

## ЭЛЕКТРОННАЯ СТАБИЛИЗАЦИЯ ВИДЕОИЗОБРАЖЕНИЯ С НЕСТАЦИОНАРНОЙ КАМЕРЫ

Белорусский государственный университет информатики и радиоэлектроники  
г. Минск, Республика Беларусь

Пчёлкин А.С.

Черная И.И. – к.т.н., доцент каф. ЗИ

Техники видеостабилизации имеют принципиальное значение для большинства видеопоследовательностей снятых с нестационарных камер из-за высокочастотных помех. Некоторые алгоритмы стабилизации, основанные на 2D и 3D преобразованиях, хорошо изучены, однако было предложено мало решений на базе глубоких нейронных сетей. В данной статье предложен алгоритм видеостабилизации с использованием сверточных нейронных сетей.

Предложенная модель *StabNet* является сверточной нейронной сетью (СНС), которая обучается прогнозировать параметры трансформации для каждого входящего необработанного кадра на основе набора обработанных кадров. Применение полученных параметров трансформации к входному необработанному кадру генерирует стабилизированный выходной кадр. Стабилизированный кадр помещается в набор обработанных кадров для стабилизации последующих необработанных кадров. На рисунке 1 показана схема предложенной модели *StabNet*.

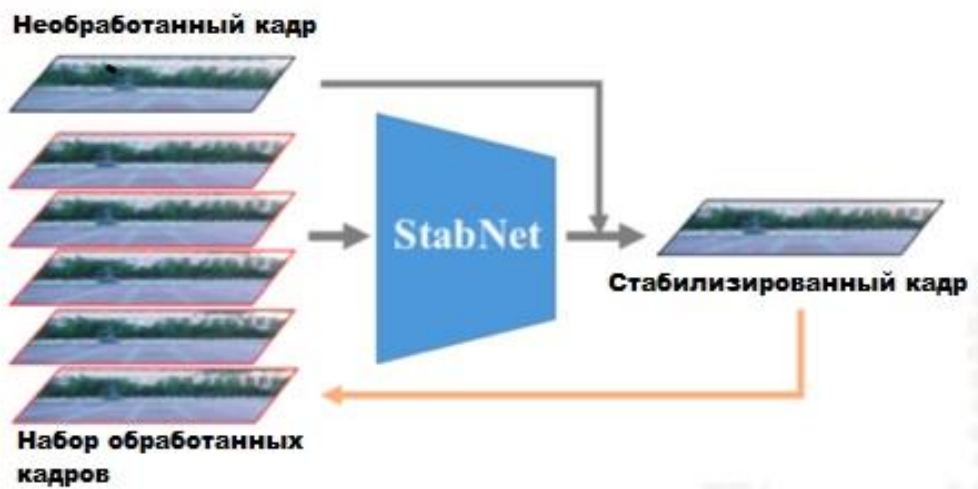


Рис.1 – Схема модели *StabNet*

Архитектура сети показана на рисунке 2. *StabNet* это сиамская сеть с двумя ветвями с обзими параметрами для обеих ветвей. Сеть состоит из кодировщика и гомографического регрессора. Гомографический регрессор это 8-канальный выходной сверточный слой с размером ядра 1×1 и шагом 1 пиксел(k1n8s1). Во время обучения, 2 последовательных кадра подаются на вход сети и прогнозируются параметры трансформации. При обучении используются функция потерь устойчивости и функция временных потерь.

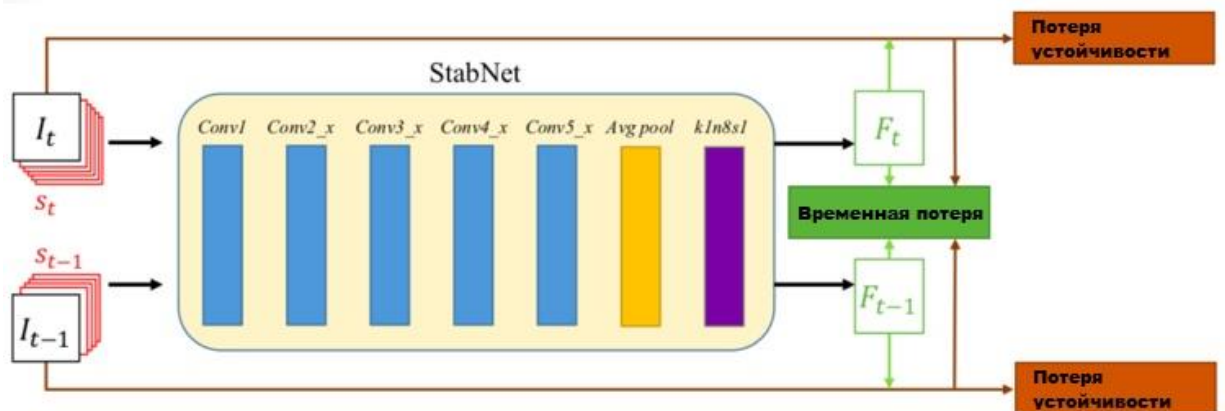


Рис. 2 – Архитектура сети

Функция потерь:

$$L = \sum_{i \in \{t, t-1\}} L_{stab}(F_i, I_i) + \delta L_{temp}(F_t, F_{t-1}, I_t, I_{t-1})$$

Функция потерь устойчивости:

$$L_{stab}(F_t, I_t) = L_{pixel}(F_t, I_t) + \alpha L_{feature}(F_t, I_t),$$

$$L_{pixel}(F_t, I_t) = \frac{1}{D} \|I'_t - F_t * I_t\|_2^2.$$

где  $I'_t$  – прямые наблюдения,

$D$  – размерность кадра.

$$L_{feature}(F_t, I_t) = \frac{1}{m} \sum_{i=1}^m \|p_t^i - F_t * p_t^i\|_2^2,$$

$$P_t = \{(p_t^i, p_t^i) | i \in \{1, \dots, m\}\},$$

где  $m$  – пары совпавших особых точек между необработанным/обработанным кадром.

Функция временных потерь:

$$L_{temp}(F_t, F_{t-1}, I_t, I_{t-1}) = \frac{1}{D} \|F_t * I_t - \omega(F_{t-1} * I_{t-1})\|_2^2,$$

где  $\omega(\cdot)$  - функция, преобразующая обработанный кадр t-1 к обработанному кадру t согласно заранее вычисленному оптическому потоку.

Результаты экспериментов показывают, что предложенная модель демонстрирует сравнимые результаты с более изученными методами, при этом работая в 30 раз быстрее. Также StabNet справляется с ночными и размытыми видеопоследовательностями, с которыми у традиционных методов часто возникают трудности.

Список использованных источников:

1. Ганина Я.В. Семантическая сегментация изображений на основе метода машинного обучения // Дипломная работа. - Москва, 2011-63с.

2. Фурман Я.А. Введение в контурный анализ; приложения к обработке изображений и сигналов / Я. А. Фурман, А. К. Кревецкий, А.К. Передреев // Научное издание. – Москва, 2003-592с