

# Rozpoznawanie emocji w filmach i grach

Filip Rynkiewicz

**Streszczenie**—W ciągu ostatnich dekad rozwój technologii tworzenia gier i filmów doprowadził do potrzeby sprawdzenia reakcji widza na obraz. Organizm człowieka reaguje na bodźce zewnętrzne, poprzez mikrozmiany twarzy, zmiany w EEG czy źrenicy oka. Procesy te mogą być odbierane przez odpowiednie urządzenia, zatem zautomatyzowanie poprawnej analizy tych danych jest możliwe. Dzięki temu twórcy są w stanie sprawdzić reakcję widza na ich obraz, by w końcu stworzyć programy które same sprawdzały by jakie emocje wywołują.

**Słowa kluczowe**—rozpoznawanie emocji, odruch źrenicy, EEG, analiza obrazu, klasyfikacja emocji

## 1 WSTĘP

Emocje pozwalają człowiekowi zdecydować czy dany obraz im się podoba czy nie. Daje to możliwość decydowania czy widz obejrzy go do końca, bądź nawet będzie chciał te emocje powtórzyć. Dla twórcy taka informacja jest bardzo ważna, ponieważ pozwala dopracować dzieło pod wpływem reakcji widza. Dzięki takim badaniom twórca jest w stanie znaleźć momenty w których widz będzie bardziej przejęty akcją, co poskutkuje jego większym zainteresowaniem.

Emocje są odczuwane według pewnego algorytmu [1]. Najpierw następuje postrzeżenie wydarzenia. Następnie jego analiza w kontekście własnych doświadczeń i norm, tak aby ostatecznym wynikiem była klasyfikacja wydarzenia w jakąś emocję.

Do rozpoznawania emocji zostały przystosowane różne sygnały, które można zakwalifikować do dwóch grup [2] takich jak:

- psychologiczne
  - EEG(Elektroencefalografia),
  - EMG(Elektromiografia),
  - EKG(Elektrokardiografia),
  - rozmiar źrenicy.
- nie-psychologiczne:
  - tekst,
  - mowa,
  - gesty,
  - mimika twarzy.

## 2 EEG

Liczne badania dotyczące ewaluacji sygnałów EEG, które rejestrują działalność mózgu w centralnym systemie nerwowym, pokazały że podczas oddziaływania emocji na nasze ciało mózg wysyła charakterystyczne informacje, dzięki którym można rozpoznać daną emocję [4] [5]. Poniżej zostanie przedstawione dwa podejścia do przetwarzania sygnału EEG.

W [3] na głowie badanego zostało umieszczone 32 elektrody w systemie 10-20, sygnały nagrywane w 1,024Hz zostały zmniejszone do 256 Hz, aby zmniejszyć koszty oraz oszczędzić pamięć. W przetwarzaniu tego sygnału zostało wzięte pod uwagę tylko **PSD**<sup>1</sup> z różnych częstotliwości, obliczona przy pomocy **FFT** oraz algorytmu **Welcha**, który powoduje wygładzenie spektrum. W tym podejściu zostały wykorzystane fale z zakresów theta (4 Hz < f < 8 Hz), slow alpha (8 Hz < f < 10 Hz), alpha (8 Hz < f < 12 Hz), beta (12 Hz < f < 30 Hz), oraz gamma (30 Hz < f) Do obliczeń uwzględniono także potencjalną asymetrię działania mózgu.

W drugim podejściu [6] użyto rozkładu elektrod 10-20, sygnał był zapisywany z częstotliwością 1000Hz. Na dane został nałożony filtr pasmowy pomiędzy 0.5Hz a 70Hz oraz sygnał został zmniejszony do 200Hz. Następnie został użyty **STFT** z oknem **Hanninga** 4s. Tym razem, oprócz **PSD**, zostały wzięte pod uwagę także **DE**<sup>2</sup>, **DAMS**<sup>3</sup> oraz **RASM**<sup>4</sup>. **DE** zostało obliczone za pomocą:

$$h(X) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \log \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \frac{1}{2} \log 2\pi e \sigma^2 \quad (1)$$

gdzie X jest rozkładem gausa  $N(\mu, \sigma^2)$ , x jest zmienną a  $\pi$  oraz e są stałymi. **DASM** oraz **RASM** są obliczane w następujący sposób:

$$DASM = h(X_{LEFT}) - h(X_{RIGHT}) \quad (2)$$

$$RASM = h(X_{LEFT})/h(X_{RIGHT}) \quad (3)$$

gdzie  $X_{LEFT}$  są **DE** dla lewej półkuli mózgu, a  $X_{RIGHT}$  z prawej. W tym badaniu wykorzystano fale delta (1Hz-3Hz), theta (4Hz-7Hz), alpha (8Hz-13Hz), beta (14Hz-30Hz) oraz gamma (31Hz-50Hz).

• Politechnika Łódzka, Łódź, Polska,  
filip.rynkiewicz@dokt.p.lodz.pl

1. Power Spectral Density  
2. Differential Entropy  
3. Differential Assymetry  
4. Rational Assymetry

### 3 ŻRENICIA OKA

Oko ludzkie może dostarczyć dużo informacji. Jedną z nich są odczuwane przez człowieka emocje. Możliwe jest to dzięki pomiarom aktywności żrenicy oka, a w szczególności jej średnicy. Urządzenia wykorzystywane do takich badań to *Eye-Trackers*. Zmiany te są jednak zależne także od światła padającego oraz mogą być zaszumiane poprzez mruganie.

Aby zmniejszyć wpływ światła na badania w obydwu przypadkach zostały zbudowane modele refleksji przy pomocy metod statystycznych. Zakładając że  $Y$  jest macierzą  $M \times N$ , reakcji żrenicy dla tego samego obrazu dla  $N$  badanych przy  $M$  próbkach,  $Y$  będzie składało się z trzech parametrów.

$$Y = A + B + C \quad (4)$$

gdzie  $A$  to wpływ światła na siatkówkę,  $B$  wpływ emocji,  $C$  szum. Aby wykluczyć wpływ światła oraz szumów skorzystano z PCA<sup>5</sup> i wyznaczono  $Y_{rest} = Y - Y_1$  gdzie  $Y_{rest}$  jest reakcją żrenicy na emocje. W obydwu przypadkach zbieranie informacji o reakcji oka były takie same.

### 4 KLASYFIKACJA EMOCJI

Po zebraniu danych od użytkowników należy je sklasyfikować jako emocje. Do tego celu, w obydwu przypadkach, zostały użyte SVM<sup>6</sup> jako klasyfikatory. SVM poddawane są uczeniu a następnie można wprowadzać dane które klasyfikator ma skategoryzować.

### 5 ŁĄCZENIE METOD

Używając różnych metod zostało sprawdzone która z nich najlepiej połączy dane EEG oraz ze żrenicy oka. Metoda *Feature Fusion* w Tabeli.1 jest najlepszą z wynikiem 73.59% poprawności. Dla Tabeli.2 najdokładniejszą metodą okazała się metoda *Decision Level Fusion* która dała wyniki 76.4%.

Tabela 1: Dokładność metod dla [6]

	EEG (DE)	Max Strategy	Sum Strategy	Feature Fusion
1	83.09	83.09	83.09	<b>93.59</b>
2	68.22	68.22	51.31	<b>78.72</b>
3	68.22	67.93	51.02	<b>68.22</b>
4	85.13	68.22	<b>85.13</b>	83.97
5	51.31	51.31	51.31	<b>77.55</b>
6	83.09	<b>83.09</b>	<b>83.09</b>	<b>83.09</b>
7	51.31	<b>68.22</b>	<b>68.22</b>	58.02
8	83.09	83.09	83.09	<b>83.38</b>
9	68.22	<b>83.09</b>	68.22	63.56
10	68.22	68.22	68.22	<b>69.10</b>
11	<b>68.22</b>	<b>68.22</b>	<b>68.22</b>	40.82
12	<b>83.09</b>	<b>83.09</b>	65.89	<b>83.09</b>
Mean	71.77	72.98	68.90	<b>73.59</b>
Std	12.03	<b>10.09</b>	12.85	14.43

5. Principal Component Analysis

6. Support Vector Machine

Tabela 2: Dokładność metod dla [3]

Modality dimension	Classification rate		Average F1	
	arousal	valence	arousal	valence
EEG	62.1%	50.5%	0.60	0.50
Eye gaze	71.1%	66.6%	0.71	0.66
Feature level fusion (FLF)	66.4%	58.4%	0.65	0.55
Decision level fusion (DLF)	76.4%	68.5%	0.76	0.68
Self-reports with SAM manikins	55.7%	69.4%	0.57	0.70
Random level	33.3%	33.3%	0.36	0.40

### 6 PODSUMOWANIE

Emocje są skomplikowanym mechanizmem w ciele człowieka, jednakże ich odczytywanie może być pomocne w wielu dziedzinach. Analiza sygnałów pozwala na sklasyfikowanie jakie emocje w danej chwili zostały wykryte, oraz stworzyć narzędzia do automatycznego wykrywania emocji w filmach bez uczestnictwa człowieka. W opisanych artykułach najlepszym wynikiem było 76.4% dokładności, co pozwala z dużym prawdopodobieństwem założyć że wynik takiej analizy będzie poprawny.

### LITERATURA

- [1] A. Ortony, G.L. Clore, A. Collins *The Cognitive Structure of Emotions.*, Cambridge University Press, July 1988.
- [2] R. A. Calvo, S.D'Mello *Affect detection: An interdisciplinary review of models, methods and their applications*, IEEE Transactions on Affective Computing, vol 1, no1, pp 18-37, 2010
- [3] M. Soleymani, M. Pantic, T. Pun *Multimodal Emotion Recognition in Responce to Videos*, IEEE Transaction on Affective Computing, vol. 3, no. 2, april-June 2012
- [4] R. Adolphs, D. Tranel, A.R. Damasio, *Dissociable Neural Systems for Recognizing Emotions*, Brain and Cognition, vol. 52, no. 1, pp. 61-69, June 2003.
- [5] A.R. Damasio, T.J. Grabowski, A. Bechara, H. Damasio, L.L.B. Ponto, J. Parvizi, and R.D. Hichwa, *Subcortical and Cortical Brain Activity during the Feeling of Self-Generated Emotions*, Nature Neuroscience, vol. 3, no. 10, pp. 1049-1056, Oct. 2000.
- [6] Z. Wei-Long, D. Bo-Nan, L. Bao-Liang *Multimodal Emotion Recognition using EEG and Eye Tracking Data*, IEEE, 2014