Problem: Still-to-Video face recognition
1. Gallery: high quality still face images (e.g., sharp and high face resolution ones)
2. Probe: low quality video face frames (e.g., blur and low face resolution ones)

Solution: Couple alignments with recognition
1. Quality Alignment (QA): select the frames of best quality from videos
2. Geometric Alignment (GA): jointly align the selected frames to the still faces
3. Sparse Representation (SR): sparsely represent the frames on the still faces

Formulation
\[
\hat{L}, \hat{T} = \arg \min_{L,T} \|Z\|_1 + \sum_{i=1}^{n} \|B_{S_i} L_i + E\|_1, \quad s.t. \quad Y \circ T = B + E, \quad B = AZ, \quad S_i = \{j|L_j = i\}.
\]
- Couple GA with SR: \(Y\circ T = B + E, \quad B = AZ, \quad \|Z\|_1 \leq t\)
  - \(Y\): Video faces, \(A\): dictionary (still faces)
  - \(\circ\) and \(T\): Alignment operator and parameters
  - \(B\): Sparse representations, \(E\): residual errors

Optimization
Linearization:
\[
\{T, L\} = \arg \min_{Y, T} \|Z\|_1 + \sum_{i=1}^{n} \|B_{S_i} L_i + E\|_1, \quad s.t. \quad Y \circ T + \Delta T = B + E, \quad B = AZ, \quad S_i = \{j|L_j = i\}.
\]
- Set the segments at coarse stage: \(S_i = \{1, \ldots, L\}, \quad S_i \neq \phi, \quad i = 2, \ldots, c\)
- Apply Augmented Lagrange Multiplier to solve:
  \[
  \{\hat{T}, \hat{Z}\} = \arg \min_{Y} \|Z\|_1 + \sum_{i=1}^{n} \|B_{S_i} L_i + E\|_1, \quad s.t. \quad Y \circ T + \Delta T = B + E, \quad B = AZ;
  \]
- Update transformations: \(T = T + \Delta T\)
- Update segments at fine search stage: \(S_i = \arg \min_{y} \|y\|_1 + \|A_{S_i} z_{y, j}\|_2\)
- Compute \(\phi_i\) of \(S_i\), \(i = 1, \ldots, n\) for voting class label.

Datasets
1. YouTube-S2V dataset: 100 subjects, privately collected from YouTube Face DB [Wolf et al., CVPR' 11]
2. COX-S2V dataset: 1,000 subjects, publicly released in our prior work [Huang et al., ACCV' 12]

Results
Comparative methods:
1. Baseline: SRC[1], CRC[2]
4. Our method: Coupling Alignments with Recognition (CAR)

Evaluation terms:
1. Face Alignments (QA and GA)
2. Sparse Representation (SR) for Face Recognition

Conclusions
- The proposed method jointly performs GA, QA and SR in a unified optimization.
- We employ an iterative EM-like algorithm to jointly optimize the three tasks.
- Experimental results demonstrate that GA, QA and SR benefit from each other.