

# **Video Motion Magnification using Spatio-Temporal Algorithm**

Lokesh Sarode  
M.E. Scholar  
G.H.R.C.E.M., Amravati

N.N.Mandaogade  
H.O.D  
G.H.R.C.E.M., Amravati

## **ABSTRACT**

The aim is to disclose temporal variations in videos that are hard or nearly impossible to view with the naked eye and display them in an indicative fashion. Our method, which is based on spatio-temporal processing, takes a standard video sequence as input, and applies spatial processing / decomposition, followed by temporal processing / filtering to the frames where the frames are filtered to select desired band of frequencies to be amplified. These bands of frequencies are empirically decided by user and also user is allowed to decide amplification factor by which motion is to be amplified. After the processing and enhancement of the low frequencies, the spatial pyramid is reconstructed back to get improved video sequence where the subtle motions are significantly visible. Or in other words, the resulting signal is then amplified and added back to the original signal to reveal hidden information. Using this method, it is able to visualize the pulsation/movement of ULNAR artery when a blood flows though it and also to amplify and reveal subtle motions in day to day life.

## **Keywords**

Decomposition, subtle motion, spatial, temporal, spatial pyramid

## **1. INTRODUCTION**

The human visual system has limited spatio-temporal sensitivity, but many signals that fall below this capacity can be informative. For example, human wrist's blood vessels vary slightly with blood circulation. This variation, while invisible to the naked eye, can be exploited to extract pulse rate. Similarly, motion with low spatial amplitude, while hard or impossible for humans to see, can be magnified to reveal interesting mechanical behavior. Hence the efficient tools are needed in order to analyze this un-approached signals usually containing significant information. As these signals corresponds to the hidden motions in the video sequences, the analysis of motion makes us enable for extraction of visual information from the temporal and spatial changes which occurs in an image sequence, and is a fundamental task in computer vision and image processing. Assuming illumination conditions remain constant, image sequence changes because of a relative motion between the scene and the camera; either by the viewing camera moving relative to a static scene, elements of the scene being in motion, or in the general case, both camera and objects moving independently. The problem of motion analysis may be divided into two sub problems; that of feature correspondence and reconstruction. The correspondence problem concerns finding features' pair in two or more perspective views of a scene so that each pair corresponds to the same scene point. Because of this ill-posed and inherent combinatorial complexity nature, feature correspondence is claimed as the hardest and poor-level image analysis tasks. The solubility of the correspondence problems

also influenced by factors such as image noise, periodic textures and object occlusion. The reconstruction problem states that, given a number of corresponding elements, and possibly knowledge of the camera's intrinsic parameters, what may be inferred about the 3D motion and structure of the observed world. The extraction of motion information from an image sequence has many important applications. For example, changes in blood circulation causes slight changes in human skin color. This variation is invisible to the naked eye but it can be exploited to extract pulse rate of a person [2]. Similarly, it is very hard or nearly impossible for human eye to see the motion with low spatial amplitude, hence it should be magnified to reveal interesting mechanical behavior [3]. The success of these tools induces the development of new techniques to manifest inscrutable signals in videos. In this paper, the approach is presented in which an annexation of temporal and spatial processing of videos can magnify elusive variations that evince important phase of the world.

The concept is to contemplate the time series of pixel values at any spatial position i.e. pixel and raise the shift in a given temporal frequency band of interest. The spatio-temporal filtering presented in this approach is capable of revealing motions with low-amplitude. For example, it is feasible to increase the scrupulous motions around the chest of a breathing baby.

## **2. RELATED WORK**

In last decade, several methods have been developed to reveal unperceivable motions in videos. Some methods found out to be very effective and build the base for this area of work. [3] Presented a technique, motion magnification that analyzes and amplify motions that would otherwise be invisible or very difficult to see by naked eye. The input was a sequence of images from a stationary camera. The system automatically segments a reference frame into region of common property, grouped by proximity, similar color, and correlated motions. Analogous to focusing a microscope, the user identifies the segment to modify, and specifies the motion magnification factor. The video sequence was then re-rendered with the motions of the selected layer magnified as desired. The output sequence allows the user to see the form and characteristics of the magnified motions in an intuitive display, as if the physical movements themselves had been magnified, then recorded. [4] Presented the CAF (Cartoon Animation Filter), a normal filter that takes an licentious input motion signal and harmonize it in such a way that the output motion is more "animated" or "alive". The filter adds a salved, inverted, and may be time shifted version of the second derivative of the signal back into the prime (original) signal. Almost all the parameters of such filter are automated. Only the desired strength of the filter needs to set by the user. The effectiveness of the animation filter false in its simplicity & generality. The filter had been applied to motions ranging from hand striate trajectories, to simple animations within Power Point

presentations, to motion captured Depth of Field (DOF) curves, to video exfoliation results. Experimental results showed that the filtered motion exhibits expectancy, follow-through, hyperbole (exaggeration) and squash-and-stretch effects which are absent in the original input motion data. The approach is made to maintain a balance between simplicity and control that favored simplicity. The authors have found out that the application of the cartoon animation filter probably is not satisfactory for hand crafted off-line animation systems although it may be useful for previews. The author believe the value of such a simple approach will be in either real time applications such as games, or in less professional settings such as a child focused animation tool, or in 2D presentation systems such as PowerPoint.

The above two methods follow a Lagrangian approach, in accordance to fluid dynamics in which the trajectory of particles is tracked over time. As such, they believe in accurate motion calculation, which is reckoning expensive and hitch to make artifact-free, especially at regions of mazy motions and occlusion boundaries. However, [4] have shown that additional technologies, including motion fragmentation and image in-painting, are required to generate good quality compilation. But this increases the involution of the algorithm further.

One of the robust and prime tool that we are adopting in our project is temporal processing/filtering which has been used previously, temporal processing has been used to smooth motions [5] and to chuck out invisible signals [2]. [2] Presented a novel methodology for extracting the pulse rate from video recordings of the human face and demonstrated an execution using a simple webcam with extensive daylight providing illumination. They introduced a real low-cost method for non-contact heart rate measurement that is automated and motion-staying. Moreover, they have shown how the approach is gauge able for simultaneous assessment of multiple people in front of a camera easily. This technology appeared to be promising for extending and improving access to medical care, there is an availability of the low cost and widespread webcams.

[5] Introduced a system with controlled temporal sampling behavior. It transforms a high frames per second input stream into a conventional speed output video in real-time. Making use of these filters, we present a novel computational imaging system which performs real-time temporal pre-filtering to dampen temporal aliasing. Their system allows for temporally overlapping filters, which are a prerequisite for successful anti-aliasing. They have made the provision to arbitrarily choose the shape and extent of the temporal filter, hence different filtering operations can be performed, such as optimally pre-filtering for a given output kernel or artistically emphasizing or modulating motion blur. Furthermore, they have demonstrate specialized filter banks for analysing the signal in the Fourier domain, in order to understand and enhance video content based on its temporal behaviour, e.g. emphasizing or de-emphasizing motion. After the experimentation, the authors found out some of the limitations of their systems such as cost involved in Performing temporal pre-filtering by starting from a super-sampled sequence, As each sub-frame is exposed for a very short period <2ms, the number of recorded photons is limited. The signal to noise ratio is weaker compared to a single

exposure for the entire frame but they have managed to reduce it after certain modifications.

One group of researcher, [1] had extracted the cardiovascular pulse variation using mass light and a simple consumer level digital camera in movie mode. Respiration and Heart rates could be quantified up to various harmonics. Even though the green channel featuring the strongest plethysmographic signal, analogous to absorption peak by (oxy-) hemoglobin, the red and blue channels also contained plethysmographic data or information. Their results show that ambient light photo-plethysmography can be useful for medical purposes such as remote sensing of vital signs (e.g., heart and respiration rates) for triage or sports purposes and characterization of vascular skin lesions (e.g., port wine stains). This research became one of the backbones for using the motion analysis in video sequence for the measurement of various medical parameters related to human body.

### **3. PROPOSED APPROACH**

The approach proposed in this paper for the motion magnification is emboldened by the Eulerian perspective, where properties of fluid, such as pressure and velocity, evolve over time. In this case, the variation of pixel values over time is amplified and studied, in a spatially-multiscale manner. In Eulerian approach to motion magnification, motion is not explicitly estimated, but rather exaggerated by amplifying temporal color changes at fixed positions. The approach relied on the same differential approximations that form the groundwork of optical flow algorithms [6] [7].

The approach combines temporal and spatial processing to intone subtle temporal changes in a video. The process is illustrated in Figure 1. The video sequence is first decomposed into multiple spatial frequency bands. These bands might be magnified differently because either they might contain spatial frequencies for which the linear approximation used in our motion magnification does not hold or they might exhibit different signal-to-noise ratios. In the latter case, the amplification is reduced for these bands to suppress artifacts. When the goal of spatial processing is simply to enhance temporal signal-to-noise ratio by pooling multiple pixels, then the video frames are spatially low-pass filtered and down sampled for computational efficiency. In the general case, a full Laplacian pyramid is computed [8].

Then temporal processing is performed on each spatial band. The time series is considered corresponding to the value of a pixel in a frequency band and a bandpass filter is applied to extract the frequency bands of interest. The temporal processing is identical for all spatial levels, and for all pixels within each level. After that the extracted bandpassed signal is multiplied by a magnification factor that can be specified by the user, and may be assuage automatically depending on certain parameters. Then the magnified signal is added to the original and collapses the spatial pyramid to obtain the final output. But it is found that there was amplification in noise also. So to reduce/suppress that noise up to some extent and to get nearly smoothed video output, bilateral filtering is used in which a combination of domain and range filter is used to preserve the edges present in amplified output video.

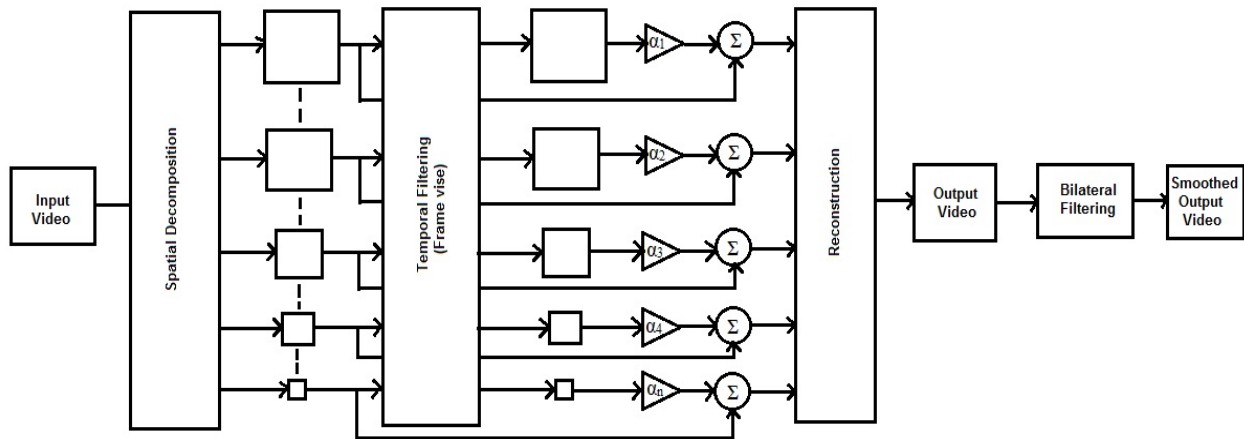


Figure 1: Overview of the Video Motion Magnification

### 3.1 Spatial Decomposition/Laplacian Pyramid

The Laplacian pyramid is a versatile data structure with many attractive features of image processing. It represents an image as a series of quasi-bandpassed images, each sampled at successively sparser densities. The resulting code element which form a self-similar structure, are localised in both space and spatial frequency. Also it is observed that the Laplacian encoding pyramid scheme requires relatively simple computations. The computations are local and may be performed in parallel, and same computations are iterated to build each pyramid level from its predecessors. An additional benefit is that in computing Laplacian pyramid, one automatically has access to quasi-bandpassed copies of image.

Over-complete decomposition is based on difference of lowpass filters; it means the image is recursively decomposed into low-pass and high-pass bands. So, the input video is totally decomposed into number of frames and each frame is processed separately to form a Laplacian pyramid. The pyramid is formed as follows,

- Original frame is initially low pass filtered and down sampled with predefined set of filter coefficient.
- The processed frame is treated as low-pass coefficient then it has been up sampled and passed through low pass filtering.
- Now this processed frame has to be subtracted from the original frame to get difference / high-pass / residue coefficient.
- The above process is repeated on the low-pass coefficient obtained from the 1st low passes filtering process for a set of levels/scale as shown in following figure.

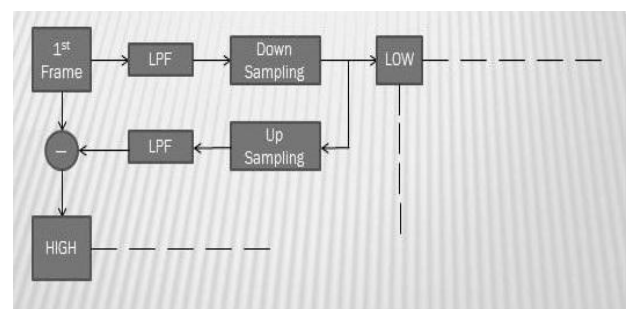


Figure 2: Overview of Spatial Decomposition/Laplacian pyramid.

- As a result, Laplacian pyramid for each and every frame of the input video can be built as shown in following figure. Here an example of “Lady” image is taken and built the Laplacian pyramid by using the above approach/algorithm.

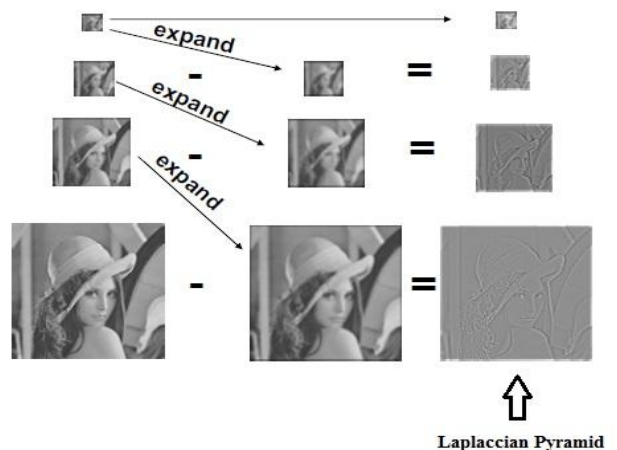


Figure 3: Example: Construction of Laplacian Pyramid of “Lady” image.

### 3.2 Temporal Filtering / Processing

- The above process will results in set of two Laplacian pyramids (one for each frame).

- The two Laplacian pyramids that are obtained from 1st step is then IIR filtered using two different set of filter coefficient.
- Mathematically,

$$y_1(n) = r_1 \cdot x(n) + (1 - r_1) \cdot y_1(n - 1) \text{ and}$$

$$y_2(n) = r_2 \cdot x(n) + (1 - r_2) \cdot y_2(n - 1) \text{ then, the}$$

difference between two is calculated as,

$$y(n) = y_1(n) - y_2(n)$$

- The output  $y(n)$  obtained in above step is then amplified by some amount and replace the difference output  $y(n)$  by its amplified output  $y(n)$  i.e.

$$y(n) = \alpha \cdot y(n)$$

Where  $\alpha$  is amplification factor.

### 3.3 Reconstruction

- The Laplacian pyramid reconstructed back with exactly the reverse procedure described in section of spatial decomposition/Laplacian pyramid.
- The procedure can be viewed with the help of following figure:

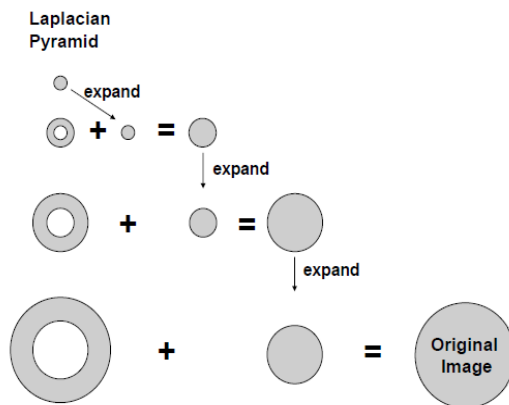


Figure 4: Reconstruction Process.

### 4. MATHEMATICAL ANALYSIS

To explain the relationship between temporal processing and motion magnification, we consider the simple case of a 1D signal undergoing translational motion. This analysis generalizes directly to locally-translational motion in 2D and 3D also.

Let  $I(x, t)$  denote the intensity at position  $x$  and time  $t$ . such that,

$$I(x, t) = f(x).$$

Since the signal undergoes translational motion, we can express the observed intensities with respect to a displacement function  $\delta(t)$ , such that,

$$I(x, t) = f(x + \delta(t))$$

The goal of motion magnification is to synthesize the signal for some amplification factor  $\alpha$ ,

$$\tilde{I}(x, t) = f(x + (1 + \alpha)\delta(t)) \text{-----(1)}$$

Assuming the signal can be approximated by a first-order Taylor series expansion,  $f(x + \delta(t))$  in a first-order Taylor expansion about  $x$ , as

$$I(x, t) = f(x) + \delta(t) \cdot \frac{\delta f(x)}{\delta(x)} \text{-----(2)}$$

Now whenever we will apply Broadband Temporal Bandpass filter to  $I(x, t)$ , the result will be  $B(x, t)$ . Then we have,

$$B(x, t) = \delta(t) \cdot \frac{\delta f(x)}{\delta(x)} \text{-----(3)}$$

Now amplify that bandpass signal with a factor  $\alpha$  and add it back to  $I(x, t)$ ,

$$\tilde{I}(x, t) = I(x, t) + \alpha B(x, t) \text{-----(4)}$$

Combining Eq. (2), (3) & (4), we have

$$\tilde{I}(x, t) = f(x + (1 + \alpha)\delta(t))$$

This shows that the processing magnifies motions—the spatial displacement  $\delta(t)$  of the local image  $f(x)$  at time  $t$ , has been amplified to a magnitude of  $(1 + \alpha)$ . This process is illustrated for a single sinusoid in Figure 2 For a low frequency cosine wave and a relatively small displacement.

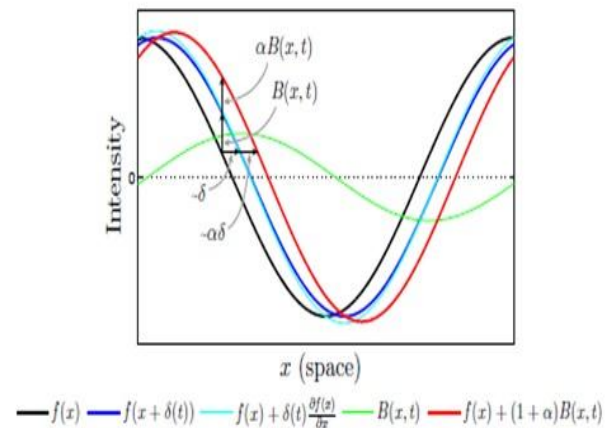


Figure 5: Temporal filtering on 1D Signal.

Following is the illustration of motion amplification on a 1D signal for different  $\alpha$  values.

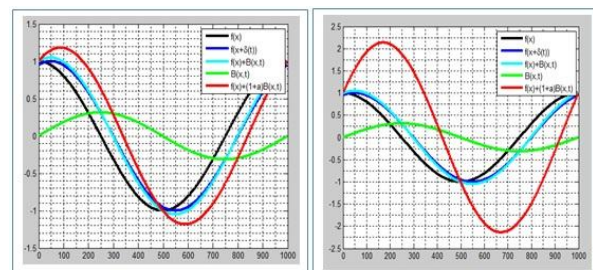
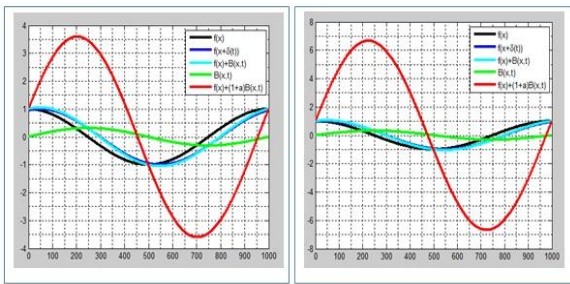
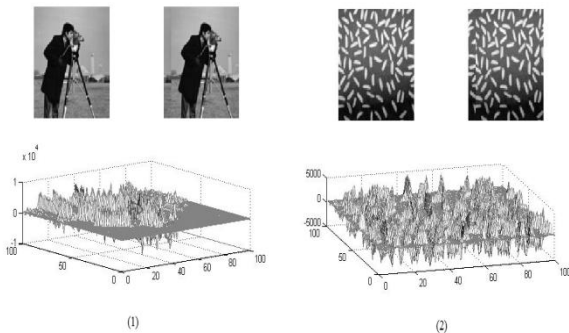


Figure 6: MATLAB output for different values of  $\alpha$ ,  
1) when  $\alpha = 1$ . 2) when  $\alpha = 5$ .



**Figure 7: MATLAB output for different values of  $\alpha$ ,**  
1)when  $\alpha = 10$ . 2) when  $\alpha = 20$

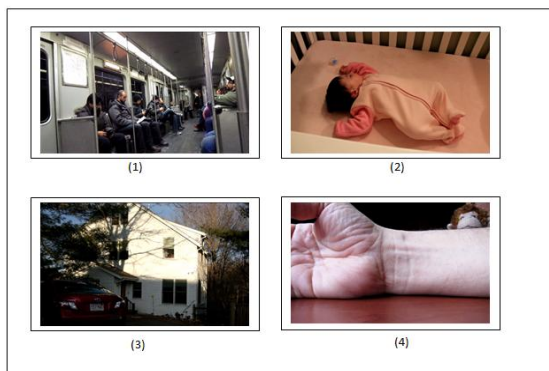
As discussed, this analysis can be generalized directly to locally-translational motion in 2D also. Here we have taken 2 input images named as cameraman.tif and rice.png, and their respective translated images. After amplification with suitable amplification factor we got the following results.



**Figure 8: Temporal filtering on 2D signal (1) cameraman.tif (2) rice.png**

## 5. RESULT ANALYSIS

We have used total 4 videos for demonstration of this project which includes baby breathing, shadow, human wrist and subway. All experiments are implemented using MATLAB (Version 2012b) on a system with 4GB RAM and Intel core i5 CPU with 2.40GHz processor. Every video is taking time in order of few minutes to compute. Filter parameters, magnification factor and cut-off wavelength are predefined by user.



**Figure 9: Sample frames from videos used in experimentation (1) subway (2) baby breathing (3) shadow (4) human wrist.**

## 5.1 Eulerian vs. Lagrangian Processing

Because the 2 methods take different approaches to motion i.e. Lagrangian approaches distinctly track motions, while the Eulerian approach does not. They can be used for complementary motion domains. In other words one can say that, the Lagrangian approaches to motion magnification [1] are complementary or opposite to the Eulerian approach proposed in this work. Such methods can amplify or boost up the motion in a video arbitrarily, but confide on accurate optical flow estimates, image segmentation, and in-painting. Such processing is difficult to do well and requires long computation times.

## 6. CONCLUSION AND FUTURE WORK

This algorithm proposed a straightforward method which takes a video as input and exaggerates subtle or imperceptible motions which is impossible to see with human naked eyes. To amplify motion, our method does not adopt a process like feature tracking or optical flow computation, but merely amplifies temporal changes using spatio-temporal processing. This eulerian based method, which temporally processes pixels in a fixed spatial region, successfully impart informative signals and amplifies small motions in real-world videos; also we believe that this method can have potential applications in medical imaging. In future, we would modify the proposed method to reduce the processing time as well as noise and attempt to make this process fully automatic.

## 7. REFERENCES

- [1] Verkrusye, w., Svaasand, l. o., and Nelson, J. S. 2008. Remote plethysmographic imaging using ambient light Opt. Express 16, 26, 21434–21445.
- [2] Poh, M. Z., Mcduff, D. J., and Picard, R. W. 2010. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Opt. Express 18, 10, 10762–10774.
- [3] Liu, C., Torralba, A., Freeman, W. T., Durand, F., and Adelson, E. H. 2005. Motion magnification .ACM Trans.Graph. 24, 519–526.
- [4] Wang, J., Drucker, S. M., Agrawala, M., and Cohen, M. F. 2006. The cartoon animation filter. ACM Trans. Graph.25, 1169–1173.
- [5] Fuchs, M., Chen, T., Wang, O., Raskar, R., Seidel, H.-P., and Lensch, H. P. 2010. Real-time temporal shaping of highspeedvideo streams. Computers & Graphics 34, 5, 575–584.
- [6] lucas, B. D., and Kanade, T. 1981. An iterative image registration technique with an application to stereo vision. In Proceedings of IJCAI, 674–679.
- [7] Horn, B., and Schunck, B. 1981. Determining optical flow. Artificial intelligence 17, 1-3, 185–203.
- [8] Burt, P., and Adelson, E. 1983. The laplacian pyramid as a compact image code .IEEE Trans. Comm. 31, 4, 532–540.
- [9] C. Tomasi, R. Manduchi, Bilateral Filtering for Gray and Color Images, Proceedings of the 1998 IEEE International Conference on Computer Vision, Bombay, India.