A Digital Image Stabilization Method Based on the Hilbert–Huang Transform

Konstantinos Ioannidis and Ioannis Andreadis

IEEE TRANSACTIONS on Instrumentation And Measurement,
VOL. 61, NO. 9, SEPTEMBER 2012
An innovative technique for digital image stabilization (DIS) based on the Hilbert Huang transform (HHT)
Proposed method in a Nutshell

- Local Motion Vector (LMV) Estimation
- Hilbert Huang Transform (HHT)
  - Empirical Mode Decomposition (EMD) process
  - Hilbert Transform
  - Detect the unwanted Camera Motion
Local Motion Vector (LMV) Estimation

- LMV represents the offset of specific image regions between two consecutive images (displacement vector)
- LMV includes intentional and unwanted motion of the camera
- Block based motion estimation used in LMV
Local Motion Vector (LMV) Estimation

- The criteria for selecting the best match is sum of absolute difference (SAD)

\[
SAD(i, j) = \frac{1}{N^2} \sum_{n1=0}^{N-1} \sum_{n2=0}^{N-1} [s(n1, n2, k) - s(n1 + i, n2 + j, k - 1)]
\]
Local Motion Vector (LMV) Estimation

- Minimum SAD value defines the displacement vector

\[ [d1, d2] = \arg \min [SAD(i, j)] \]

- Full search (FS) algorithm is used in this method
Local Motion Vector (LMV) Estimation

- Introduction
- Proposed method in a Nutshell
- Local Motion Vector (LMV) Estimation
- Hilbert Huang Transform (HHT)
  - Empirical Mode Decomposition (EMD) process
  - Hilbert Transform
  - Detect the unwanted Camera Motion
- Experimental Results
- Conclusion

- Comparison of Full search (FS) algorithm with other search method

- Four step search algorithm
- Three step search algorithm
- Diamond search algorithm
Local Motion Vector (LMV) Estimation

- Using LMV estimation - \( x(t) \) (displacement vector)

\[
SAD(i, j) = \frac{1}{N^2} \sum_{n1=0}^{N-1} \sum_{n2=0}^{N-1} [s(n1, n2, k) - s(n1 + i, n2 + j, k - 1)]
\]

\[
[d1, d2] = \arg \min [SAD(i, j)]
\]
Hilbert Huang Transform (HHT)

- Empirical Mode Decomposition (EMD) process

- Introduction
- Proposed method in a Nutshell
- Local Motion Vector (LMV) Estimation
- Hilbert Huang Transform (HHT)
- Empirical Mode Decomposition (EMD) process
- Hilbert Transform
- Detect the unwanted Camera Motion
- Experimental Results
- Conclusion
Hilbert Huang Transform (HHT)

- Introduction
- Proposed method in a Nutshell
- Local Motion Vector (LMV) Estimation
- Hilbert Huang Transform (HHT)
  - Empirical Mode Decomposition (EMD) process
  \[ m_{i1}(t) = \frac{U(t) + L(t)}{2} \]
- Hilbert Transform
- Detect the unwanted Camera Motion
- Experimental Results
- Conclusion
Hilbert Huang Transform (HHT)

- Empirical Mode Decomposition (EMD) process
- Shifting Process

\[ h_1(t) = x(t) - m_1(t) \]

\[ h_{11}(t) = h_1(t) - m_{11}(t) \]

\[ h_{12}(t) = h_{11}(t) - m_{12}(t) \]

\[ \vdots \]

\[ h_{1k}(t) = h_{1(k-1)}(t) - m_{1k}(t) \]

- This will continue until value $SD$ of difference (SD) reaches between 0.2 and 0.3

\[ SD = \sum_{t=0}^{T} \left| h_{k-1}(t) - h_k(t) \right|^2 \]

\[ SD = \frac{\sum_{t=0}^{T} h_{k-1}^2(t)}{\sum_{t=0}^{T} h_{k-1}^2(t)} \]
Hilbert Huang Transform (HHT)

- Empirical Mode Decomposition (EMD) process
  - If $0.2 \leq SD \leq 0.3$
    - $c_1 = h_{1k}$ be the first IMF (Intrinsic Mode function)
    - $c_1(t)$ is removed from the image to get residue
      \[ r_1(t) = x(t) - c_1(t) \]
      - $r_1$ is the residue

- Residue is treated as the new data and subjected to same shifting process to give $c_2(t)$
  \[ r_2(t) = r_1(t) - c_2(t) \]

\[ r_w(t) = r_{w-1}(t) - c_w(t) \]

$C_w(t)$ is the $w^{th}$ IMF
Hilbert Huang Transform (HHT)

- Empirical Mode Decomposition (EMD) process
- When the residue $r_w$ becomes a monotonic function from which no IMF can be extracted.
Hilbert Huang Transform (HHT)

- Empirical Mode Decomposition (EMD) process
- Sum of the IMFs and the residue recovers the original signal.

\[ x(t) = \sum_{j=1}^{w} c_j + r_w \]

\[ x(t) = \sum_{j=1}^{w-1} c_j + c_w + r_w = \sum_{j=1}^{w-1} c_j + r_{w-1} \]

\[ = \sum_{j=1}^{w-2} c_j + c_{w-1} + r_{w-1} = \sum_{j=1}^{w-1} c_j + r_{w-2} \]

\[ = \sum_{j=1}^{1} c_j + c_2 + r_2 = c_1 + r_1 = x(t) \]
Hilbert Huang Transform (HHT)

- Hilbert Transform
  - It is used to compute instantaneous frequencies and amplitudes
  
  \[ H(x(t)) = y(t) = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{x(\tau)}{t - \tau} d\tau \]

  - Where P denotes Cauchy principal value

- z(t) analytical signal is given by
  
  \[ z(t) = x(t) + iy(t) = \alpha(t)e^{i\theta(t)} \]

Where \( \alpha(t) \) – Instantaneous Amplitude

\( \theta(t) \) – Instantaneous Phase
Hilbert Huang Transform (HHT)

- **Hilbert Transform**
  - Instantaneous Amplitude
    \[ \alpha(t) = \sqrt{x^2(t) + y^2(t)} \]
  - Instantaneous Phase
    \[ \theta(t) = \tan^{-1}\left(\frac{y(t)}{x(t)}\right) \]
  - Instantaneous Frequency
    \[ f(t) = \frac{1}{2\pi} \frac{d\theta(t)}{dt} \]
Hilbert Huang Transform (HHT)

- Detect the unwanted Camera Motion
- EMD process divides the initial signal into finite number of sub signals based on their frequencies
Hilbert Huang Transform (HHT)

- Detect the unwanted Camera Motion
- The power of each IMF is proportional to the amplitude of its sample.

\[ P_i = \sum_{t=0}^{K} \alpha_i^2(t) \]

Where \( \alpha_i \) – amplitude of a IMF’s

\[ i = 1, 2, \ldots, w+1 \]

\[ t \] – frame number
Hilbert Huang Transform (HHT)

- Detect the unwanted Camera Motion
  - IMF with lower indices include high frequency component (jitter)
  - From lower to higher IMFs, the energy content is reduced
  - After a certain IMF, a sharp increase of the energy occur due to intentional camera motion
Hilbert Huang Transform (HHT)

- Detect the unwanted Camera Motion
- IMF with higher index and lower energy content to the last IMF which includes jitter components

![Graph showing power vs. number of IMF](image)
Hilbert Huang Transform (HHT)

- Detect the unwanted Camera Motion
- Jitter and intentional camera motion

\[
X_J(t) = \sum_{i=1}^{d} c_i(t) \quad X_G(t) = \sum_{i=d}^{w} c_i(t) + r_w
\]

where \( d = \arg \min[P_i] \)
Experimental Results

\[ e_{rms} = \frac{1}{N} \sqrt{\sum_{n=1}^{N} (\overline{x_n} - x_n)^2 + (\overline{y_n} - y_n)^2} \]

Where \( N \) is the number of frames

\((x_n, y_n)\)  Optimal camera motion

\((\overline{x_n}, \overline{y_n})\)  Resulting camera motion
Conclusion

- DIS method based on HHT has been presented
- Jitter signal is defined based on its two principle feature
  - High frequency
  - Low energy
- Experimental results have proven that the proposed method can successfully decompose two camera motions