Handwriting Database Panel

Danica M. Ommen, SDSU

JoAnn Buscaglia, FBI CTFR Unit

Ted Burkes, FBI QD Unit

Christopher Saunders, SDSU

Peter Tytell, NYC

Handwriting Comparison

- Handwriting Examinations
  cursive, print, numerals, signatures

- Generally side by side comparison,

- Questioned writing to Known writing
  Q to K   *specific source*

- Questioned writing to Questioned writing
  Q to Q   *common source*

Operational Databases
used to associate documents and cases

FISH (Forensic Identification System for Handwriting)
   US Secret Service (BKA)
      threats to the POTUS and some others
      uses actual evidence images
      requires manual input of handwriting characteristics
      associates based on handwriting properties.

Anonymous Letter File

Bank Robbery Note File
   FBI Lab
      use actual evidence images
      significant text entered into the database
      associations made by wording
      more by HW comparisons of multiple submissions

Purposes of Research Datasets

- Evaluation of automated technology
- Evaluation of human performance ("black box")
- Assessment of rarity of writer profile and random match probabilities
  - quality and quantity of features for conclusions
- Research on
  kinematics of handwriting
  variability of handwriting with age of writer
  (*i.e.*, long-term and short-term within-writer variability)

West Virginia University.

HANDWRITING
COLLECTION

The way we think is the way we write

Extant Research Datasets

FBI 500 dataset (~ 500)

    IRB approval did not include public release

    Very difficult if not impossible to get retroactive approval
        FBI IRB requires subject/writer to consent to new use

FBI – WVU dataset

    (currently sequestered for ongoing research, but some
        will be available for release in the future)

Others

    NIST has 2 listed on its catalog
        for the Census (*handprinting and numerals*)
        Chinese

    Cedar Fox collection (NIJ funded)

    Lisa Hanson (NIJ funded)

Collecting Datasets
that are fit-for-purpose

- Collection design depends on intended use
  - What/how much data should it contain?
  - What should it be representative of?
  - Are convenience samples adequate?
- Existing data vs targeted collection
  - Can we use public collections (accumulations) for extracting features?
    - Library of Congress & National Archives collections
    - Federal Forms (*e.g.*, Census, INS, Voter)
      - No ground truth
      - Could be outdated handwriting (*i.e.*, an old copybook)
      - Limited instances per writer to characterize writer profile
  - Prison populations

Distorted Writing and Signatures Simulation or Disguise?

Simulations
- – freehand drawn copies from a model
- – freehand written from memory
- – tracing
- Choice of disguise strategy
- Skill level of the Simulators/Disguisers
- Complexity and fluency of the model
- Disease/intoxication/old age/dementia

Controlled Capture Mechanisms

Pen & paper capture
and digital conversion for dataset

– dynamic features that QDE uses lost

– limit to QDE conclusion with non-original documents

• Digital capture

– No Conversion

– Captures dynamic features
(*e.g.*, acceleration, pressure patterns)

– Needs paper for writer feedback FDE examination