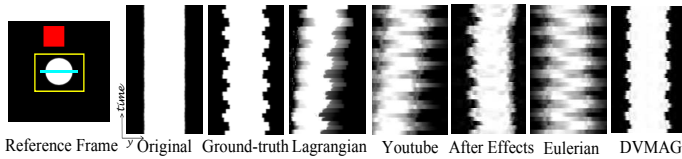
We process **Sim1** using different magnification techniques. The aim is to assess the ability of magnifying the vibration of the white circle. We examine different amplification factors  $\alpha$  and we compare against ground-truth. Ground-truth is generated using the same method of generating **Sim1**. For an amplification factor  $\alpha$  the corresponding ground-truth is calculated by shifting the white circle of the reference frame by  $d_j = A(\alpha + 1) \sin(2\pi \frac{f}{f_s} j)$ . As in **Sim1**  $j$  indexes the frames,  $A = 0.25$  pixels,  $f = 5$  cycle/frame and  $f_s = 24$  frame/second. For ground-truth we do not add a global motion as all comparisons are done against a temporally aligned version of the generated magnifications.

Fig. 9 and Fig. 10 (left) shows the magnifications of **Sim1** as generated by different techniques. Here we use an amplification factor of  $\alpha = 20$  and we examine vibra-

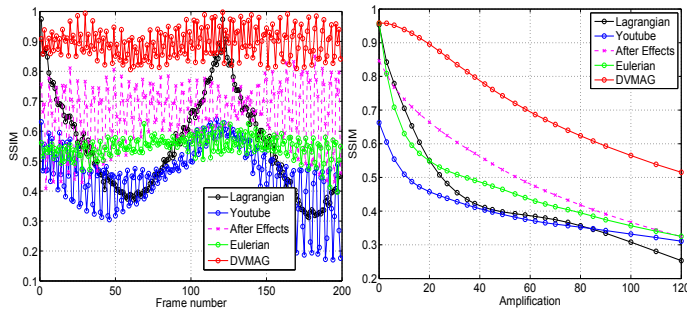
tions in the range of 4.9 – 5 Hz. Motion candidates are generated using KLT tracks [12] on the red rectangle. In Fig. 9 we show the spatio-temporal slice for the blue line (see reference frame). In Fig. 10 (left) we show SSIM [16] with ground-truth for each examined frame.  $SSIM(I_1, I_2)$  measures the structure similarity between the two images  $I_1$  and  $I_2$ . Here  $SSIM = 1$  denotes exact ground-truth similarity and a value of 0 denote no similarity at all. Note that in **Sim1** only sites inside the yellow rectangle of Fig. 9 are taken into consideration while estimating SSIM.

Fig. 9 shows that DVMAG best resembles ground-truth. Eulerian and Youtube generate significant blurring artifacts, while Lagrangian is sensitive to motion errors. After Effects generate poor stabilization due to the absence of enough long feature point trajectories. Fig. 10 (left) shows that the Lagrangian error follows the same profile of the global motion. The error increases as the magnitude of the global motion increases. In addition it reaches its minimum at frame 0 and 120 as global motion is minimal. This shows that Lagrangian magnification is sensitive to motion estimation errors. Examining the remaining techniques in Fig. 10 (left) shows that DVMAG outperforms all other approaches.

Fig. 10 (right) shows how **Sim1** behaves with different amplifications  $\alpha$ . Here at each  $\alpha$  we estimate the mean SSIM of the entire magnified sequence against ground-truth. Fig. 10 (right) shows that DVMAG can handle larger amplifications with less errors over all other techniques. For instance errors in DVMAG with  $\alpha = 20$  are almost equivalent to errors in Eulerian, After Effects and Lagrangian with  $\alpha = 1$ . In addition Fig. 10 (right) shows that DVMAG has the slowest rate of degradation among most techniques. For instance in the range  $\alpha = 0 - 40$  the slopes of Youtube,



**Figure 9:** Ground-truth verification of DVMAG compared against other techniques. Left: A frame from **Sim1**, where we generate motion candidates from the red rectangle movement. Sites inside the yellow rectangle are the only sites used in the quantitative assessment of Fig. 10. Right: Spatio-temporal slices for the blue line (see left) for different magnification techniques. DVMAG best resembles ground-truth and does not generate blurring artifacts as other techniques.

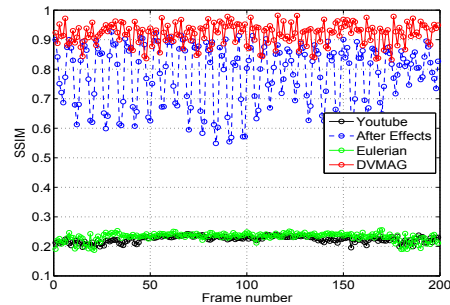
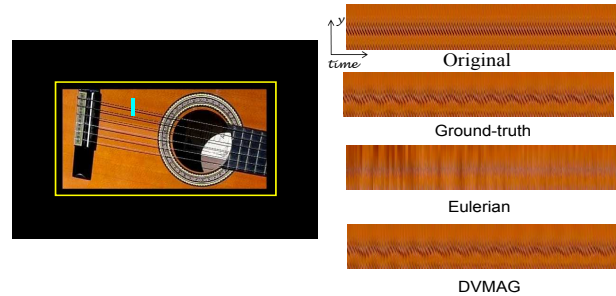


**Figure 10:** Left: SSIM with ground-truth for each frame of **Sim1**. The larger the SSIM the better. DVMAG outperforms all examined techniques. Right: SSIM with ground-truth over different amplification factors. Our approach handles large amplifications with less magnification artifacts over all other techniques. It also has the lowest degradation rate as a function of amplification.

Eulerian and Lagrangian are steeper than DVMAG. This shows that DVMAG is more robust to magnification artifacts over all other approaches.

**Sim2:** We explore how large motion degrades magnification and how such degradation is handled by DVMAG. We examine the **Guitar** sequence from Wu et al. [17]. This sequence does not have any large motion, only small motion due to the subtle movement of guitar strings. We use the phase technique [14] to magnify the Low E note. We treat this magnification as the ground-truth. We then added a large global motion in the same way global motion was added to **Sim1**. Here we use  $A = 50$ ,  $f = 0.2$  both in the horizontal and vertical directions. The generated sequence is **Sim2**. We process **Sim2** with the Eulerian approach, After Effects and DVMAG. We set  $\alpha = 50$  and we examine the frequency spectrum 72–92 Hz. For DVMAG we generate motion candidates with Liu et al. [6] optical flow. Here the entire frame is treated as the motion mask.

Fig. 11 (first row) shows an original frame from **Sim2** and its magnification with different techniques. We show the spatio-temporal slices of the blue line (see top, left). The string vibration due to Low E note is evident in the ground-



**Figure 11:** A frame from **Sim2** (top-left) and its magnification using different techniques (top-right). Here we zoom on the spatio-temporal slice of the blue line (see top-left). DVMAG outperforms Eulerian in resembling the ground-truth. Bottom: SSIM with ground-truth for different techniques. Here SSIM is estimated for only sites included in the yellow rectangle (see top-left). SSIM shows that our approach outperforms all other techniques.

truth. The Eulerian approach generates significant blurring and does not reveal the string vibration. DVMAG correctly resembles the ground-truth and does not generate blurring artifacts. Fig. 11 (bottom) shows the SSIM of each magnified frame against ground-truth. Here we only consider sites included in the yellow rectangle (Fig. 11, see top-left). After Effects stabilization suffered from temporal shakiness due to erroneous tracks. This shakiness became more apparent after magnification. DVMAG outperformed all techniques. In addition it outperformed the Eulerian approach by a factor of around 200% (see Fig. 11 bottom).

## 5. Conclusion

We presented a video magnification approach for amplifying small motions within large ones. Current magnification techniques generate significant artifacts when large motions are present. Our approach is based on temporal stabilization followed by layer-based magnification. Matting is used to magnify only region of interest while maintaining integrity of nearby sites. Results show handling larger motions, larger amplification factors and significant reduction in artifacts over state of the art. Future work can address handling multiple different motions in an examined object.



## References

- [1] J. Bai, A. Agarwala, M. Agrawala, and R. Ramamoorthi. Selectively de-animating video. *ACM Trans. Graph.*, 31(4):66:1–66:10, 2012. [2](#)
- [2] G. Balakrishnan, F. Durand, and J. Guttag. Detecting pulse from head motions in video. In *CVPR*, pages 3430–3437, 2013. [1](#), [2](#)
- [3] M. Grundmann, V. Kwatra, and I. Essa. Auto-directed video stabilization with robust l1 optimal camera paths. In *CVPR*, pages 225–232, 2011. [1](#), [5](#)
- [4] A. Levin, D. Lischinski, and Y. Weiss. A closed-form solution to natural image matting. *IEEE Transactions on PAMI*, 30(2):228–242, 2008. [3](#), [4](#)
- [5] X. Li. Image recovery via hybrid sparse representations: A deterministic annealing approach. *IEEE Journal of Selected Topics in Signal Processing*, 5(5):953–962, 2011. [2](#), [4](#)
- [6] C. Liu. *Beyond Pixels: Exploring New Representations and Applications for Motion Analysis*. PhD thesis, Massachusetts Institute of Technology, 2009. [2](#), [3](#), [5](#), [6](#), [7](#), [8](#)
- [7] C. Liu, A. Torralba, W. T. Freeman, F. Durand, and E. H. Adelson. Motion magnification. *ACM Trans. Graph.*, 24(3):519–526, 2005. [1](#), [2](#), [6](#)
- [8] F. Liu, M. Gleicher, H. Jin, and A. Agarwala. Content-preserving warps for 3d video stabilization. *ACM Trans. Graph.*, 28(3):44:1–44:9, 2009. [1](#)
- [9] F. Liu, M. Gleicher, J. Wang, H. Jin, and A. Agarwala. Subspace video stabilization. *ACM Trans. Graph.*, 30(1):4:1–4:10, 2011. [1](#), [5](#)
- [10] S. Liu, L. Yuan, P. Tan, and J. Sun. Bundled camera paths for video stabilization. *ACM Trans. Graph.*, 32(4):78:1–78:10, 2013. [1](#)
- [11] M. Rubinstein, N. Wadhwa, F. Durand, W. Freeman, H. Eugene, and J. Guttag. Revealing invisible changes in the world. *Science*, 339(6119):518–519, February 2013. [1](#), [2](#)
- [12] J. Shi and C. Tomasi. Good features to track. In *CVPR*, pages 593–600, 1994. [2](#), [3](#), [7](#)
- [13] E. Simoncelli and W. Freeman. The steerable pyramid: a flexible architecture for multi-scale derivative computation. In *International Conference on Image Processing (ICIP)*, volume 3, pages 444–447, 1995. [2](#)
- [14] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman. Phase-based video motion processing. *ACM Trans. Graph.*, 32(4), 2013. [1](#), [2](#), [3](#), [4](#), [5](#), [8](#)
- [15] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman. Riesz pyramids for fast phase-based video magnification. In *International Conference on Computational Photography*, 2014. [1](#), [2](#)
- [16] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. [7](#)
- [17] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. T. Freeman. Eulerian video magnification for revealing subtle changes in the world. *ACM Trans. Graph.*, 31(4):65:1–65:8, 2012. [1](#), [2](#), [3](#), [5](#), [8](#)