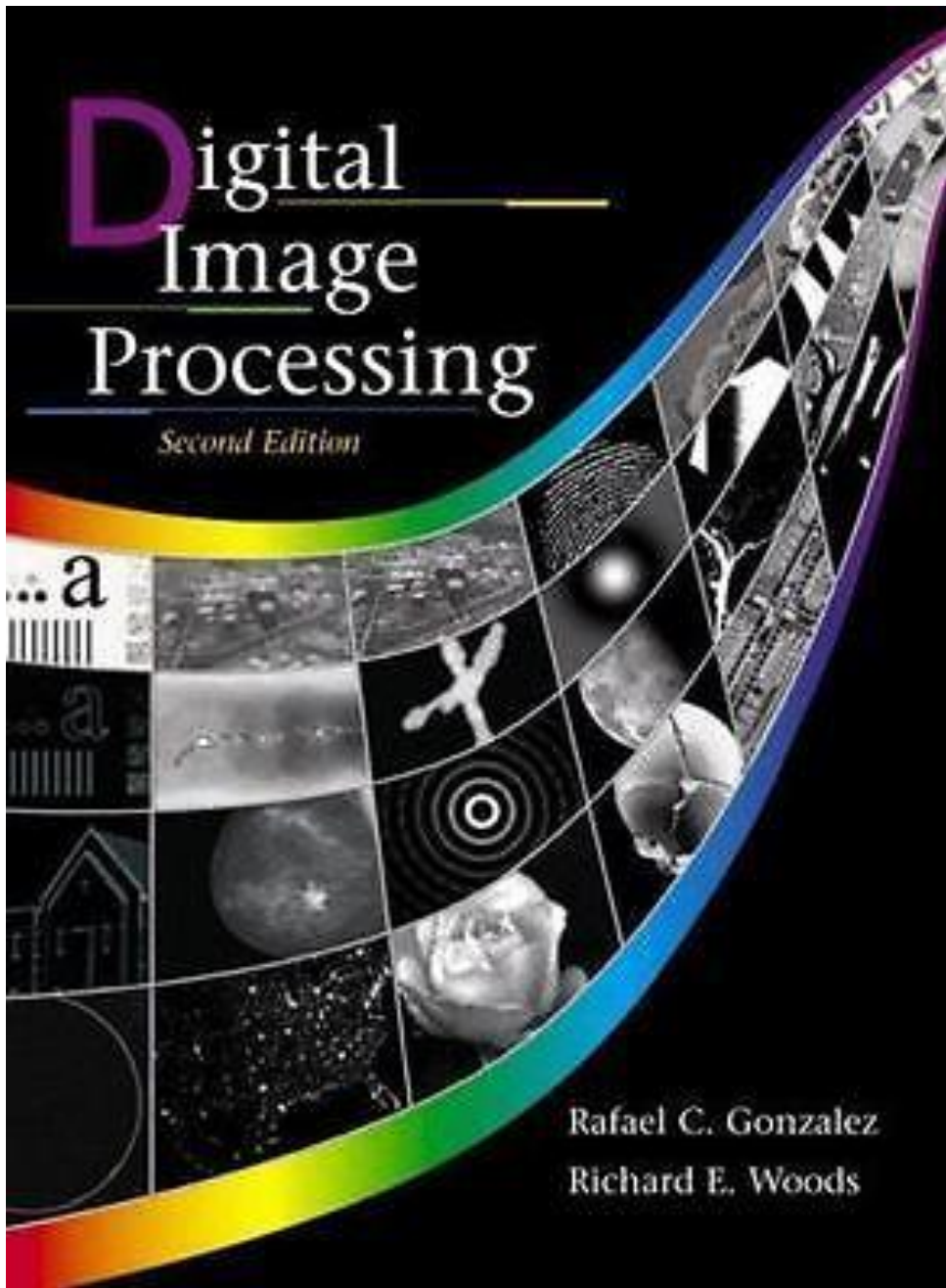


# Ping Pong Tracker using Computer Vision

Vîrtopeanu Sebastian



# Digital Image Processing

Rafael C Gonzalez  
Richard E. Woods

# Demo:

<https://www.youtube.com/watch?v=4sYsQbKgLV0>



Demo: <https://youtu.be/wg4dRa-nRz4>



# Cum am obținut aceste rezultate?

1. Identificarea și urmărirea traiectoriei mingii de ping-pong
2. Detectarea și delimitarea mesei de ping-pong
3. Sincronizarea video-urilor pe baza analizei sunetului
4. Combinarea datelor pentru reconstrucția traiectoriei în spațiul 3D

## Identificarea și urmărirea traiectoriei mingii de ping-pong

- Detectarea unei mingi de ping-pong pare, la prima vedere, un task simplu pentru noi, motiv pentru care am ales să aplicăm tehnici de computer vision pe acest obiect.
- Totuși, s-a dovedit mai dificilă decât ne-am așteptat, din cauza mai multor factori:
  - distorsionarea mingii la viteze mari
  - iluminarea artificială cu frecvențe de 50-60 Hz, care necesită o cameră de cel puțin 120 Hz pentru a respecta teorema Nyquist
  - prezența unor obiecte similare în fundal
  - reflexiile de pe suprafețe precum geamurile.





















- Astfel, tehnicile clasice de Computer Vision nu sunt suficiente pentru a detecta o minge de ping-pong cu precizie. O mare parte din încercările mele s-au concentrat pe:
- Analiza diferenței dintre frame-uri consecutive.
- Filtrarea pe baza culorii în spațiul HSV și extragerea formelor rotunde.
- Din acest motiv, am recurs la utilizarea rețelelor neuronale preantrenate, în special YOLO, pentru a îmbunătăți detectarea.

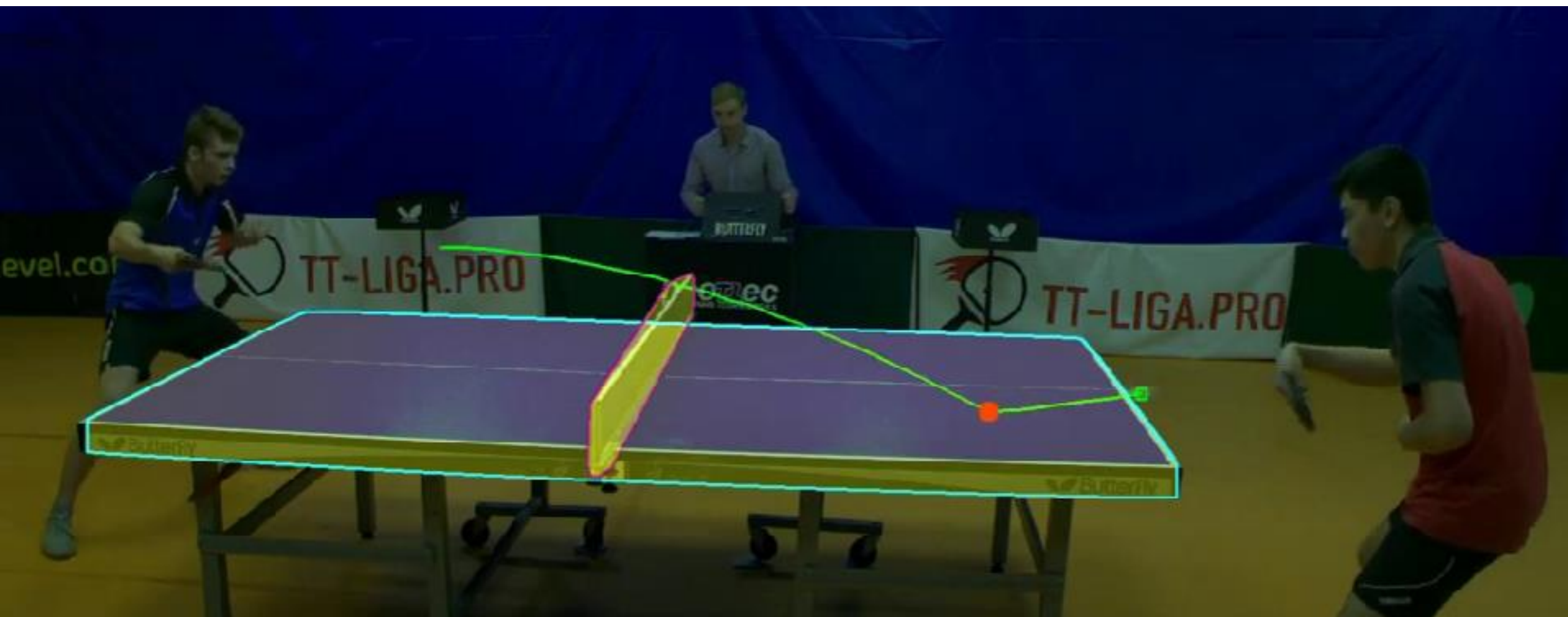






# Crearea traiectoriei mingii

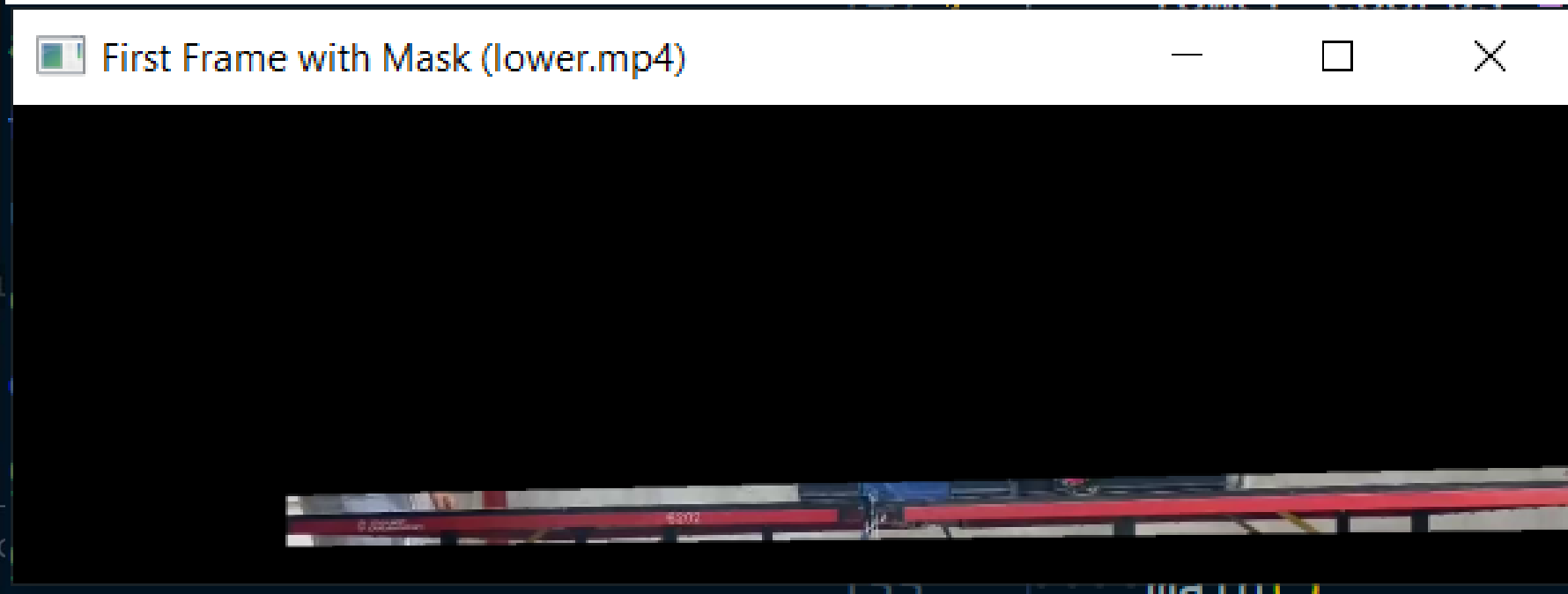
- Acum, având un instrument capabil să detecteze mingea de ping-pong în majoritatea imaginilor, pot trece la procesarea traiectoriei. Aceasta se realizează utilizând o coadă în care salvez coordonatele mingii pentru fiecare frame.
- În cazul în care mingea nu este detectată într-un anumit frame, nu este o problemă. Pot conecta punctele discrete existente pentru a genera segmente de dreaptă care să sugereze o traiectorie continuă. Totuși, a fost nevoie de un plus de atenție aici, deoarece nu mi-am dorit să trasez o simplă linie, ci să creez o reprezentare mai realistă și naturală a traiectoriei.



## Detectarea și delimitarea mesei de ping-pong

- În prima parte a proiectului, filtrarea în spațiul HSV pentru detectarea mingii de ping-pong a funcționat bine, atât pentru perspectivele de sus, cât și din lateral. Cu ajutorul unor convoluții, eroziuni și dilatări, am reușit să obțin rezultate satisfăcătoare.
- Totuși, problema a apărut atunci când am testat pe setul de date de la OSAI, unde performanța a scăzut considerabil. Aceasta s-a datorat, în principal, necesității de a detecta și plasa, pe lângă mingea de ping-pong. Din acest motiv, am optat pentru utilizarea SAM implementat cu GDINO, combinat cu diverse prelucrări suplimentare ale coordonatelor, pentru a obține o detecție mai precisă.





# Masca obtinuta din SAM cu GDINO



# Aplicarea operației de eroziune asupra măștii



Masca finală obținută după aplicarea dilatării,  
extragerea contururilor și unificarea celor două  
contururi într-unul singur.



Masca obținută prin diferența.

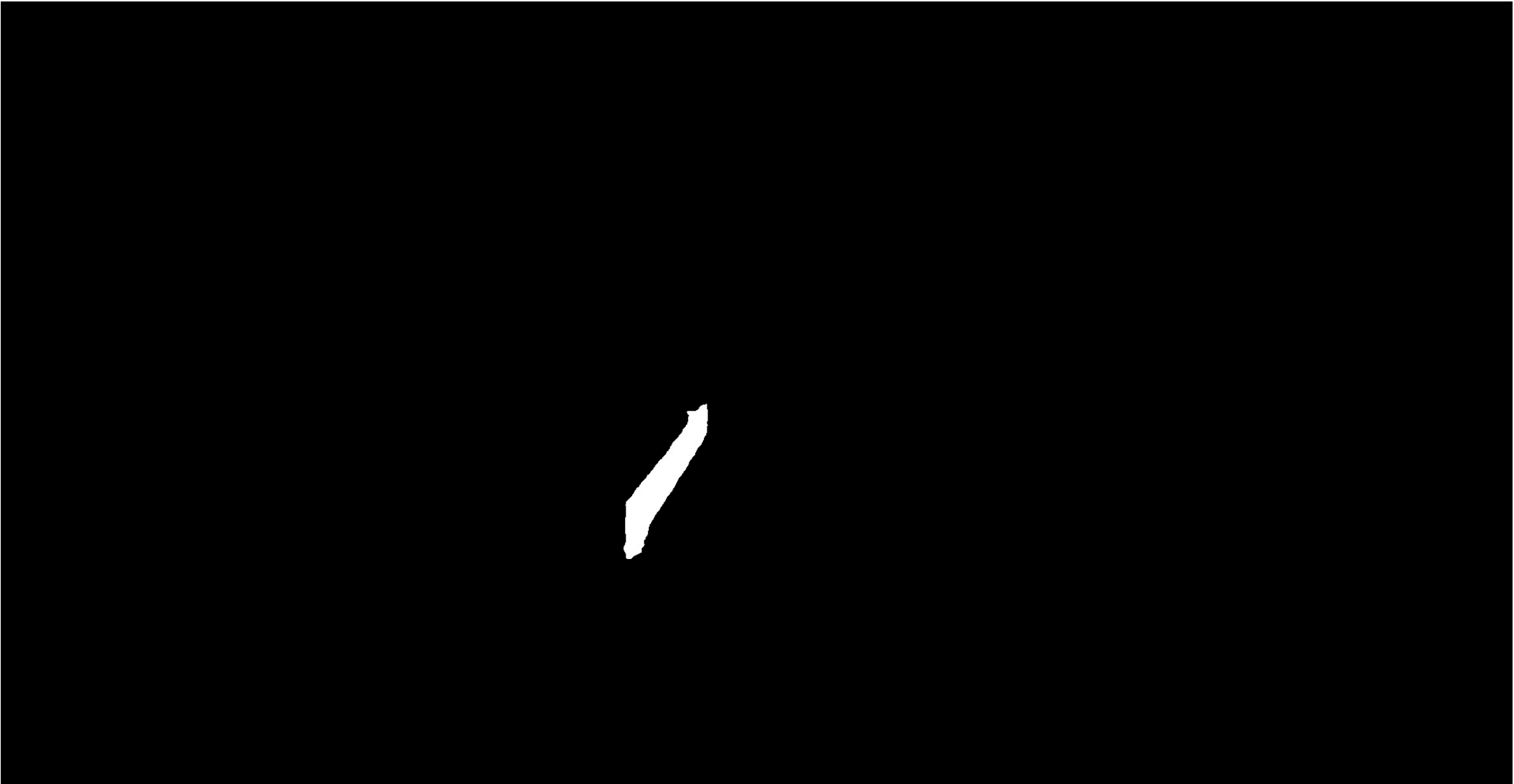


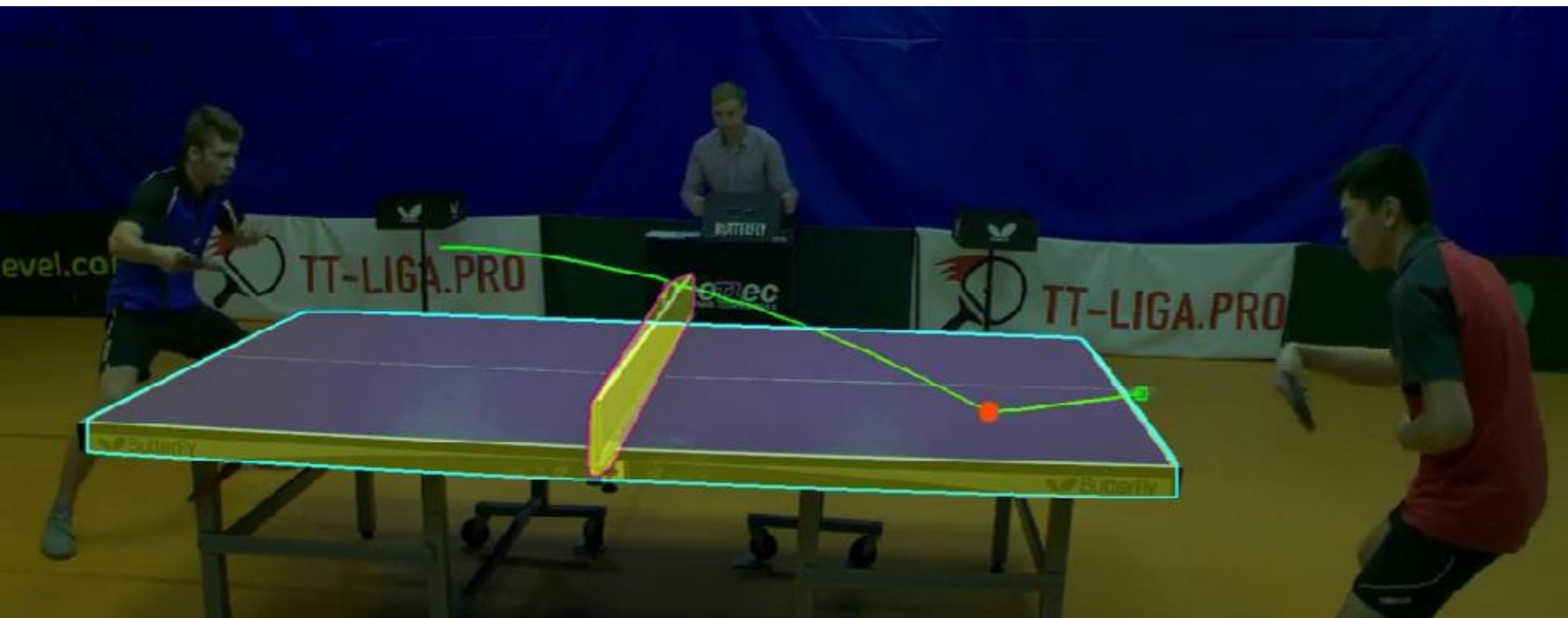


# Afișarea combinată a celor două măști



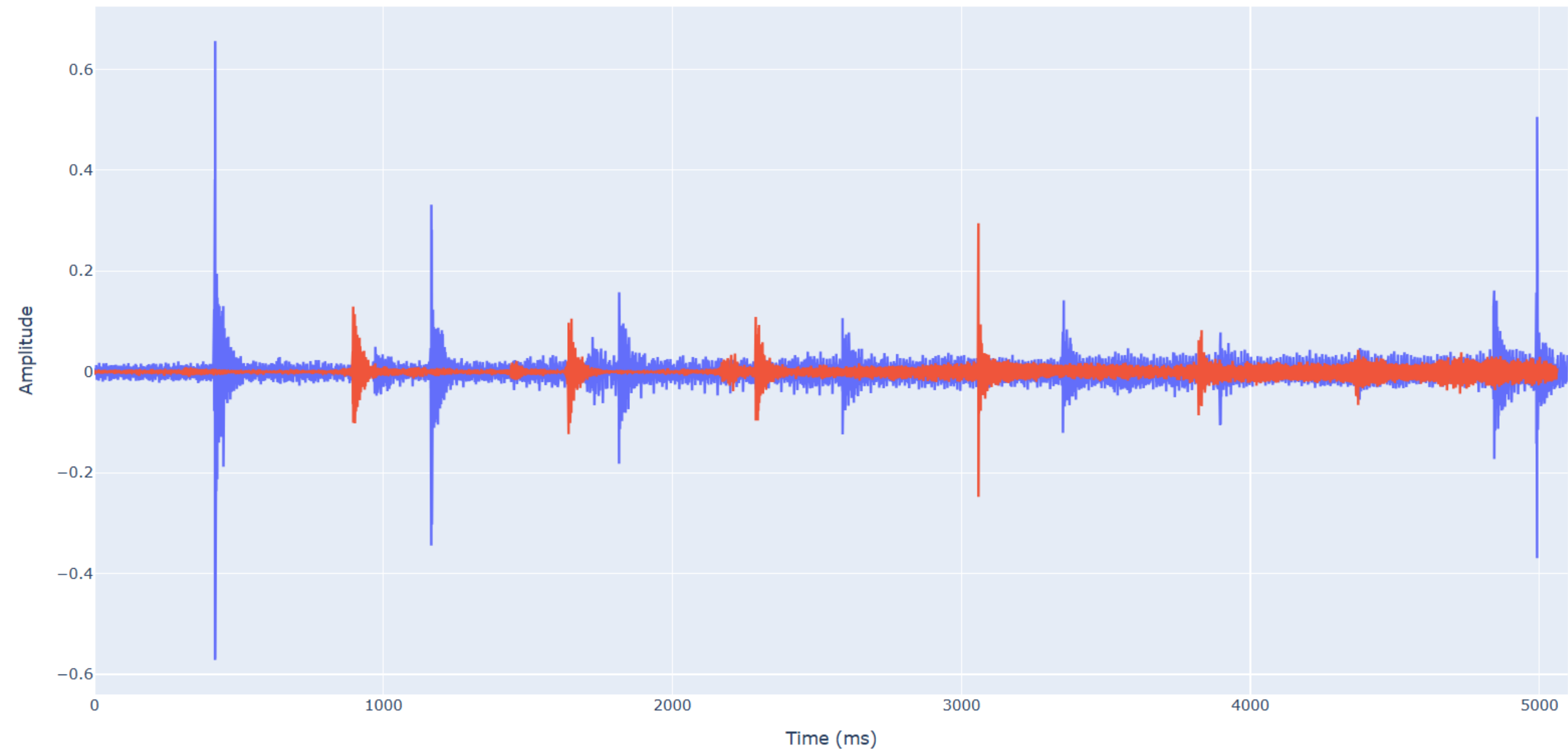
Extragerea centroidului și generarea unei noi măști rafinate utilizând SAM.



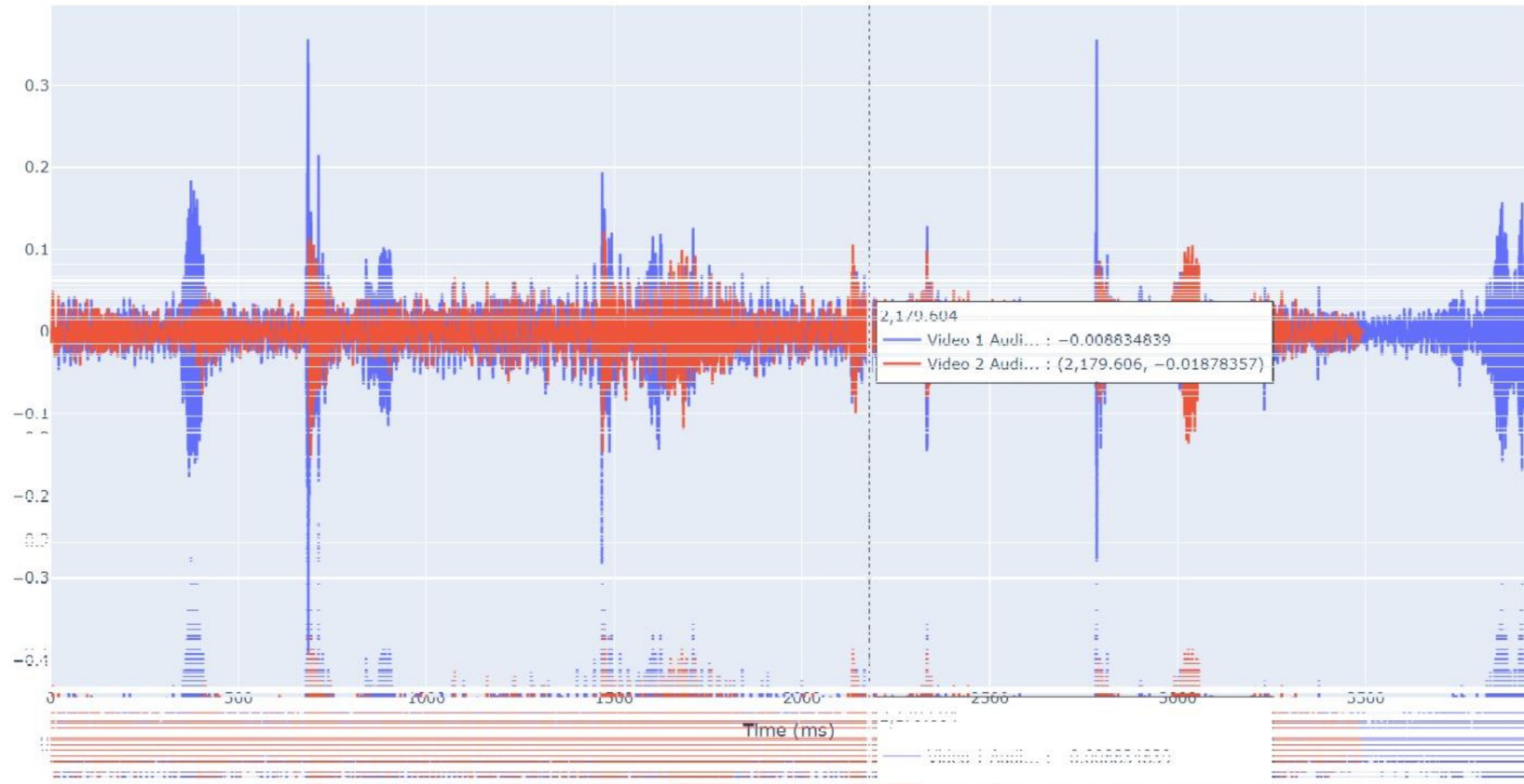


## Sincronizarea video-urilor pe baza analizei sunetului

- O problemă pe care am întâlnit-o în analiza unui meci de ping-pong din două perspective a fost sincronizarea acestor perspective.
- O soluție pe care am explorat-o a fost analiza semnalelor audio din cele două videoclipuri pentru a determina un offset care să ofere cel mai bun cross-correlation. În acest proces, a fost necesar să reduc zgomotul audio pentru a obține un semnal mai clar și a îmbunătăți precizia sincronizării.







Video 1 Audio (True)

Video 2 Audio (Noise)

Video 1 Audio (True)

Video 2 Audio (Noise)

2,179.604

Video 1 Audio (True) : -0.008834839

Video 2 Audio (Noise) : (2,179.606, -0.01878357)

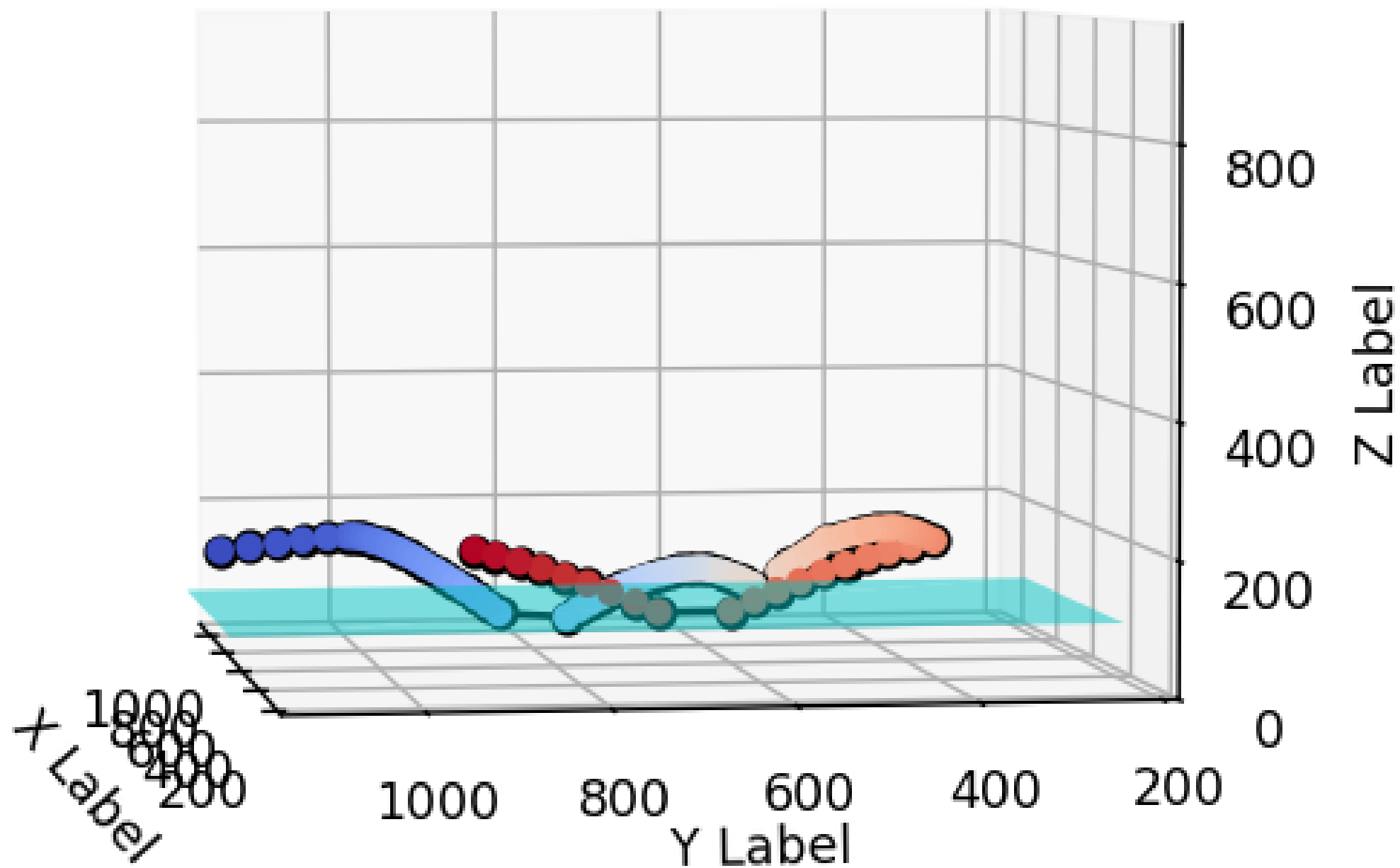
Time (ms)

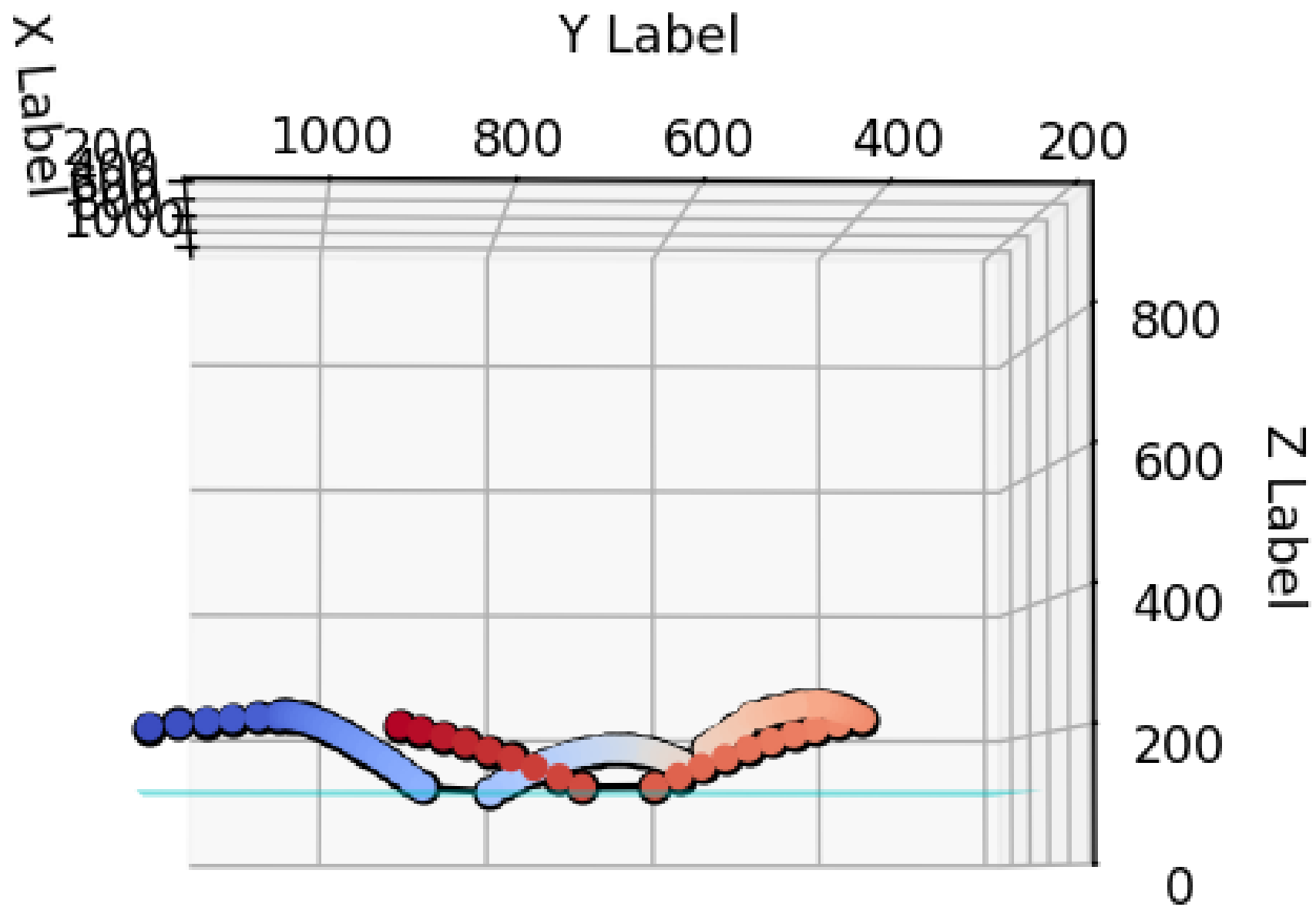
Video 1 Audio (True) : -0.008834839

Video 2 Audio (Noise) : (2,179.606, -0.01878357)

## Combinarea datelor pentru reconstrucția traiectoriei în spațiul 3D

- O primă idee care mi-a venit după ce am realizat analiza din cele două perspective a fost să transform coordonatele în 3D. Astfel, pentru videoclipul filmat din lateral, extrag coordonata Z în raport cu înălțimea mesei de ping-pong, iar pentru cel filmat de deasupra, extrag coordonatele X și Y, având ca referință colțul din dreapta sus al mesei.





Sfârșit

