Design/Practical Experience [EEN1010]
Department of Electrical Engineering
FINAL REPORT

Academic Year: 2021-2022
Semester: 1

Date of Submission of Report: 18th Nov 2021

1. **Name of the Student:** KRISHI PATEL
2. **Roll Number:** B20EE30
3. **Title of the Project:** VISUAL MICROPHONE
4. **Project Category:** 3
5. **Targeted Deliverables:** An implementation of a method to recover sound of an object from its video.
6. **Work Done:**
    - **Theory**

      A sound wave is a fluctuation of pressure created in the medium in which it travels (in our case air). When sound hits an object, the pressure variations create vibrations on the surface of the object. These vibrations are not visible to a naked eye.

      Visual microphone is a method that extracts the sound from an object by filming the object with a high fps video camera and observing the vibrations created on the surface of the object.

      Each frame of the video is initially partitioned into complex valued sub-bands corresponding to different scales and orientations. This is accomplished by using the complex steerable pyramid representation of video $V(x,y,t)$. Specifically, each scale r and orientation θ is a complex image that can be expressed in terms of amplitude A and phase ϕ as

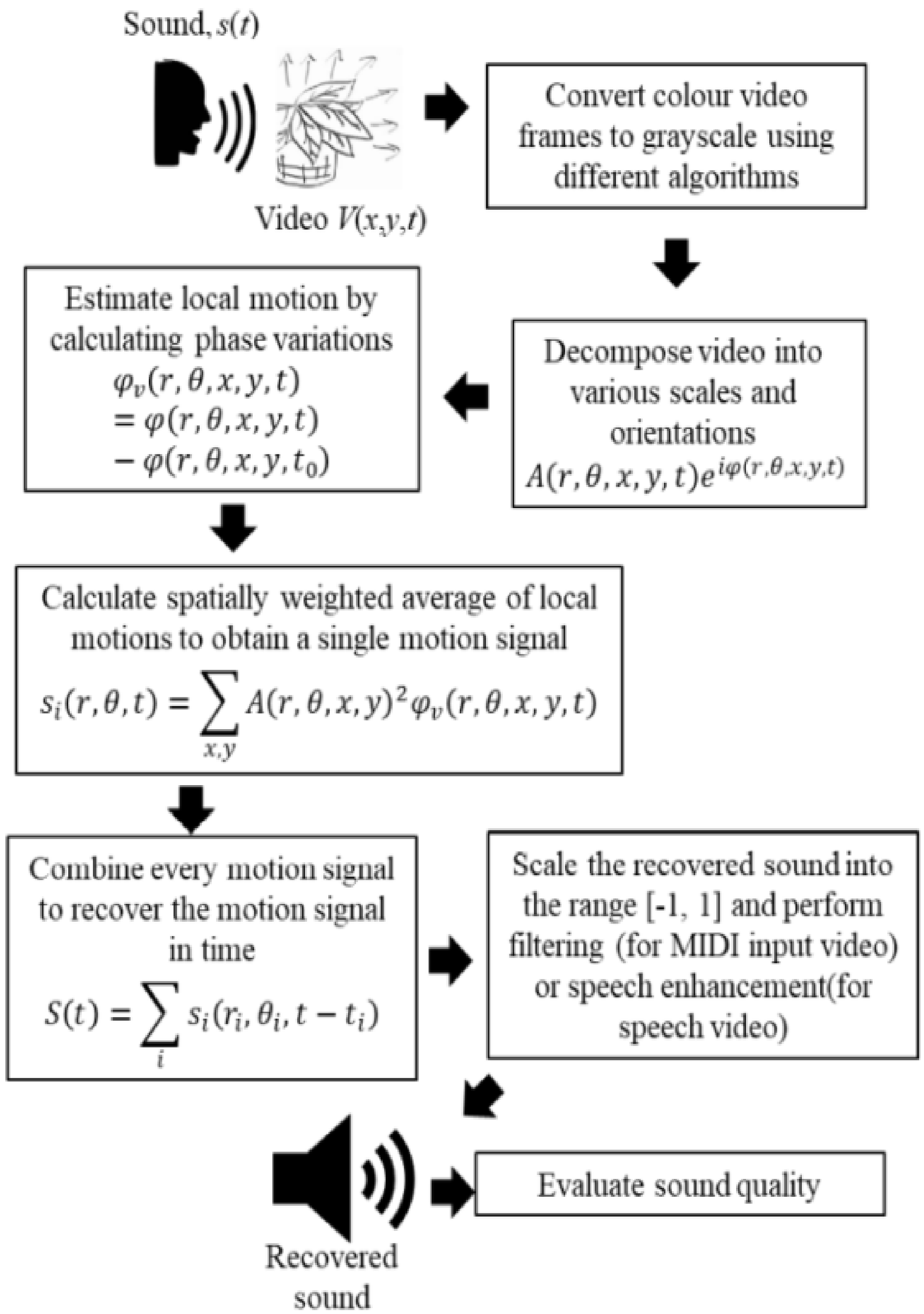
      $$A(r, \theta, x, y, t)e^{i\phi(r,\theta,x,y,t)} .$$

      Local motion signals are calculated at every pixel, orientation and scale. Local phases calculated from this equation are then subtracted from the local phases of the reference frame corresponding to $t = t_0$ (typically the first frame of the video) to compute the phase variations.

      $$\phi_v(r, \theta, x, y, t) = \phi(r, \theta, x, y, t) - \phi(r, \theta, x, y, t_0).$$

For each orientation θ and scale r in the complex steerable pyramid decomposition of the video, we temporarily align the local motion signals (to avoid destructive interference) and then compute a spatially weighted average of the local motion signals to produce a single global motion signal.

- **Methodology**



1. Filming the the object with high speed camera
2. Computing the local motion signal across the dimension of complex steerable pyramid built on the recorded video.
    - **Computing local motion signals**
      Initially partitioning of the video is done with the help of complex steerable pyramid. Specifically, each scale r and orientation θ is a complex image that can be expressed in terms of amplitude A and phase ϕ as

      A(r, θ, x, y, t)e^(iϕ(r,θ,x,y,t)) .

      Local phases calculated from this equation are then subtracted from the local phases of the reference frame corresponding to t = t_0 (typically the first frame of the video) to compute the phase variations.

      ϕ_v(r, θ, x, y, t) = ϕ(r, θ, x, y, t) − ϕ(r, θ, x, y, t_0).

3. Aligning the local motion signals and then averaging them to get a global motion signal
    - **Computing global motion signals**
      For computing the global motion signal we first take the weighted average of the motion signal associated with different scales and orientation. While calculating the weighted average we take the square of amplitude of the steerable pyramid as the weight. So the motion signal is given by

      $$\Phi_i(r,\theta,t) = \sum_{x,y} A(r,\theta,x,y)^2 \varphi_v(r,\theta,x,y,t).$$

      Now, in order to get the global motion signal we take the average of the motion signals after aligning them (to avoid destructive interference). Thus global motion signal *s* is given by

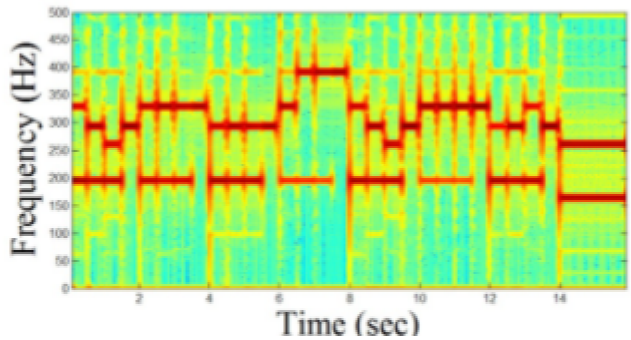
      $$\hat{s}(t) = \sum_i \Phi_i(r_i,\theta_i,t-t_i),$$

      The global motion signal is normalised and centered between [-1,1].
4. Denoising

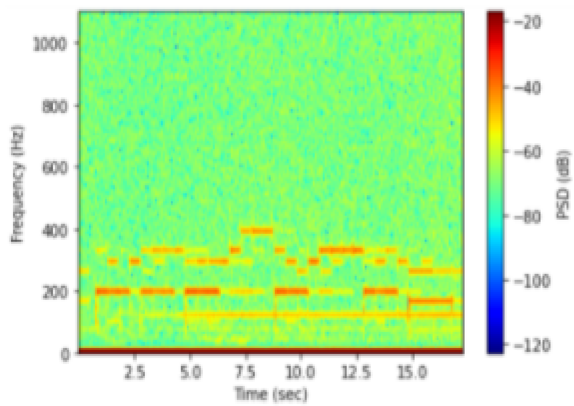
- ## Result

  The code implementation can be found here .
  The code has been tested on a video, filmed on a bag of chips. The video was filmed with a high speed camera behind a soundproof glass.
  The corresponding input and output spectrogram are as follows:

  - **INPUT SOUND SPECTROGRAM**



  - **OUTPUT SOUND SPECTROGRAM**



7. **Concluding Remarks:**
   - This report consists of the code implementation of the method described in **Davis, Abe, Michael Rubinstein, Neal Wadhwa, Gautham J. Mysore, Frédo Durand, and William T. Freeman. "The visual microphone: passive recovery of sound from video."** ***ACM Transactions on Graphics (TOG)*** **33, no. 4 (2014): 1-10.**
   - The language used for the implementation of the code is python. We learned the use of various libraries provided by python for image and video processing. The main libraries used are
     1. NUMPY - for array implementation
     2. PYRTOOLS - For the complex steerable pyramid representation of video
     3. OpenCV - for grayscalling and different video related operations.
     4. Scipy - for different operations to be done on audio

- A lot of theoretical knowledge was also been provided in this project, for eg. many video processing techniques, complex steerable pyramid representation, computing local and global motion signals, etc.

8. **References:**

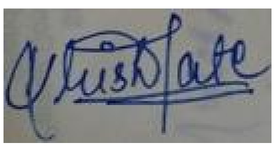
[Davis, Abe, Michael Rubinstein, Neal Wadhwa, Gautham J. Mysore, Frédo Durand, and William T. Freeman. "The visual microphone: passive recovery of sound from video." *ACM Transactions on Graphics (TOG)* 33, no. 4 (2014): 1-10.]

[Github reference]

[For openCV library]

[For pyrtools]

9. **Declaration:** I declare that no part of this report is copied from other sources. All the references are properly cited in this report.



Signature of the Student

Rajendra Nagar
Signature of the Supervisor

This section is mandatorily to be filled for the End Term Report by the supervisor. The section is to be omitted for the mid term report.

<u>Supervisor's Recommendation for the Evaluation</u>

Please tick any one of the following

1. The work done is satisfactory, and sufficient time has been spent by the student. The submission by the student should be evaluated in this term.
2. The work is not complete. Continuity Grade should be given to the student. The student would need to be evaluated in the next semester for the same Design Project with me.
3. The work is not satisfactory. There is no need for evaluation. The students should look for another Design Credit Project for the next semester.
4. [Other Comment, if 1-3 are not valid] ____________________________________

Rajendra Nagar
Signature of the Supervisor