

Towards Digitizing Filled Indic Handwritten Forms *

MOTIVATION

- Streamline conversion of handwritten forms to digital format.
- Automate digitization of forms in English and four Indian languages: Hindi, Bengali
- Significantly reduce time needed for manual digitization.

METHODOLOGY

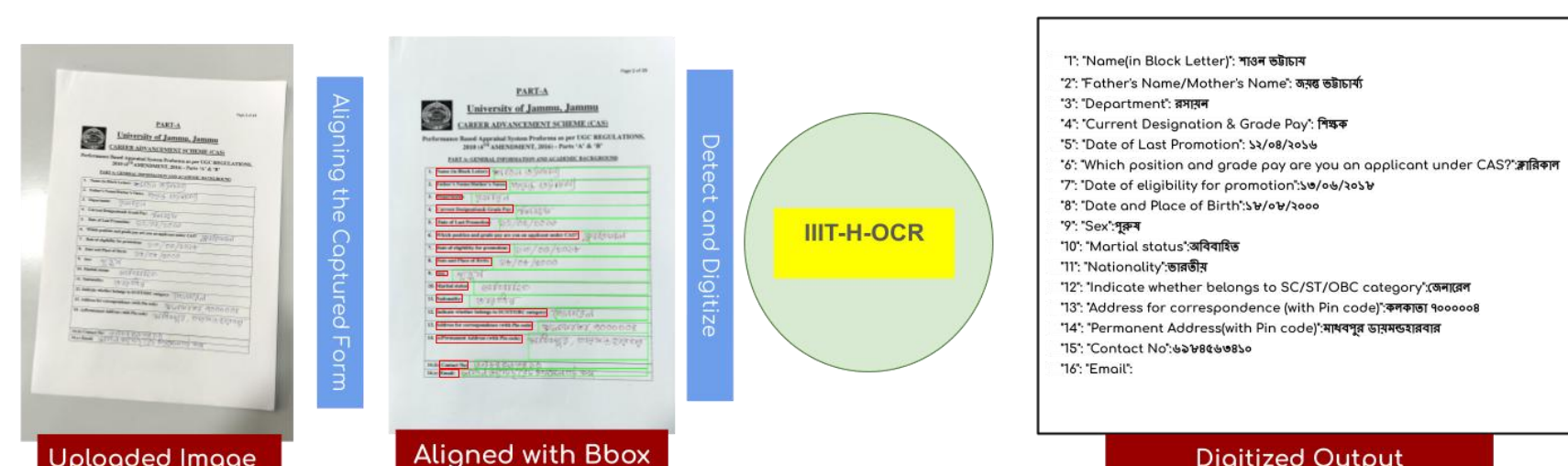
Step 1: Accept forms captured by smartphones in any orientation.

Step 2: Canonical alignment of forms.

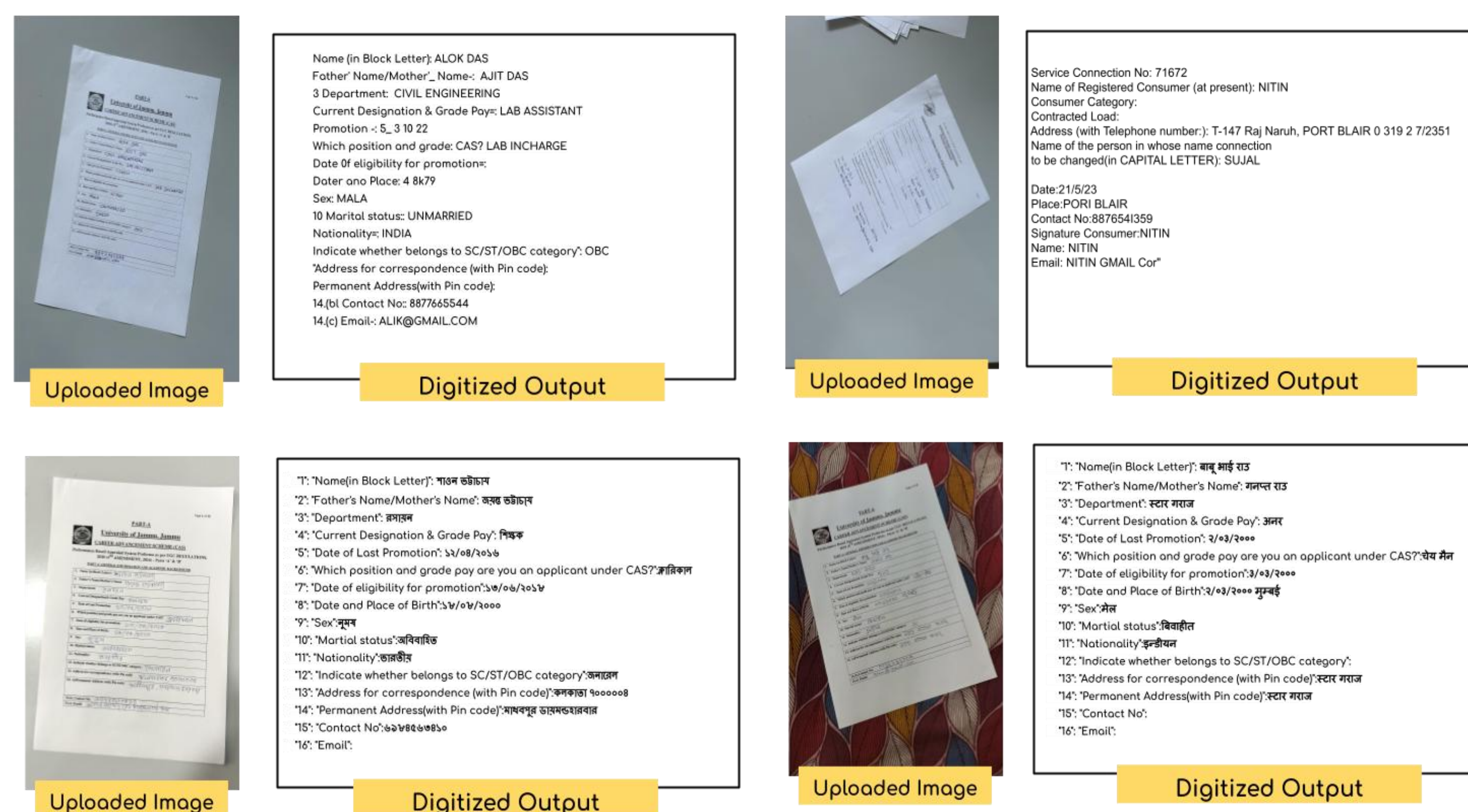
Step 3: Annotate form's template to identify keys and corresponding values.

Step 4: Detect and extract fields from the aligned forms.

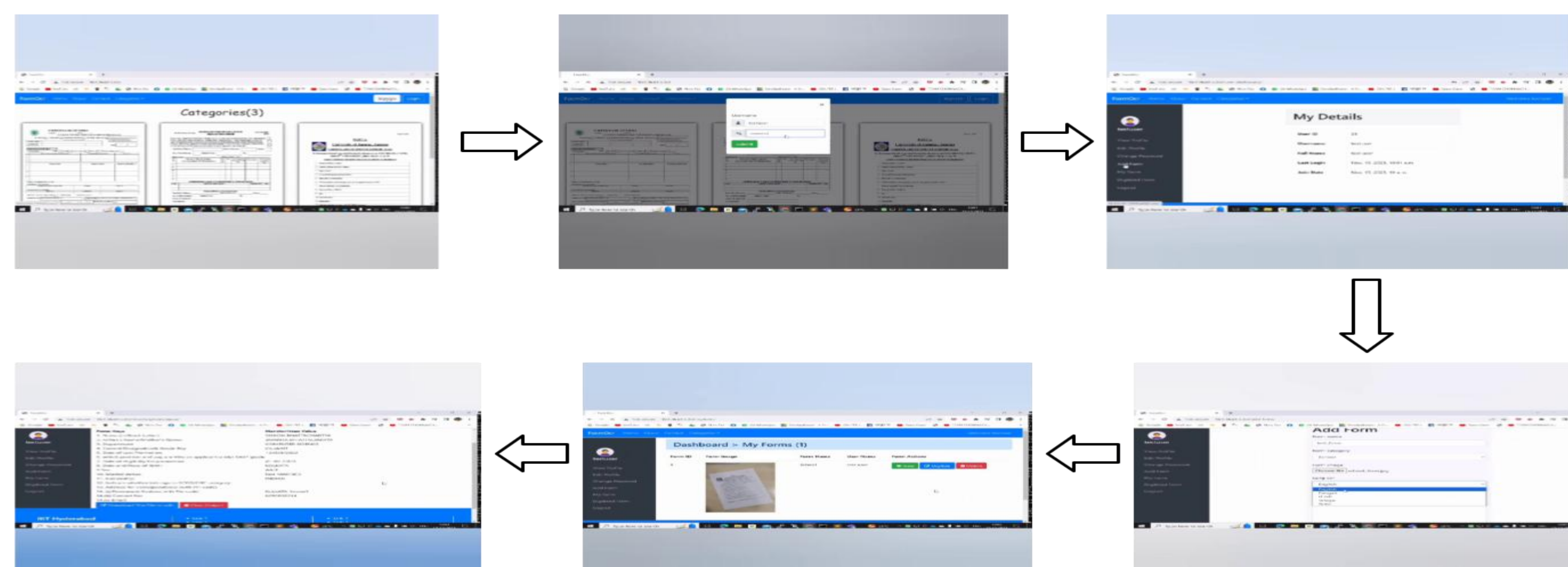
Step 5: Send the extracted fields to Optical Character Recognition (OCR) for recognition.



RESULTS



Demo of Form Digitization Process



* Accepted in 9th International Conference on Computer Vision & Image Processing 2024

CVIT

Research center

Summary

- Introduces an end-to-end pipeline for digitizing handwritten forms.
- Process includes alignment, content detection, and recognition and maintaining order.
- Create a new dataset of handwritten forms.
- Need to more focus on multilingual form processing.

REFERENCES

- JaidedAI: EasyOCR
- Smith, R. Tesseract Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)
- Youngmin Baek and Bado Lee and Dongyoon Han and Sangdoo Yun and Hwalsuk Lee: craft