

Podsłuchiwanie obrazem

7 sierpnia 2014

Ekspersi z MIT-u, Microsoftu i Adobe stworzyli algorytm, który potrafi odtworzyć dźwięk za podstawie analizy drgań obiektu widocznego na filmie. Podczas eksperymentów algorytm odtworzył zdania analizując vibracje paczki chipsów, którą filmowano z odległości 4,5 metra przez szybę nieprzepuszczającą dźwięku. W czasie innych eksperymentów dźwięk został odtworzony z obrazu drgań folii aluminiowej, wody oraz liści rośliny doniczkowej.

„Gdy dźwięk uderza w obiekt, wywołuje jego vibracje. Dostarczają one subtelnych informacji wizualnych, które są zwykle niewidoczne dla gołego oka. Ludzie nie zdają sobie sprawy z obecności tych informacji” – mówi magistrant Abe Davis z MIT-u. Wraz z profesorami Fredo Durandem i Billem Freemanem napisał on artykuł, który zostanie zaprezentowany podczas konferencji Siggraph. W pracy nad algorytmem brali też udział Neal Wadhwa z MIT-u, Michael Rubinstein z Microsoft Research oraz Gautham Mysore z Adobe Research.

Rekonstrukcja audio z obrazu wideo wymaga, by częstotliwość próbek wideo była wyższa niż częstotliwość sygnału audio. To z kolei oznacza, że konieczne jest wykorzystanie szybkich kamer. Podczas niektórych eksperymentów uczeni korzystali z urządzeń wykonujących 2000-6000 ujęć w ciągu sekundy. To znacznie więcej niż 60 klatek na sekundę (fps), którymi mogą pochwalić się kamery w zaawansowanych smartfonach. Jednak technologia, której użyli uczeni jest dostępna dla przeciętnego użytkownika. Najlepsze komercyjnie dostępne szybkie kamery wykonują nawet 100 000 klatek na sekundę.

Co ciekawe, podczas części eksperymentów użyto zwykłych kamer cyfrowych pracujących z szybkością 60 fps. W tym wypadku rekonstrukcja nie była tak dobra, jak przy użyciu szybkich kamer, jednak wystarczająca, by określić płeć rozmówców, ich liczbę oraz – o ile mamy wystarczająco dużo informacji na

temat właściwości akustycznych głosów – określić tożsamość rozmówców.

Oczywiście pierwszymi, które chętnie wykorzystają wspomniane prace, będą z pewnością służby specjalne i organy ścigania. Jednak naukowców interesuje coś innego. Dla nich nowy algorytm to okazja do poznania właściwości różnych obiektów. „Odtwarzamy dźwięk na podstawie obrazu obiektu. To daje nam dużo informacji o dźwięku, ale również o samym obiekcie, gdyż różne obiekty w różny sposób reagują na fale dźwiękowe” – mówi Davis. Uчени już rozpoczęli serię eksperymentów, podczas których starają się określić właściwości materiałów i ich strukturę na podstawie nagrań wideo ich reakcji na dźwięk. Mierzyli też ruch obiektu w rozdzielczości sięgającej dziesiątych części mikrometra. Uważają, że możliwe będzie nawet odnotowanie przesunięcia mniejszego niż piksel na obrazie. „Wystarczy” jedynie zauważyć przesunięcie piksela koloru.

Naukowcy już wykorzystali swoją pracę w praktyce. Połączyli swój algorytm z techniką wykorzystaną w algorytmach wzmacniających niewidoczne różnice w obrazach wideo i stworzyli w ten sposób algorytm, który pozwala na optyczne monitorowanie oddechu i pulsu noworodków w szpitalach.

Autor: Mariusz Błoński

Na podstawie: MIT

Źródło: [Kopalnia Wiedzy](#)