


Numer zlecenia oraz nazwa i akronim projektu: <i>System rozpoznawania przemocy z nagrań video - VRS</i>	Zleceniodawca: <i>PJATK</i>	Zleceniobiorca: 
Zespół projektowy: <i>Ola Piętka Jakub Kulaszewicz Benedykt Kościński Mateusz Chodyna</i>	Kierownik projektu: <i>Benedykt Kościński</i>	
Nazwa dokumentu: <i>Dokument Założeń Wstępnych</i>	Odpowiedzialny za dokument: <i>Ola Piętka</i>	Opiekun projektu: <i>prof. dr hab. Marek Bednarczyk</i>

Historia dokumentu

Wersja	Opis modyfikacji	Rozdział / strona	Autor modyfikacji	Data
1.0	Wersja wstępna	Całość	<i>Ola Piętka Jakub Kulaszewicz Benedykt Kościński Mateusz Chodyna</i>	26.04.2020
1.1	Poprawienie błędów	1, 2, 5, 8, 9	<i>Ola Piętka Jakub Kulaszewicz Benedykt Kościński Mateusz Chodyna</i>	11.05.2020
1.1.0.1	Poprawienie błędów	1, 8, 9	<i>Ola Piętka Jakub Kulaszewicz Benedykt Kościński Mateusz Chodyna</i>	13.05.2020

1 Opis problemu

Przemoc stanowi poważne zagrożenie dla bezpieczeństwa osobistego i stabilności społecznej. Jest ona obecna w miejscach publicznych, szkołach, domach jak i Internecie. Według statystyk dokonanych przez Bureau of Justice Statistics [1] w latach 2004-2008 na obszarze Stanów Zjednoczonych liczba zarejestrowanych aktów przemocy wynosiła około 5.5 mln, w tym ponad 1.6 mln wydarzyła się w miejscach publicznych, a ponad 700 tys. w szkołach. Dlatego niezwykle ważne jest automatyczne wykrywanie zdarzeń związanych z przemocą na podstawie ogromnej ilości danych wideo. Rozważając różne aplikacje, w tym adnotacje wideo, wyszukiwanie wideo i monitorowanie w czasie rzeczywistym, koncentrujemy się na trudnym zadaniu polegającym na wykrywaniu brutalnych działań w nagraniach wideo. To zadanie obejmuje wiele powiązanych technik wizji komputerowej, na przykład wykrywanie obiektów, rozpoznawanie akcji i klasyfikację. Wykrywanie przemocy ma na celu automatyczne i skuteczne określenie czy przemoc występuje w sekwencji wideo.

2 Cele systemu

Celem systemu jest klasyfikacja przemocy z nagrań wideo. Po zrealizowaniu spodziewamy się gotowego modułu klasyfikacji, który w 80 na 100 przypadków poprawnie rozpozna przemoc lub jej brak. Dzięki naszemu systemowi w miejscach jego używania zakładamy spadek aktów przemocy. System jest przeznaczony dla firm prywatnych pożądaną filtrowania nagrań wideo w celu wyłapywania aktów agresji, jak również miast i firm prywatnych chcących posiadać system raportowania aktów przemocy z kamer i możliwość ich szybkiej weryfikacji.

3 Kontekst systemu

Wyuczony model klasyfikacji przemocy który nie narzuca środowiska zastosowania oraz integracji z innymi systemami.

Naszym użytkownikiem jest klient docelowy. Pozwalamy użytkownikowi (klientowi) na dowolność w kwestii zastosowania systemu i dostosowania go do swoich potrzeb. Użytkownik ma dostęp do wyniku końcowego klasyfikacji z wideo.

Przykładem rozwiązania konkurencyjnego jest „Abto Violence Detection Technology” od firmy Abto Software, które również charakteryzuje się wykrywaniem aktów przemocy. Udało im się osiągnąć dokładność klasyfikacji na poziomie 91%.

4 Zakres systemu (funkcjonalność)

System ma jedną główną funkcjonalność, którą jest klasyfikacja przemocy z podanej wcześniej sekwencji wideo.

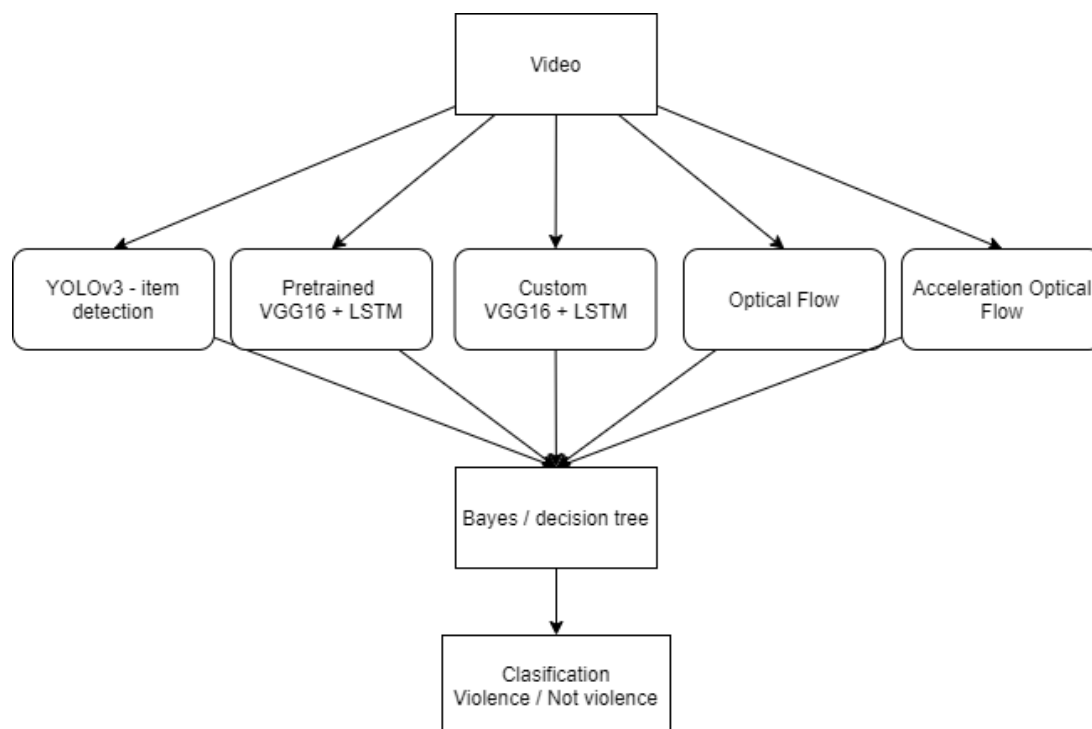
5 Wymagania jakościowe i inne

- Model klasyfikujący który w 80 na 100 przypadków poprawnie rozpozna przemoc lub jej brak.
- Klasyfikowanie w czasie bliskim rzeczywistego
- System musi mieć możliwość postawienia na lokalnym serwerze jak i chmurze

- System musi być elastyczny, tj. przyjmować nagrania wideo w różnych rozdzielczościach i formatach

6 Wizja konstrukcyjna

Zbudowanie systemu w języku Python na podstawie wielu wyuczonych modułów klasyfikujących akty przemocy na różne sposoby (tj. rozpoznawanie przedmiotów niebezpiecznych, znajdowanie cech przemocy oraz cech przyspieszenia i ruchu)



7 Ograniczenia

- Czas
- Wiedza
- Moc obliczeniowa
- Złożoność problemu
- Dataset

8 Słownik pojęć

Dataset – zbiór danych wideo potrzebnych do wyuczenia sieci

Moc obliczeniowa – ilość użytecznej pracy wykonanej przez system komputerowy

YOLOv3 – wariant algorytmu wykrywania obiektów „YOLO” [2]. Model rozpoznaje 80 różnych obiektów na zdjęciach i filmach

VGG16 – model splotowej sieci neuronowej [3].

LSTM – architektura rekurencyjnej sieci neuronowej (RNN) [4] stosowana w dziedzinie głębokiego uczenia się

Optical flow – wzór widocznego ruchu: obiektów, powierzchni i krawędzi w scenie wizualnej, spowodowany względnym ruchem między obserwatorem a sceną

9 Bibliografia

- [1] Bureau of Justice Statistics, „Detailed place of occurrence for violent and property crimes, average annual 2004-2008,” [Online]. Available: <https://www.bjs.gov/index.cfm?ty=tp&tid=44>.
- [2] J. Redmon i A. Farhadi, „YOLOv3: An Incremental Improvement,” 2018. [Online]. Available: <https://pjreddie.com/media/files/papers/YOLOv3.pdf>.
- [3] K. Simonyan i A. Zisserman, „Very Deep Convolutional Networks for Large Scale Image Recognition,” 2015. [Online]. Available: <https://arxiv.org/pdf/1409.1556.pdf>.

- [4] S. Hochreiter i J. Schmidhuber, „LONG SHORT-TERM MEMORY,” 1997. [Online]. Available: <https://www.bioinf.jku.at/publications/older/2604.pdf>.