

Sistem za sledenje govornicu v zaprtih prostorih na osnovi termovizije v realnem času

Mark Breznik, izr. prof. dr. Matevž Pogačnik, as. mag. Klemen Pečnik

Univerza v Ljubljani, Fakulteta za elektrotehniko, Tržaška cesta 25, 1000 Ljubljana

E-pošta: mark.breznik@lmmfe.org, matevz.pogacnik@lmmfe.org, klemen.pecnik@lmmfe.org

Real-Time Indoor Speaker Tracking System

Abstract. The article describes the design and development of a real-time indoor positioning and tracking system using a thermal camera and a micro-controller. During the COVID-19 pandemic, there has been an increase in the use of distance learning for lectures. Distance learning requires a reliable audio-video system that does more than just record video and audio. With the lack of affordable and reliable autonomous indoor positioning and tracking systems for the educational environment, we set out to create a robust but not cost-prohibitive solution. Throughout the article, we detail the base hardware technologies required for streaming lectures, the different types of autonomous indoor tracking, including their benefits and drawbacks, analyze the differences between a complex ML-powered and a simple highest-temperature approach to thermal tracking. The article is concluded with the strengths and weaknesses of the major systems.

1 Uvod

V članku je predstavljen razvoj sistema za usmerjanje pogleda kamer(e) v zaprtih prostorih v realnem času s pomočjo uporabe mikroročunalnika in termovizijske kamere. Poleg same predstavitve razvoja in implementacije sistema za sledenje na osnovi termovizijske detekcije je opisana in predstavljena tudi oprema za snemanje predavanj, različne vrste oz. pristopi sledenja govornicu, integracija sistema v obstoječe produkcijske sisteme in analiza ter predvideni koraki za nadaljnji razvoj.

V času epidemije je prišlo do porasta izvajanja poučevanja na daljavo, kar je posledično vodilo k preoblikovanju načina izvajanja predavanj ter povečalo potrebe po učinkovitih audio-video sistemih, ki nudijo več kot samo statičen zajem slike in zvoka.

Pri izvedbi predavanj na daljavo je zelo pomembno, da je tehnologija učinkovita in enostavna za uporabo, tako da je za predavatelja praktično neopazna, nemoteča in zahteva čim manj dodatnih korakov pri sami izvedbi. Pri profesionalnih športnih prireditvah se tehnologija za avtonomno sledenje uporablja že desetletja [1], medtem ko za manjše produkcije konfiguracija vse potrebne opreme predstavlja izziv, ki zahteva previsok finančni vložek in specifična znanja s področja multimedije.

Projekt je bil zasnovan v času epidemije, ker je v tem obdobju primanjkovalo ustreznih in cenovno dostopnih rešitev, ki bi omogočale integracijo nove opreme z obstoječo brez večjih posegov v samo infrastrukturo. Cilj

projekta je bil izdelati cenovno dostopen in zanesljiv sistem za določanje lokacije predavatelja v predavalnici ter s pomočjo aplikacijskih vmesnikov gibljivih kamer ustrezno upravljati smer kamere in njeno goriščno razdaljo (ang. Pan, Tilt, Zoom camera – PTZ camera). Prototip rešitve se preizkuša v različnih pogojih in kombinacijah, ki so najbolj značilne za predavatelje na Fakulteti za Elektrotehniko, Univerze v Ljubljani.

2 Oprema za snemanje predavanj



Slika 1: Panasonic AW-HE42 PTZ kamera [3]

Pri izvajanju predavanj na daljavo ni možno imeti samo ene rešitve za zajem AV vsebin, saj je vsaka predavalnica unikatna v mnogih pogledih. Posebej je potrebno paziti na velikost predavalnice, saj bi uporaba navadne spletne kamere v večini večjih prostorov enostavno prikazala le oblike in neprepoznavne krivulje. Prav tako je potrebno paziti na dinamični razpon kamer, saj je razlika med najsvetlejšo (npr. projektor) in najtemnejšo točko (npr. tabla pod projektorjem) ogromna in posledično večina cenejših kamer ne zmore prikazati celotnega razpona slike [2].

V zadnjih letih je močno narasla uporaba PTZ kamer za različne dogodke, manjše produkcije, predavanja, vaje itd. Velika prednost takih kamer je možnost motoriziranega premikanja, nagibanja in približevanja pogleda, kar pomeni, da ni potrebna fizična prisotnost kamermana v predavalnici, saj je kamero mogoče krmiliti na daljavo. Prav tako večina PTZ kamer, kot so npr. Panasonic AW-HE42, AW-UE100 [3] in HuddleCamHD SimpliTrack2 [4], nudi veliko boljšo kvaliteto slike ob različnih svetlobnih pogojih, predvsem zaradi večjih senzorjev in kvalitetnejših objektivov.

Večina PTZ kamer nudi HDMI (ang. High Definition Multimedia Interface), SDI (ang. Serial Device Interface), USB (ang. Universal Serial Bus), NDI (ang. Network Device Interface) in NDI|hx (ang. Network Device Interface high efficiency) video izhode, preko katerih lahko posredujejo zvok in sliko do različnih platform za učenje na daljavo, kot so Zoom, Microsoft Teams, Cisco WebEx in eFePlusMM.

PTZ kamere, kot je Panasonic AW-HE42, ki je prikazana na Slika 1, sodijo v cenovni razred 4.000 €, kar je za izobraževalne ustanove kar precejšnji zalogaj in postavi relativno visok prag vstopa. Obstajajo tudi cenovno ugodnejše rešitve, kot sta AVer DL30 in HuddleCamHD SimpliTrack2 [5][4], ki poleg same kamere nudijo tudi že vgrajeno sledenje osebam v kadru in sodijo v cenovni razred med 1.000 € in 3.000 €, ampak se pri testih niso izkazale kot natančne pri premikih in nudijo manj možnosti za kontroliranje takega sistema.

3 Vrste sledenja

Za sledenje osebam v kadru je možno uporabiti različne sisteme za določanje položaja v zaprtih prostorih v realnem času. Vsak sistem ima svoje prednosti ter slabosti in posledično vsak sistem ni primeren za vsako predavalnico oz. prostor.

3.1 Sistem slikovnega sledenja

Najbolj znan in uporabljen sistem sledenja je slikovni sistem, ki temelji na prepoznavi karakteristik oseb. Ameriške tajne obveščevalne službe že pet desetletij razvijajo algoritme in programe za mehansko prepoznavo obrazov, saj se zavedajo, da človek ni zmožen dovolj hitro zaznati, ali je na sliki iz kamere iskana oseba ali ne. Zaradi tehnoloških omejitev pa je tovrstna tehnologija postala dostopna povprečnim uporabnikom šele v zadnjih 15 letih.

Slikovni sistemi temeljijo na obdelavi posnetka s posebnimi algoritmi in sistemi prepoznave značilnosti, kot so histogrami, SIFT (ang. Scale-Invariant Feature Transform) algoritmi za zaznavo lastnosti [6], Laplacian piramide [7] in tehnike centroidov, ki zmorejo s pomočjo algoritmov kategorizirati sliko in tako ugotoviti, kje v kadru se nahaja oseba.

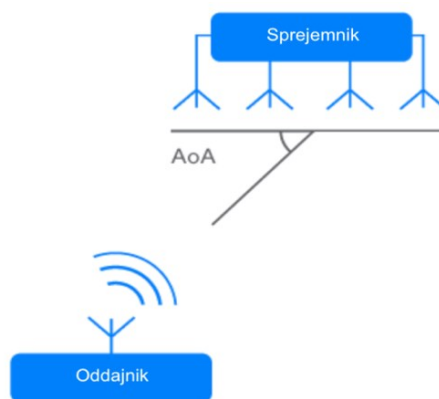
Slikovni sistemi so zaradi uporabe zahtevnih tehnik in algoritmov prepoznavne procesorsko zelo zahtevni, kar pomeni, da je potrebno za izvedbo takega sistema, poleg kvalitetne kamere, dodati tudi dodatno zmogljivo strojno opremo, kar vpliva na ceno.

Sistemi, ki bazirajo na slikovni tehnologiji sledenja, so zelo zanesljivi in natančni, vendar ob slabih svetlobnih pogojih težko prepoznajo obraze oz. podobe ljudi.

3.2 Sistem radio frekvenčnega sledenja

Radiofrekvenčni sistemi sledenja nudijo mnogo različnih tehnik za določanje lokacije znotraj prostora. Najbolj znana je tehnika indikacije sprejete moči signala oz. RSSI (ang. Received Signal Strength Indicator), ki se pretežno uporablja v nakupovalnih centrih in zabaviščnih parkih za podajanje dodatnih informacij glede izdelka preko nizko energijskih Bluetooth oddajnikov (ang. Bluetooth Low Energy Beacon) [8]. Na prostem (oz. v neskončnem praznem prostoru) moč signala pada s kvadratom razdalje od oddajnika, vendar v zaprtih prostorih to ne drži, saj pride do težav z odboji in slabljenjem signala zaradi ovir v samem prostoru.

Novejši sistemi temeljijo na tehnologijah vhodnega kota sprejetega Bluetooth signala – Bluetooth AoA (ang. Angle of Arrival) in ultra širokem frekvenčnem pasu – UWB (ang. Ultra-Wide Band), ki poleg prejete moči omogočata merjenje vhodnega kota signala in posledično določanje smeri, v kateri se nahaja oddajnik. AoA deluje na principu izračuna zamika faz med sprejemnimi antenami, kot je to razvidno na Slika 2. Na podlagi posnetkov prejetega lociranega signala z večimi IQ (ang. in-phase in quadrature-phase) vzorci na različnih sprejemnih antenah je mogoče izračunati smer in oddaljenost prejetega signala [9].



Slika 2: AoA metoda za zaznavo vpadnega kota prihajajočega Bluetooth signala [9]

UWB sistem za večjo natančnost pri določanju lokacije v EU (Evropska Unija) deluje na frekvencah med 6,0 in 8,5 GHz, kar omogoča tudi bistveno višjo zanesljivost pri izvajanju meritev [10].

Pri vseh radiofrekvenčnih sistemih mora predavatelj nositi oddajnik, ki mora biti pravilno nameščen, občasno pa ga je potrebno tudi polniti, kar posledično zahteva dodatno tehnično osebje, ki skrbi za delovanje takšnega sistema. Prav tako se pri teh sistemih pojavljajo težave pri določanju pozicije v notranjih prostorih, kjer pride do povečanih odbojev in interferenc, npr. od večjih kovinskih površin (npr. bele table).

3.3 Sistem termovizijskega sledenja

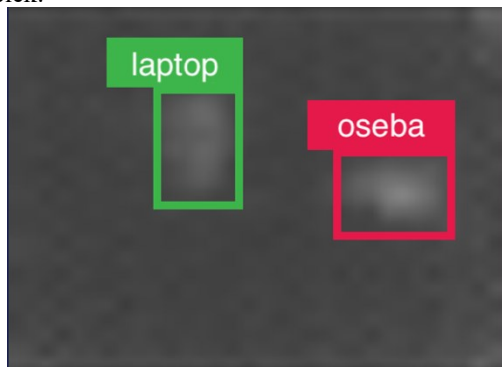
Termovizijske kamere ne zaznavajo vidne svetlobe, ki se odbije v objektiv, ampak ustvarijo sliko na podlagi infrardeče svetlobe, ki jo oddajajo topla telesa.

Povprečna temperatura človeškega telesa se giblje med 36,1 °C in 37,2 °C [11], kar je precej višje od povprečne sobne temperature in pasivnih objektov v sobi, ki se gibljejo okoli 20 °C. Edina naprava v prostoru, ki ima podobno temperaturo, je računalnik, kar je razvidno na Slika 3.

Prednost takšnega sistema je predvsem zanesljivost ob različnih svetlobnih pogojih in različnih vrstah oblačil predavateljev, saj v infrardečem spektru ni opaznih razlik v sceni med na primer rdečo in modro majico.

Pri izdelavi prototipa sta bili izvedeni dve rešitvi s termovizijskim sledenjem, ki sta opisani v nadaljevanju. Uporabljena kamera MLX90640 je nizkocenovnega razreda z nizko ločljivostjo 32x24px [12], kar zadošča za

potrebe sledenja in omogoča cenovno bolj dostopen izdelek.



Slika 3: Klasifikacija osebe in računalnika na zajemu toplotne kamere MLX90640

4 Termovizijsko sledenje oseb

4.1 Vrste sistemov za termovizijsko sledenje

Pri zasnovi termovizijskega sistema sledenja je mogoče uporabiti dva različna principa detekcije. Z uporabo modelov strojnega učenja (ang. Machine learning model - ML) je mogoče doseči bolj robustno zaznavo in razlikovanje med človekom, računalnikom in drugimi morebitnimi toplimi telesi, saj omogoča prepoznavo toplotnega vzorca osebe, kot prikazuje Slika 3. Pri določanju položaja lahko tudi zgolj omejimo dele kadra, v katerih se išče višja temperatura, in se ne obremenjujemo, ali je zaznani objekt z najvišji temperaturo oseba, ali kaj drugega.

Predlagani ML model je bil zgrajen na podlagi 400 vzorcev, od tega je 100 vzorcev vsebovalo osebo in računalnik, 100 samo osebo, 100 samo računalnik in 100 samo prazen prostor. Razpon iskane temperature je bil omejen med 20 °C in 38 °C ter predstavljen v 8-bitni sivinski sliki, kjer 0 predstavlja 20 °C in 255 predstavlja 38 °C. Zaradi omejitev platforme EdgeImpulse [13], kjer se je izvajalo učenje ML modela, je bilo potrebno vse vzorce bi-kubično interpolirati do kvadratne oblike 320x320 slikovnih točk [14].

Po 25-ih dobah (ang. epoch) strojnega učenja je model dosegel natančnost pravilne indentifikacije govorca (s tem, da je cel v kadru) 87,36 % oz. 98,67 % [15], če je bilo sledenje omejeno na notranjih 80 % kadra, saj zaradi karakteristik širokokotnega objektivna na kameri MLX90640BAA pride do popačenja slike na robovih [12].

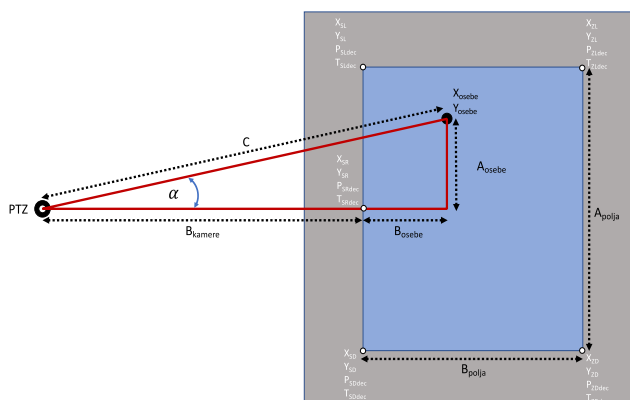
Kljub visoki natančnosti ML modela je bila za prototip izbrana metoda najvišje temperature v kadru, saj je metoda z uporabo ML modela pri trenutni tehnologiji in neoptimizirani programski opremi potrebovala preveč časa za analizo ene same slike – čas obdelave se je namreč gibal med 600–900 ms za ML model in 200–300 ms za obdelavo brez uporabe ML modela. Pri tem so bila uvedena »temna območja« (npr. kateder z računalnikom) kjer se je nastavila temperatura na 0 °C, da objekti znotraj teh območij ne vplivajo na odločitve sistema.

4.2 Preslikava iz ptičje perspektive

Večina PTZ aplikacijskih vmesnikov – API (ang. Application Programming Interface), sprejema PAN,

TILT in ZOOM vrednosti preko GET HTTP (ang. Hyper-text Transfer Protocol) zahtev. Pri Panasonic kamerah so vrednosti parametrov, posredovanih preko API vmesnika, zapisane v šestnajstiškem sistemu, kjer vrednost 8000 predstavlja odklon 0°, vrednost FFFF odklon +175° ter vrednost 0000 odklon -175°.

Pri prototipni rešitvi je bila termovizijska kamera nameščena na stropu, zato je bila zajeta slika iz ptičje perspektive. Posledično je potrebno preslikati lokacijo zaznanega subjekta na termovizijski kameri na pogled kamere in temu primerno izračunati pripadajoče PAN in TILT vrednosti. Na podlagi tehnične specifikacije kamere, območja premikanja govorca in meritev prostora, kot prikazuje Slika 4, so bila izračunana polja potrebnega pokrivanja ter izpeljane enačbe (1–8) za izračun ustreznih PAN in TILT vrednosti kamer.



Slika 4: Potrebne meritve za izračun PAN in TILT

$$A_{osebe} = \frac{|X_{osebe} - X_{SR}| \times \frac{A_{polja}}{2}}{X_{SL} - X_{SR}} \quad (1)$$

$$B_{osebe} = \frac{(Y_{osebe} - Y_{SR}) \times B_{polja}}{Y_{ZL} - Y_{SL}} \quad (2)$$

$$\alpha = \tan^{-1} \frac{A_{osebe}}{B_{osebe}} \quad (3)$$

$$PAN = \alpha \times \frac{P_{SLdec} - P_{SRdec}}{\tan^{-1} \frac{A_{polja}}{2} \times \frac{A_{polja}}{B_{kamere}}} \quad (4)$$

$$C = \sqrt{A_{osebe}^2 + B_{osebe}^2} \quad (5)$$

$$C_1 = \sqrt{B_{kamere}^2 + \left(\frac{A_{polja}}{2}\right)^2} \quad (6)$$

$$C_2 = \sqrt{(B_{kamere} + B_{polja})^2 + \left(\frac{A_{polja}}{2}\right)^2} \quad (7)$$

$$TILT = C \times \frac{T_{ZLdec} - T_{SLdec}}{C_2 - C_1} \quad (8)$$

Pri premiku predavatelja v predavalnici ni vedno zaželeno sprotno zvezno sledenje kamere, saj bi to ob pisanju na tablo bilo zelo moteče, zato so bile na območju

gibanja definirane cone, znotraj katerih se PTZ kamere ne premikajo, vse dokler predavatelj ne zapusti cone.

4.3 Programska oprema

Programska koda za avtonomni sistem, spisana v jeziku Python, je na voljo na Github repozitoriju https://github.com/mbdev2/RTLS_PTZ_Tracking.

Program prejema podatke preko I²C vmesnika in nato s pomočjo Flask strežnika in uporabniškega vmesnika, kot prikazuje Slika 5, omogoča uporabniku izbiro med pred nastavljenimi statičnimi kadri in avtonomnim sledenjem. Pri avtonomnem sledenju sistem sproti v ozadju preračunava pozicijo predavatelja in pošilja GET HTTP ukaze z uporabo Requests knjižnice na API PTZ kamere, kar omogoča sprotno sledenje predavatelju po predavalnici, vse dokler ne zapusti vidnega polja termovizijske kamere.



Slika 5: Uporabniški vmesnik spletne aplikacije

5 Sklep

Avtonomno sledenje omogoča predavateljem in gledalcem veliko boljše uporabniško izkušnjo kot samo statični kadri, ampak vse rešitve niso enakovredne. Na zajem in upravljanje PTZ kamer vpliva mnogo dejavnikov, ki so odvisni od fizičnih lastnosti prostora, na katere pogosto ni mogoče vplivati v tolikšni meri, kot bi bilo potrebno za optimalno delovanje sistemov za sledenje predavatelja.

Vizualni sistemi so zanesljivi in zelo natančni ob dobrih svetlobnih pogojih, ampak pogosto za delovanje le teh potrebujemo zmogljivo in drago strojno opremo. Radiofrekvenčni sistemi delujejo neodvisno od svetlobnih pogojev, vendar so občutljivi na radiofrekvenčne motnje, odboje in interference, ki so pogosto posledica več kovinskih površin v predavalnicah, kot npr. bele table. Prav tako zahtevajo namestitve dodatne strojne opreme v prostoru in vzdrževalno osebje. Termovizijski sistemi sicer zahtevajo namestitve dodatne strojne opreme, vendar so mnogo enostavnejši za vzdrževanje in uporabo ter hkrati omogočajo enostavnejšo integracijo in z izbranimi komponentami predstavljajo veliko nižji finančni vložek kot ostali integrirani ali ločeni sistemi. Ne glede na izbrano metodo je končni rezultat odvisen tudi od načina implementacije algoritma za premikanje kamere in natančnosti sistema za premik pogleda kamere.

Literatura

- [1] Tyco, „Automated Production“, *Pixellot*. <https://www.pixellot.tv/solutions/automated/> (pridobljeno jun. 30, 2021).
- [2] „Why webcams aren't good enough“. <https://reincubate.com/support/how-to/why-are-webcams-bad/> (pridobljeno jul. 23, 2021).
- [3] „AW-HE42W/K | PTZ Camera Systems | Broadcast and Professional AV | Panasonic Global“. <https://pro-av.panasonic.net/en/products/aw-he42/> (pridobljeno jun. 30, 2021).
- [4] „Auto Tracking Camera“, *HuddleCamHD*. <https://huddlecamed.com/simpltrack-2/> (pridobljeno jun. 30, 2021).
- [5] „Aver DL30 - Distance Learning Tracking Camera“. <https://presentation.aver.com/model/dl30> (pridobljeno jun. 30, 2021).
- [6] „OpenCV: Introduction to SIFT“. https://docs.opencv.org/master/da/df5/tutorial_py_sift_intro.html (pridobljeno jun. 30, 2021).
- [7] W.-S. Lai, J.-B. Huang, N. Ahuja, in M.-H. Yang, „Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution“, *ArXiv170403915 Cs*, okt. 2017, Pridobljeno: jun. 30, 2021. [Na spletu]. Dostopno na: <http://arxiv.org/abs/1704.03915>
- [8] „Beacon - What is it and how it work? BLE Bluetooth Beacon Technology Guide | Kontakt.io“. <https://kontakt.io/what-is-a-beacon/> (pridobljeno jul. 23, 2021).
- [9] „How AoA & AoD Changed the Direction of Bluetooth Location Services“, *Bluetooth® Technology Website*, mar. 27, 2019. <https://www.bluetooth.com/blog/new-aoa-aod-bluetooth-capabilities/> (pridobljeno jul. 02, 2021).
- [10] „Ultra-Wideband (UWB) in Europe - Telecom ABC“. <http://www.telecomabc.com/u/uwb-europe.html> (pridobljeno jul. 02, 2021).
- [11] „Body temperature norms: MedlinePlus Medical Encyclopedia“. <https://medlineplus.gov/ency/article/001982.htm> (pridobljeno jul. 03, 2021).
- [12] „MLX90640-Datasheet-Melexis“, *MLX90640-Datasheet-Melexis*. Pridobljeno: jul. 17, 2021. [Na spletu]. Dostopno na: <https://mel-prd-cdn.azureedge.net/-/media/files/documents/datasheets/mlx90640-datasheet-melexis.pdf>
- [13] „Ingestion service“, *Edge Impulse Docs*. <https://docs.edgeimpulse.com/reference#data-acquisition-format> (pridobljeno jun. 30, 2021).
- [14] „Bicubic interpolation“, *Wikipedia*. jul. 01, 2021. Pridobljeno: jul. 03, 2021. [Na spletu]. Dostopno na: https://en.wikipedia.org/w/index.php?title=Bicubic_interpolation&oldid=1031409060
- [15] F. Gaillard, „Epoch (machine learning) | Radiology Reference Article | Radiopaedia.org“, *Radiopaedia*. <https://radiopaedia.org/articles/epoch-machine-learning> (pridobljeno jul. 04, 2021).