VACE Video Text Recognition Evaluation Plans

John Garofolo
NIST Information Access Division

SDIUT 2005

Introduction

- Goal: Develop technology to detect text in video and transcribe the words
- Approach: Supporting progress by providing research and evaluation infrastructure
  - task and evaluation specifications
  - training, development, and evaluation data, human reference annotations, and scoring tools
  - publicly available tools and data to support intrinsic developmental evaluation
  - open common extrinsic evaluations and workshops

Evaluation Team

- NIST
  - Coordination
- University of South Florida
  - Evaluation specs, scoring software, and implementation
- Advanced Interfaces
  - Reference annotations
- University of Maryland
  - Annotation tool development and support
- BBN/SRI
  - Contributed to development of task definitions
  - Annotation quality control and enrichment

Task Definitions

- Detection Task: Spatially locate the blocks of text in each video frame in a video sequence
  - Text blocks (objects) contain all words in a particular line of text where the font and size are the same
- Tracking Task: Spatially/temporally locate and track the text objects in a video sequence
- Recognition Task: Transcribe the words in each frame, including their spatial location (detection implied)
- Currently working with broadcast news data

Annotation Examples

Line Level Annotation for Detection

Word Level Annotation for Recognition

ARDA
Video Analysis
Content Extraction

NIST
National Institute of
Standards and Technology

Recognition Task Excluded Text
(not scored via "don't care" areas)

Low-readability text

Dynamic text (scoreboard)

Scrolling text (News ticker)

Highly stylized text (logos)

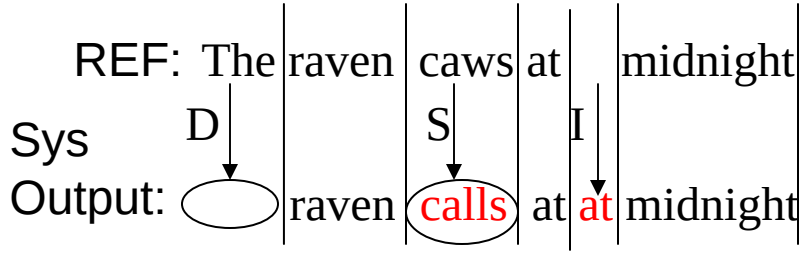
ARDA
Video Analysis Content Extraction

NIST
National Institute of Standards and Technology

Recognition Scoring

- Spatially map system output detected words to reference words, then compare the strings for mapped words
  - An unmapped word in system output incurs an Insertion (I) error
  - An unmapped word in reference incurs a Deletion (D) error
  - A mapped word with a character mismatch incurs a Substitution (S) error

$$WER = \frac{(I + D + S)}{(\text{Total \# Words in Ref})}$$



WER = (1 + 1 + 1)/5 = 3/5 (60%)

- Errors are accumulated over entire test set
- Character Error Rate is also generated

ARDA Video Analysis Content Extraction

NIST National Institute of Standards and Technology

Current Recognition Data Sets

English Broadcast News

- Training/Dry Run Development Set
  - 5 Clips, will be expanding
    - 14.5 minutes
    - 1181 words
- Planned Evaluation Set
  - 25 Clips
    - 62.5 minutes
    - 4178 words

Plans

- Near-term:
  - Training and test datasets for English broadcast news – currently implementing “dry run” evaluation
  - Training and test datasets for Arabic broadcast news
  - English and Arabic evaluations scheduled for Feb-Mar 2006
  - Results to be reported at MLMI-2006, (Rich Transcription Workshop, May 4-5 2006, US Location TBD)
- Longer-term:
  - Conduct annual text recognition evaluations
  - Explore object-centric text recognition evaluation
    - Errors normalized by object rather than by frame instances
  - Focus on most important text
    - subsetting on particular text (e.g., named entities)
  - Explore semantic text clustering and word ordering
  - Explore clustering of captions with images and speech
  - Explore new domains (meetings, surveillance)

We Welcome Your Participation

1. Send an email with your contact information, subject: Text Recognition Evaluation to vace-info@nist.gov
2. You will receive
   a) a participation agreement that must be signed in order to receive the data
   b) Instructions on how to contact the Linguistic Data Consortium (LDC) to receive data.
   c) training data and reference annotations
3. You will be required to
   a) Participate in periodic planning teleconferences
   b) Adhere to the requirements in the participation agreement form, agreeing to have your results published and your use of those results and the datasets
   c) Present your work at the designated evaluation workshop

Questions?