

Перетворення відеозапису з дошки у слайд-шоу

Доповідач: Максим Шило¹ Науковий керівник: Водолазський Є. В.¹

¹Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

16 червня 2022

Мотивація



Теорема 2. Нехай $A = T_1 \sqcup T_2 \sqcup T_3$

Тоді може відношуватися один з

що $A \sim T_1 \sqcup T_2 \sqcup T_3$

• Нехай $a \sim b \Leftrightarrow \exists t_1: a \in t_1, b \in t_1$

тобто $a \in t_1, b \in t_1 \Rightarrow a \in t_1$

суми $a + b = a \in t_1, b \in t_1 \Rightarrow a + b \in t_1$

також $a \sim b \Leftrightarrow \exists t_2: a \in t_2, b \in t_2$

тобто $a \in t_2, b \in t_2 \Rightarrow a \in t_2$

$\Rightarrow (a \sim b) \Leftrightarrow a \sim c$

• Тимчасом:

$A^2 = \{(a, b) | a \in A, b \in A\}$

R - більшість відношування на A^2

$(a, b) R (x, y) \Leftrightarrow \begin{cases} (a, b) \sim (x, y) \\ (a, b) = (x, y) \end{cases}$

$(x, y) R (u, v) \Leftrightarrow \begin{cases} (x, y) \sim (u, v) \\ (x, y) = (u, v) \end{cases}$

$(a, b) = \{(a, b), (b, a)\}$

$(a, a) = \{(a, a)\}$

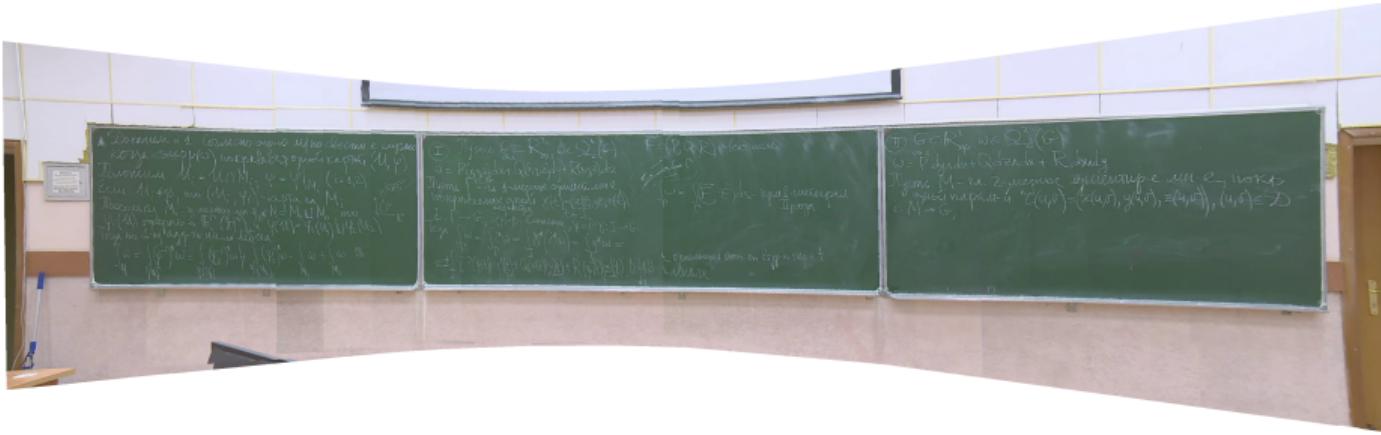
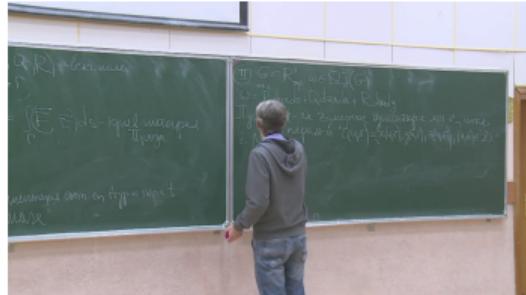
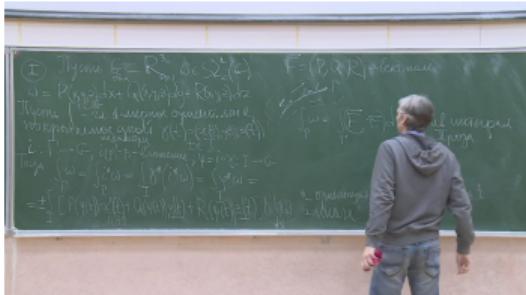
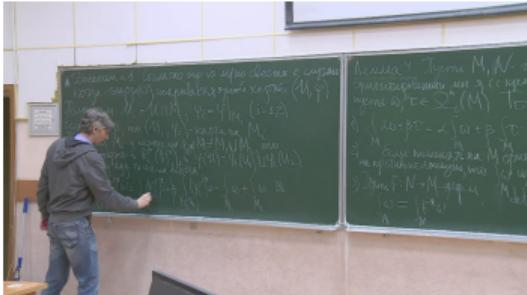
Якщо $a \neq b$, то $\{a\} \cap \{b\} = \emptyset$

$A^{(2)} = A^2 / \text{небільшість відношування}$

$A^{(2)} = A^2 / R$

$\mathbb{Z} \times \mathbb{N} \sim (m_1, n_1) \sim$

Мотивація



Вступ

Об'єкт дослідження — відеозаписи лекцій.

Предмет дослідження — автоматична обробка відеозаписів.

Метою роботи є розробка алгоритму, що перетворює відеозапис лекції у панорамні знімки без викладача за умови, що камера може рухатись.

Дана інформаційна технологія повинна мати ряд властивостей:

- здатність працювати на звичайному смартфоні у режимі реального часу;
- можливість працювати з дошками різного кольору;
- можливість працювати з рухливою камерою;
- мінімальна кількість дефектів на слайдах, таких як наявність фрагментів викладача або видимі шви у місцях склейки кадрів.

Зміст

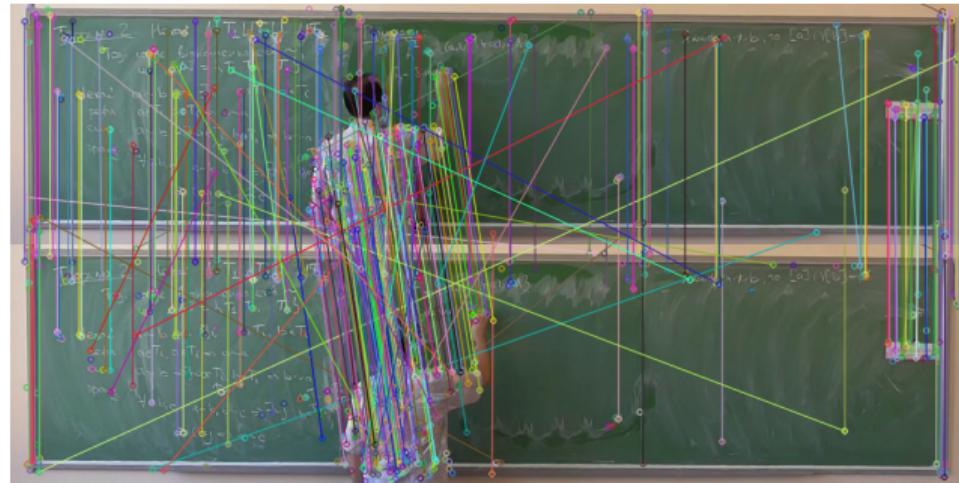
- 1 Процедура створення панорамних слайдів
- 2 Стабілізація відео
- 3 Методи локалізації людини та рухомих об'єктів
- 4 Створення панорами
- 5 Обробка слайдів
- 6 Conclusions

Процедура створення панорамних слайдів



Стабілізація відео

Оскільки дошка — це плоска поверхня, ми можемо знайти матрицю гомографії між двома кадрами. Для цього використовуємо алгоритм SIFT та RANSAC.



Кадри F^i (верхній) та F^{i+s} (нижній) із ключовими точками та лініями,
які поєднують відповідні точки з відео

Алгоритм Бойкова Колмогорова

Сформулюємо задачу максимального потоку

$$\sum_{t \in N_s} f_{st} \rightarrow \max_{f: \tau \rightarrow R}$$

з обмеженнями

$$\begin{cases} f_{tt'} \leq c_{tt'}, & \forall tt' \in \tau, \\ \sum_{p \in P_t} f_{pt} - \sum_{t' \in N_t} f_{tt'} = 0, & \forall t \in T \setminus \{s, e\}, \\ \sum f_{tt'} \geq 0, & \forall tt' \in \tau. \end{cases}$$

Це означає, що

- ① потік має не перевищувати пропускну здатність для всіх ребер;
- ② сума потоків, що входять у вузол не повинна змінитись на виході;
- ③ потік завжди додатній.

Процес створення маски рухомих об'єктів алгоритмом Б-К

(a) Попередній кадр F^i (b) Поточний кадр F^{i+s} (d) Маска рухомих об'єктів на кадрі F^{i+s} (c) Інвертована різниця F^i і F^{i+s}

Згорткові нейронні мережі для локалізації людини



(e) Алгоритм Б-К



(f) yolov5n



(g) ssdlite320-mobilenet-v3



(h) fasterrcnn-mobilenet-v3

Алгоритм створення панорами

Вхід: два кадри F^i та F^{i+s} , поточна панорама W^i ($W^1 = F^1$).

Вихід: панорама W^{i+s} .

Етап отримання відповідних точок.

1. Знаходимо набір M^i пар відповідних пікселів між кадрами F^i і F^{i+s} і будуємо множину $M'^i = \{(p^i, p^{i+s}) \in M^i : B_{p^i}^i = B_{p^{i+s}}^{i+s} = 0\}$ тих пар відповідних пікселів, координати яких не належать області рухомих об'єктів.

2. Якщо $|M'^i| < 0.5 \cdot |M^i|$ або $|M'^i| < 4$, завершуємо алгоритм з результатом $W^{i+s} = W^i$.

3. Знаходимо набір M_W^i пар відповідних пікселів між панорамою W^i і кадром F^{i+s} і будуємо множину $M'_W^i = \{(p_W^i, p^{i+s}) \in M_W^i : \exists p^i \in P : (p^i, p^{i+s}) \in M'^i\}$.

Етап обчислення матриці гомографії.

4. На базі множини M'_W^i пар відповідних точок знаходимо матрицю H_W^i гомографії, що співставляє пікселі кадру F^{i+s} та панорами W^i .

Етап обчислення розміру нової панорами.

5. Рахуємо координати крайніх точок кадру F^{i+s} після застосування до них матриці H_W^i .
 $\vec{l}_1 = H_W^i \cdot (0, 0, 1)^T, \vec{l}_2 = H_W^i \cdot (w - 1, 0, 1)^T, \vec{l}_3 = H_W^i \cdot (0, h - 1, 1)^T, \vec{l}_4 = H_W^i \cdot (h - 1, h - 1, 1)^T$

Алгоритм створення панорами

6. Для визначення множини P_W^i нової панорами W^{i+s} рахуємо величини

$$x_{\min}^i = \min_{j=1,4} \frac{(l_j^i)_x}{(l_j^i)_z}, x_{\max}^i = \max_{j=1,4} \frac{(l_j^i)_x}{(l_j^i)_z}, y_{\min}^i = \min_{j=1,4} \frac{(l_j^i)_y}{(l_j^i)_z}, y_{\max}^i = \max_{j=1,4} \frac{(l_j^i)_y}{(l_j^i)_z}.$$

Позначимо $P_W^{i+s} = 1, \dots, \max(w^i, x_{\max}^i - x_{\min}^i) \times 1, \dots, \max(h^i, y_{\max}^i - y_{\min}^i)$

Етап створення нової панорами.

7. Будуємо панораму W^{i+s} . Для зручності позначимо обернену матрицю $H' = (H_W^i)^{-1}$, числа

$$x'_{\min} = \min(x_{\max}^i, 0), y'_{\min} = \min(y_{\max}^i, 0) \text{ і відображення } f' : p \rightarrow \left(\left| \frac{(H' \cdot p)_x}{(H' \cdot p)_y} \right|, \left| \frac{(H' \cdot p)_y}{(H' \cdot p)_z} \right| \right), \text{ що перетворює}$$

координати з панорами до пікселів кадру за допомогою гомографії. Інтенсивність у пікселі p панорами W^{i+s} визначається за формулою

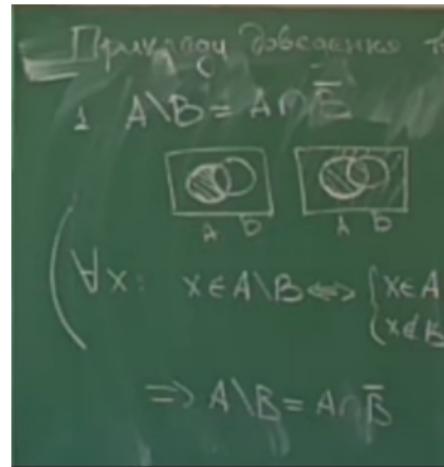
$$W^{i+s}(p) = \begin{cases} F^{i+s}(f(p)), & f(p) \in P, \\ W^i(p + (x'_{\min}, y'_{\min})), & p + (x'_{\min}, y'_{\min}) \in P^i, \\ 0, & p + (x'_{\min}, y'_{\min}) \notin P^i. \end{cases}$$

Обробка слайдів

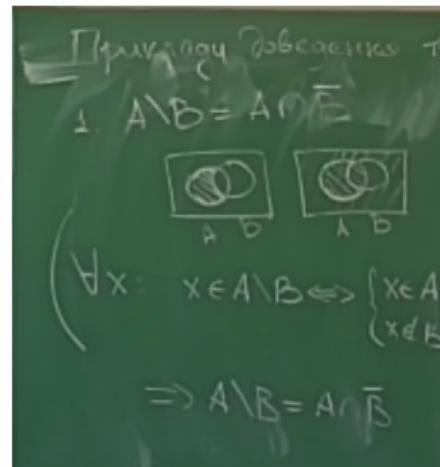


(i) Оператор Лапласа + Бінаризація Отсу

Обробка слайдів



(j) До швидкої медіани



(k) Після швидкої медіани

Conclusions

Now we have:

- stabilization based on homography estimation
- neural network for human detection
- creating slides

Future improvements:

- replace a neural network with faster method
- vectorization of the board
- improve stabilization