Lips Region Detection for Visual Speech Processing

Presented by: Adi Fridman and Ortal Shechtman

Supervised by: David Dov

29/1/2014
• Motivation- why visual speech?

- The performance of traditional *audio* algorithms deteriorates in highly *non stationary* acoustic environment.

- *Visual information* is totally *immune* for that type of noise.

- Visual information has shown to be *useful* for Speech Recognition and Voice Activity Detection.

- *Frontal videos* are commonly used for VOIP and cellphone video calls.

- *Lips* contain over 90% of visual speech data.
• Previous work- Approaches for Lips Detection

- **Color based** lips detection
  - Based on color segmentation.

- **Key points** detection
  - Key points examples: eyes and lips edges.

- Detection based on **face geometry**
  - Assume constant distances between facial features for different speakers.
Lips Region Detection for Visual Speech Processing

- System Block Scheme - incorporation with audio visual VAD

- Lips region detection + Visual features extraction
- Audio features extraction
- Audio visual detection
- Speech/non speech

Lips region detection

- Face and eyes recognition
- Lips detection
- Accurate cropping
**Proposed Algorithm - Geometry Algorithm**

- Detect eyes and face.
- Define $M =$ distance between eyes.
- Detect lips region = estimate lips region size and location using $M$ and face geometry.

**Disadvantages:**

- Different people has different face geometry.
- Slight head rotation may change lips location with respect to eyes.
Lips Region Detection for Visual Speech Processing

**Proposed Algorithm** - Geometry Algorithm - face proportions

- **L** - vertical distance between the lips center and the middle of the eyes.
- **M** - horizontal distance between the pupils.
- The ratio L/M was determined empirically.

$L = 1.11M$

$L = vertical distance$

$M = horizontal distance$

$0.7M$
• **Proposed Algorithm - Geometry + Search Algorithm**

- Detect lips region using geometry algorithm.
- Enlarge the detected region by a constant factor.
- Calculate visual feature.
- Search lips region in the feature space.
- Define **energy** based measure to locate Lips region accurately.
Lips Region Detection for Visual Speech Processing

- **Proposed Algorithm** - Geometry + Search Algorithm - Visual Features

- Use features of the VAD algorithm

- **Optical Flow:**
  - Decomposes an image into small squared cells.
  - For each cell find the matching cell in the next frame.
  - The cell change direction is the motion vector.

- **HOG Features:**
  - Decomposes an image into small squared cells.
  - Computes an histogram of oriented gradients in each cell.
  - normalizes the result and return a descriptor for each cell.
Lips Region Detection for Visual Speech Processing

- **Proposed Algorithm** - Geometry + Search Algorithm - Visual Features

- Use **HOG** Features following to David’s research.

- The cell size $D$ is chosen according to a threshold.

- Large $D$ – more information.

- Small $D$ – better resolution.

- HOGs may be efficiently scale converted.
• Assumption: the lips area contain the features with *high energy*.

• Search a bounding box with maximal *energy*.

• Bounding box size - same as in the Geometry algorithm.
Lips Region Detection for Visual Speech Processing

- Simulation Results - Qualitative Evaluation
• Conclusions

**Advantages:**

• Accurate lips region cropping
• Robust to lightning and skin color.
• Can be used with a variety of features.
• Low computational cost - no additional features extraction.
• Outperforms face geometry algorithm.

**Disadvantages:**

• Ground truth setting is still challenging.
• Compatible for front videos only.