Vital Monitor and ID Detection through Machine Vision for Improving EMS Communication Efficiency

Miller AC, Blalock TW. Augmented reality: a novel means of measurement in dermatology. J Med Eng Technol. 2021 Jan;45(1):1-5. doi: 10.1080/03091902.2020.1838641. Epub 2020 Nov 16. PMID: 33191825.

Group 11:
Robert Huang

Clinical Mentors:
Dr. Nick Dalesio
Dr. Laeben Lester

Computer Vision Mentor:
Dr. Mathias Unberath

600.456 CIS2 Spring 2021

Project Summary

Two Objectives:

- **Objective 1:** Automatically **extract and insert information** into a digital medical note from Smart Glasses Feed
- **Objective 2:** Provide View of Data Monitors Remotely with AI and Computer Vision from Smart Glasses Feed

Relevance

- Both objectives need OCR to extract and classify the textual elements from both identification and vitals monitors.
- Even state of the art OCR techniques have significant accuracy drops in **presence of poor lighting, uneven illumination or perspective distortion.**
- For emergency medicine, illumination and distortion is also unpredictable.

[1] Vuzix Corporation, 2020

AirSupport

Overview of Paper

- A novel ‘frame weighing’ technique that takes the **per-frame Optical Character Recognition** (OCR) textual results of non-ideal video feed, specifically that taken from a mobile camera, and adds them through weights in order to more accurately extract the textual elements from a document.

[1] Vuzix Corporation, 2020

AirSupport

Background

- OCRs increase in demand. Saves time, money, and effort in documentation.
- Increasing use of mobile phones in uncontrolled environments for OCR.
    - OCRs do not work natively in nonoptimal conditions.
- Mobile phones can take videos, whose OCR outputs can be combined.
    - Incorporate different illuminations, angles, and focus characteristics into the text recognition algorithm

[1] Vuzix Corporation, 2020

Johns Hopkins University

AirSupport

Sources of Error

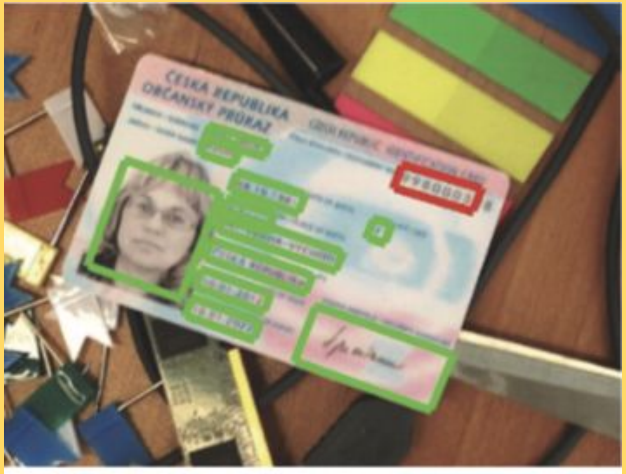
- Physical Difficulty
  - Defocus, Glare

[Figure 2, 1]

[1] O. Petrova 2021

Sources of Error

- Recognition-Stage Difficulty



[Figure 6, 1]

[1] O. Petrova 2021

Problem Statement

- Represent outputs as matrices.

$$x = (x_1, x_2, \ldots, x_K) \in [0.0, 1.0]^K, \quad \sum_{k=1}^{K} x_k = 1,$$

Character representation [1]

$$X = (x_{jk}) \in [0.0, 1.0]^{M \times K}, \quad \forall j : \sum_{k=1}^{K} x_{jk} = 1,$$

|x1|
|x2|
|x3|

String representation [1]

[1] O. Petrova 2021

ROVER

1. When two recognized texts are obtained, due to variable length recognition, the texts are first aligned using the distance equation below.
2. Then, the algorithm combines the two texts using the weighted equation below to produce a resulting estimation of the desired text.

$$\rho\left(x^1, x^2\right) = \frac{1}{2}\sum_{k=0}^{K}\left|x_k^1 - x_k^2\right|$$

Formula for alignment [1]

$$r = \left(r_k\right) \in \left[0.0, 1.0\right]^{K+1}, \quad \forall k : r_k = \frac{x_k^1 \cdot w(x^1) + x_k^2 \cdot w(x^2)}{w(x^1) + w(x^2)},$$

Formula for combining/voting strings [1]

Lower weight ABCC

Higher weight ACCA → AACCA

[1] O. Petrova 2021

Weighting Criterion - Focus Estimation

- Focus estimation: More focused image will have more weight.

$$G^{V}_{r,c}(I_i(\bar{X})) = \left|I_{r+1,c} - I_{r,c}\right|,$$
$$G^{H}_{r,c}(I_i(\bar{X})) = \left|I_{r,c+1} - I_{r,c}\right|,$$
$$G^{D_1}_{r,c}(I_i(\bar{X})) = (1/\sqrt{2})\left|I_{r+1,c+1} - I_{r,c}\right|,$$
$$G^{D_2}_{r,c}(I_i(\bar{X})) = (1/\sqrt{2})\left|I_{r,c+1} - I_{r+1,c}\right|,$$

Gradient Calculation [1]

$$F(I_i(\bar{X})) = \min\{q(G^{V}(I_i(\bar{X}))), q(G^{H}(I_i(\bar{X}))), q(G^{D_1}(I_i(\bar{X}))), q(G^{D_2}(I_i(\bar{X})))\},$$

Focus Estimation [1]

[1] O. Petrova 2021

Weighting Criterion - Confidence Level

- Confidence Level: More confident images will have more weight.

$$Q(X) = \min_{j=1}^{M} \left\{ \max_{k=1}^{K} x_{jk} \right\}$$

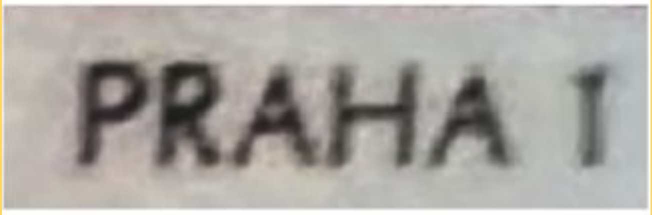
Confidence Level Weight Calculation [1]

[0,0.1,0.9]
[0,0.2,0.8]
[0,0.3,0.7]

[1] O. Petrova 2021

Per-Character Weighting

- Individual characters in a text field may contain different levels of quality
- Assign weights to each character



[Figure 6, 1]

[1] O. Petrova 2021

Weighting Model

- Question: How many frames to use?
- Firstly order the frames from the best to the worst quality (according to w), and then keep the best t results by zeroing the weight of the worst frames.

(Image, weight)

[(I1, 0.6), (I2, 0.5), (I3, 0.7)] ⟶ $\pi(i) < \pi(j) \Leftrightarrow w(I_i(\overline{X}), X_i) \geq w(I_j(\overline{X}), X_j)$ ⟶ **[(I3, 0.7),(I1, 0.6), (I2, 0.5)]**

[(I3, 0.7),(I1, 0.6), (I2, 0.5)] ⟶ $w_i^{(t)} = \begin{cases} w(I_i(\overline{X}), X_i), & \text{if } \pi(i) \leq t, \\ 0, & \text{if } \pi(i) > t. \end{cases}$ **[(I3, 0.7),(I1, 0.6), (I2, 0)]**

t=2

[1] O. Petrova 2021

Overall Model

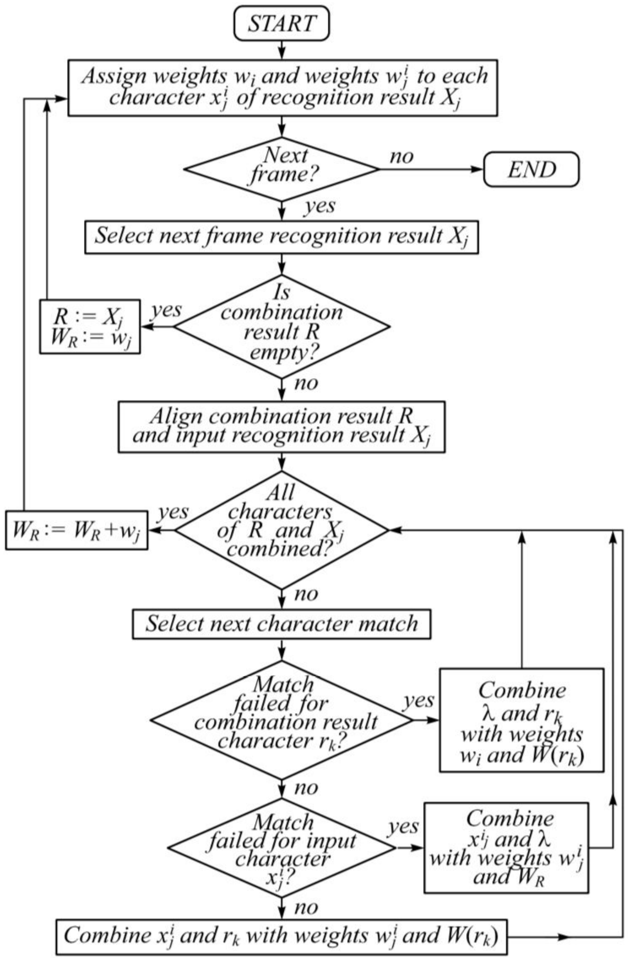


Fig. 7. Diagram of combining text recognition results with per-character weighting

Diagram of combining text recognition results with per=character weighting [Figure 7, 1]

[1] O. Petrova 2021

Experiments Conducted

- Datasets:
  - MIDV-500: 500 videos with distortion
  - MIDV-2019: 200 videos with distortion
- Four Text Fields considered:
  - Document numbers, numeric dates, latin name components, and machine-readable zone lines were read
- First Experiments:
  - Ten Categories
    - Focus weighted criteria
    - Confidence weighted criteria
    - No weighting
    - Choosing the best result
    - Weighting the three best results, weighting the top 50% of results, and Weighting every frame

Experiments Conducted

- Second Experiments:
  - Five Categories
    - Focus weighted criteria ONLY
    - No weighting
    - Full string weighting with all frames
    - Full string weighting with the best 50% of frames
    - Per-character weighting with all frames
    - Per-character weighting with the best 50% of frames

Results

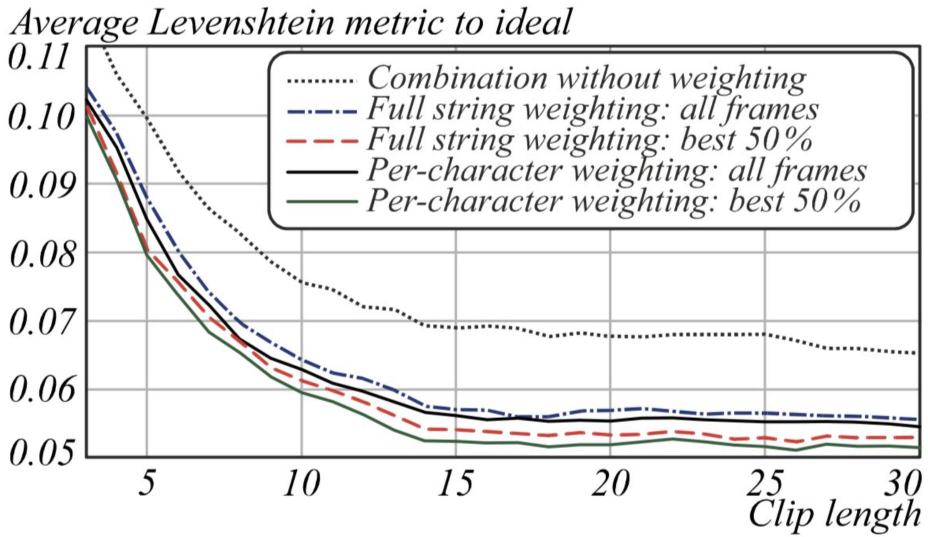


Fig. 14. Performance profiles for focus estimation weighting on MIDV-500

Performance Profile for Focus Estimation Weighting on MIDV-500 [Figure 14, 1]

Table 5. Mean Normalized Levenshtein metric distance to the correct result on MIDV-500 using focus estimation

Combination method	Mean Normalized Levenshtein metric					
	5 frames	10 frames	15 frames	20 frames	25 frames	30 frames
Without weighting	0.0995	0.0756	0.0689	0.0677	0.0680	0.0652
Full string weighting: all frames	0.0879	0.0643	0.0570	0.0569	0.0565	0.0555
Full string weighting: best 50%	0.0804	0.0612	0.0541	0.0533	0.0529	0.0529
Per-character weighting: all frames	0.0847	0.0628	0.0561	0.0553	0.0552	0.0545
Per-character weighting: best 50%	**0.0795**	**0.0595**	**0.0524**	**0.0518**	**0.0516**	**0.0515**

Mean Normalized Levenshtein metric distance to the correct result on MIDV-500 using focus estimation [Table 6, 1]

[1] O. Petrova 2021

Conclusions

- The combination of **the best 50% frames** with **per-character weighting** and **focus weighting** will result in the best performance when integrating multiple frames of a text field

Critiques

- Brief outline descriptions describes a section in isolation rather than in the flow of the paper.
- The results of some papers are referenced in a conclusive form, but explicit numbers or percentages are never produced
- Could have looked at more fields of the document.
- Only Two Criterions
- Experiments do not consider how documentation recognition errors may affect the weighting algorithm.

Key Takeaways

- best 50% frames
- use per-character weighting
- use the focus weighting criterion.

References

1. O. Petrova, K. Bulatov, V. Arlazarov, V. Arlazarov, Weighted combination of per-frame recognition results for text recognition in a video stream, Computer Optics. 45 (2021) 77–89. doi:10.18287/2412-6179-co-795.

2. O. Petrova, K. Bulatov, V.L. Arlazarov, Methods of weighted combination for text field recognition in a video stream, Twelfth International Conference on Machine Vision (ICMV 2019). (2020). doi:10.1117/12.2559378.

Thank You!
Questions?