

VILNIAUS UNIVERSITETAS
MATEMATIKOS IR INFORMATIKOS FAKULTETAS
PROGRAMŲ SISTEMŲ KATEDRA

Kursinis darbas

Gestų kalbos atpažinimas naudojant internetinę kamerą
(Sign language recognition using web camera)

Atliko: 3 kurso I grupės studentas

Pranciškus Ambrazas (parašas)

Darbo vadovas:

asist. Linas Petkevičius (parašas)

Vilnius, 2017

Turinys

Išvadas	2
1. Teorija	3
1.1. Savaime apsimokančios sistemos	3
1.2. Neuroniniai tinklai	3
1.2.1. Dirbtiniai neuroniniai tinklai	4
1.2.2. Rekurentiniai neuroniniai tinklai	5
2. Gestų kalba	6
2.1. Gestų kalbos skirstymas	6
2.2. Gestų kalbos atpažinimas	6
2.2.1. Problematika	6
3. Sistemos apmokymas	8
3.1. Kadro apdorojimas naudojant Sobel valdiklį	8
3.2. Kadro atidavimas savaime apsimokančioms sistemoms	9
4. Gestų kalbos statinių ženklų atpažinimas	10
4.1. Atpažinimas naudojant gestų nuotraukas	10
4.2. Atpažinimas naudojant internetinę kamerą	10
5. Gestų kalbos dinaminių ženklų atpažinimas	11
6. Pagrindinė tiriamoji dalis	12
6.1. Poskyris	12
6.1.1. Skirsnis	12
6.1.1.1. Straipsnis	12
6.1.2. Skirsnis	12
7. Skyrius	13
7.1. Poskyris	13
7.2. Poskyris	13
Išvados	14
Literatūra	15
Priedas Nr.1	
Priedas Nr.2	

Įvadas

Daugiau nei 360 milijonų žmonių pasaulyje kenčia dėl klausos ir kalbos įvairių problemų, o daugiau nei 32 milijonai jų yra vaikai ir šis skaičius vis auga [3]. Gestų kalba yra pagrindinis šių žmonių bendravimo įrankis. Tačiau reiktų pastebėti ir tai, jog dalis jų moka skaityti iš lūpų. Norint žmogui be šių ydų bendrauti su gestakalbiu (*gestų kalba kalbantis žmogus*) reikia vertėjo, kuris išverstų gestų kalbą į įprastinę ir atvirkščiai.

Kiekviena pasaulyje esanti kalba turi ir savo gestų kalbą. Tai reiškia, skiriasi tiek gestų kalbos gramatika, tiek netgi patys gestai. Pasaulyje randama net dialektų pagal regionus, ne tik pagal šalis. Pavyzdžiui, amerikiečių anglų kalba šnekančių žmonių pasaulyje yra apie 500 tūkstančių [remiantis **United States Census Bureau's data from the 2009-2013 American Community Survey**]. Todėl netgi bendraujant dviem žmonėms, mokantiems gestų kalbą neretai iškyla vertimo problema, tad tenka ieškoti gestų vertimų. Paiešką galima atlikti atsižvelgiant į delno padėtį, rankos judesį (ar net abiejų rankų), rankos padėtį ir kiek rankų atlieka gestą. Tuomet pagal gesto išvaizdos paveiksliukus ar kartais net vaizdo įrašus, gestakalbiai gali išsiversti gestus. Tam yra skirtos tiek internetinės svetainės - žodynai, tiek įvairūs rašytiniai žodynai.

Šio tyrimo tikslas - padėti ne tik gestakalbiams tarpusavyje, bet ir žmonėms, nesuprantantiems gestų kalbos bendrauti su gestakalbiais tam pasitelkiant technologijas, taip suteikiant šiems žmonėms pilnavertį gyvenimą bendraujant su kitais. Šiuo tyrimu siekiama apžvelgti ir įvertinti ar naudojantis įprasta kompiuterine kamera įmanoma paversti gestų kalbą rašytiniu tekstu ar net garsine kalba ir lygiai taip pat versti rašytinę ar garsinę kalbą į gestų kalbą. Taip pat siekiama, kad vėliau būtų sukurtas visiems gestakalbiams prieinamas produktas ar programinė įranga, kurią kiekvienas įsidieęs į savo įrenginį - kompiuterį, mobiliųjį telefoną ar planšetinį kompiuterį galėtų naudotis šiuo vertėju. Taip pat tai galėtų tapti ir mokomąja gestų kalbos priemone.

Šiuo metu yra gaminamas vienas iš tokių produktų, tačiau tai yra atskiras įrenginys, turintis įmontuotą kamerą, kuri be vaizdo taip pat fiksuoja ir atstumą iki tam tikrų objektų (šiuo atveju rankos), tačiau produkto kūrėjai sako, jog jų įrenginys galės versti gestų kalbą į anglų ir kitas kalbas ir lygiai taip pat versti įprastą kalbą į rašytinę kalbą. Plačiau: <http://www.motionsavvy.com/>

Šiame darbe bus pagrįste nagrinėjama amerikiečių (*angl. American Sign Language (ASL)*) ir lietuvių gestų kalbos.

1. Teorija

Šiame skyriuje bus aprašoma teorija apie apsimokančias sistemas ir neuroninius tinklus

1.1. Savaime apsimokančios sistemos

Apsimokančios sistemos (*angl. machine learning*) - tai mokslas apie tai, kaip kompiuterius užprogramuoti taip, jog jie patys darytų sprendimus be žmogaus įsikišimo neužprogramuojant kiekvienos galimos situacijos. Kitais žodžiais tariant, leisti kompiuteriui pačiam nuspręsti kaip elgtis esant tam tikroms aplinkybėms. Savaime apsimokančios sistemos yra didelis žingsnis į priekį norint sukurti dirbtinį intelektą.

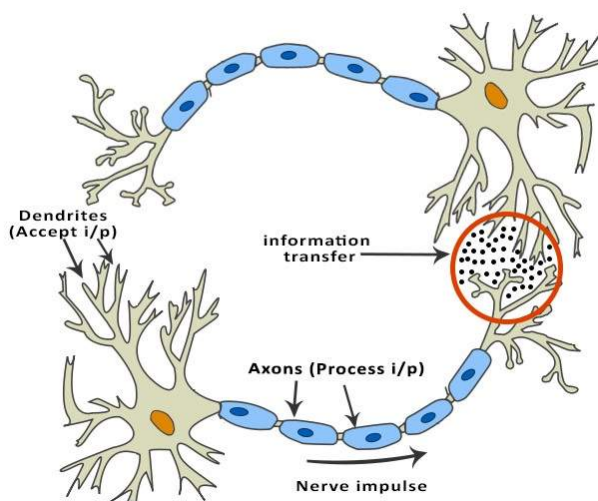
Savaime apsimokančių sistemų ir jų algoritmų sukūrimo dėka gatvėmis pradėjo važinėti patys save vairuojantys automobiliai (*angl. self-driving cars*) arba dar kitaip vadinamos autonominės transporto priemonės. Įvairūs paieškos varikliai tokie kaip „Google“ ar „Yahoo“ taikydami šiuos alogirtmus naudotojams rodo kiekvienam asmeniškai sugeneruotą turinį. Taip pat reikėtų paminėti ir kalbos atpažinimo sistemas tokias kaip „Siri“ ar „Google Assistant“, kurios iš joms duotų komandų atlieka tam tikrus veiksmus.

1.2. Neuroniniai tinklai

Neuronas (*arba - nervinė ląstelė*) - pagrindinės nervų sistemos ląstelės, sukuriančios ir/arba perduodančios elektrocheminius impulsus.

Žmogaus smegenys yra sudėtingas, nelinijinis ir paralelinis kompiuteris[2], kurias sudaro neuronai. Vienas neuronas vienu metu jungiasi su daugybe kitų neuronų per dendritus, ant kurių yra daug sinapsių, per kurias ateina informacija iš kitų neuronų. Dendritus paprasčiau galima pavadinti kaip informacijos priimėjais. Todėl vienas neuronas gali sudaryti iki 100000 sinapsių. Kiekviena sinapsė gali būti jaudinanti arba slopinanti. Visas šis mechanizmas dar nadinamas **neuroniniais tinklais** (*angl. Neural network*).

Vienas neuronas priima informaciją per dendritus, tuomet pats neuronas nusprendžia ar bus siunčiama informacija į kitus neuronus ir kokia ji bus siunčiama (*žr. 1 pav.*)



1 pav. Neuroninio tinklo struktūra [1]

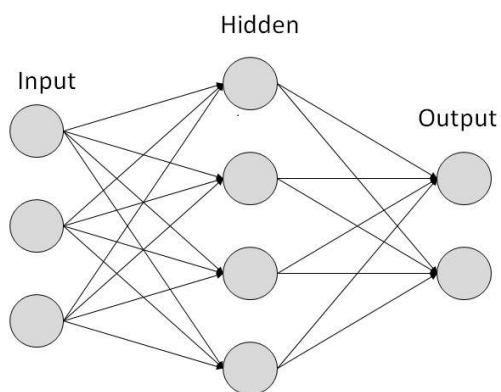
Remiantis šiais principais buvo sukurti dirbtiniai neuroniniai tinklai.

1.2.1. Dirbtiniai neuroniniai tinklai

Dirbtinis neuroninis tinklas (angl. *Artificial neural network*) - struktūra, skirta apdoroti didiam kiekiui informacijos, sukurta remiantis žmogaus nervų sistemos veikimo principu. Kitais žodžiais tariant, skaitmenizuota žmogaus smegenų veikla.

Neuronai dirbtiniame neuroniniame tinkle yra sujungti jungtimis. Taip jie tarpusavyje komunicuoja perduodami vienas kitam informaciją. Kiekvienas neuronas gali priimti atėjusią informaciją, ją apdoroti ir perduoti kitam neuronui. Kiekviena jungtis turi savo svorį, pagal kurį pasirenkama, į kurį neuroną turi būti perduodama informacija. Šių jungčių svorius galime nustatyti arba apmokyti naudodamiesi apsimokančių sistemų pagalba.

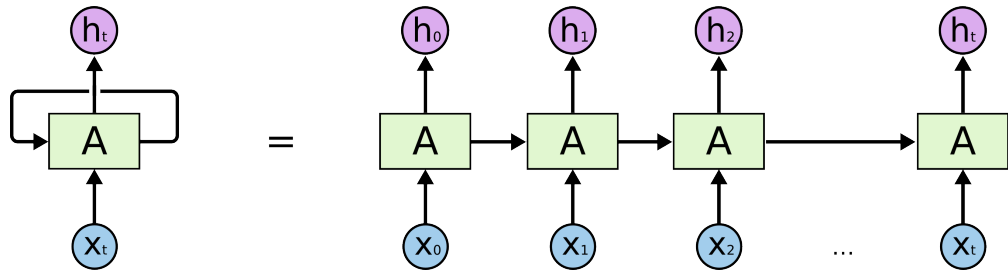
2 pav. pateikiama paprasta schema, kaip veikia dirbtiniai neuroniniai tinklai:



2 pav. Dirbtinio neuroninio tinklo struktūra [1]

1.2.2. Rekurentiniai neuroniniai tinklai

Rekurentiniai neuroniniai tinklai (*angl. Recurrent neural network*) - tai dirbtinis neuroninis tinklas, kuris saugo informaciją apie praeituose žingsniuose (neuronuose) atliktus veiksmus ar skaičiavimus.



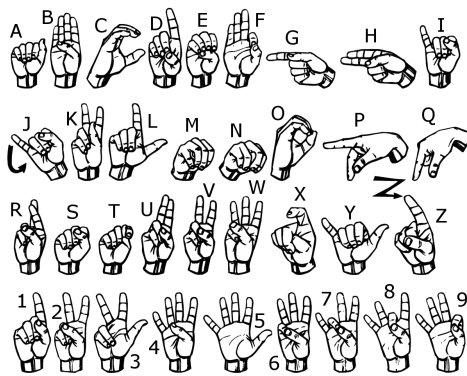
3 pav. Rekurentinių neuroninių tinklų schema [1]

- x_t - įeiga momentu t
- h_t - išeiga momentu t
- A_t - būseną momentu t

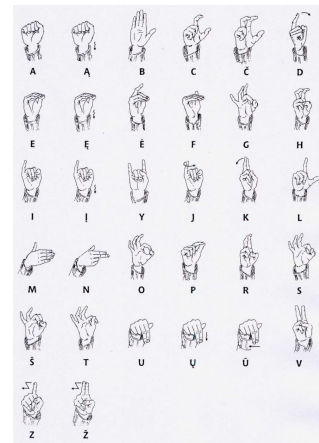
2. Gestų kalba

2.1. Gestų kalbos skirstymas

Gestų kalba susideda iš dviejų dalių - statinių ir dinaminių ženklų. Gestų kalboje kiekviena kalba turi savo abėcėlę. Statiniais ženklais atvaizduojama didžioji abėcėlių raidžių dalis. O dinaminiais - žodžiai ir kai kurios gestų abėcėlių raidės. Pavyzdžiui, amerikiečių gestų kalbos abėcėlėje J ir Z raidės atvaizduojamos dinaminiais judesiais (žr. 4 pav.), o lietuvių - A, D, I, K ir kt. bei jau minėtosios J ir Z raidės (žr. 5 pav.)



4 pav. Amerikiečių gestų kalbos abėcėlė



5 pav. Lietuvių gestų kalbos abėcėlė

2.2. Gestų kalbos atpažinimas

Gestų kalba ir jos atpažinimas naudojant kompiuterinę techniką yra viena iš daugelio apsimo-kančių sistemų sričių.

2.2.1. Problematika

Norint atpažinti gestus, paversti juos į raides, žodžius ar sakinius susiduriama su problemomis, kurios susijusios tiek su statinių, tiek su dinaminių gestų atpažinimu.

Pagrindinės problemos išskylančios atpažįstant statinius gestų kalbos ženklus yra:

1. Kiekvienos kalbos abėcėlę sudaro skirtingas raidžių (statinių ženklų) skaičius. Pavyzdžiui, lietuvių kalbos abėcėlę sudaro 32 ženklai, o amerikietiška - 26;
2. Gestų panašumai. Pavyzdžiui, raidės A, E, N, S, T yra atvaizduojamos sugniaužtus kumštį, o net trijose iš jų (A, E ir S) skiriasi tik nykščio padėtis;
3. Kampas, kuris susidaro atpažįstant gestą. Pavyzdžiui, kai A raidė rodoma ne iš priekio, o iš šono;
4. Apšvietimas. Pavyzdžiui, gestų atpažinimas esant prieblandai ir dienos šviesai.

Pagrindinės problemos išskylančios atpažįstant dinامينius gestų kalbos ženklų yra:

1. Nauji gestų kalbos žodžiai. *Pavyzdžiui*, kiekvienas uraganas turi savo pavadinimą, todėl tai gali reikšti naujo gesto atsiradimą;
2. Gesto kelios reikšmės. *Pavyzdžiui*, vienas gestas gali turėti kelias reikšmes, kaip kad lietuvių kalboje vienas žodis „kasa“ gali turėti net tris skirtingas reikšmes;
3. Kampas, kuris susidaro atpažįstant gestą. *Pavyzdžiui*, kai A raidė rodoma ne iš priekio, o iš šono;
4. Žodžių apjungimas į vieną sakinį. *Pavyzdžiui*, keli gestai einantys vienas po kito gali reikšti vieną žodį, tačiau tuo pačiu būti panašūs į vieną gestą, kuris jau reikš tik vieną žodį.

3. Sistemos apmokymas

Norint, jog sistema atpažintų gestus, svarbiausia ją apmokyti ką reiškia tam tikri gestai. Tam galime išnaudoti kadrus (*angl. frame*) ir savaime apsimokančių sistemų galimybes. Tad reiktų imti vieną kadrą ir jį paversti į kompiuteriui suprantamą kalbą.

Imkime, jog nuotrauka yra sudaryta iš $n * m$ taškų (*angl. pixels*). Reikia, jog sistema atskirtų, kurioje kadro vietoje yra rodomo gesto dalis, kurioje tik pašalinis fonas.

Apmokymui toliau aptarsime Sobel valdiklį, savaime apsimokančių sistemų metodus ir duomenų surinkimą

3.1. Kadro apdorojimas naudojant Sobel valdiklį

Sobel valdiklis (*angl. Sobel operator*) - vaizdų apdorojimo algoritmas, skirtas paversti kadrą į kontūrų žemėlapi.

Šis valdiklis naudojami šiomis funkcijomis, kad konvertuotų vaizdą į kontūrus:

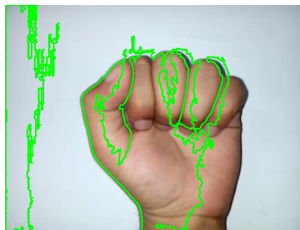
$$G_x = \begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix} * A \quad (1)$$

$$G_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * A \quad (2)$$

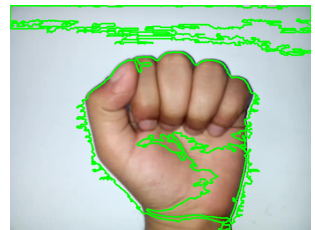
$$G = \sqrt{G_x^2 + G_y^2} \quad (3)$$



6 pav. Originalus paveikslėlis



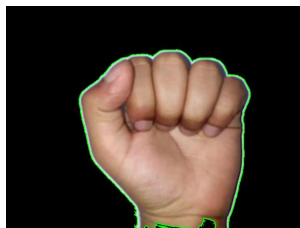
7 pav. Pritaikyta G_x



8 pav. Pritaikyta G_y



9 pav. Pritaikyta G



10 pav. Be fono



11 pav. Dviejų spalvų

Prieš kadrui taikant Sobel funkcijas, reikia jį paruošti. Kadangi Sobel funkcijos ima matricas

sudarytas iš 3 x 3 skaičių, tai reikia sušvelninti šalia esančius skaičius. Tą padaryti lengviausia pasinaudojant uždedant, pavyzdžiui, 5% miglą (*angl. blur*) kadrai. Uždėti miglą reikia tam, jog sumažintume triukšmą (*angl. noise*) kontūrams.

Toliau, atiduodant kadrai Sobel funkcijoms, kadras yra konvertuojamas į skaičių masyvą, kur kiekvienas taškas turi savo skaičių - spalvos kodą. Tuomet norint rasti visus kontūrus esančius kadre, taikome šiuos veiksmus:

1. Einame per visus taškus esančius kadre ir taikome 1 formulę, kur A - kiekvienas 3 x 3 kadro taško spalvos mastytas. Gauname tokį vaizdą - žr. 7 pav.;
2. Einame per visus taškus esančius kadre ir taikome 2 formulę, kur A - kiekvienas 3 x 3 kadro taško spalvos mastytas. Gauname tokį vaizdą - žr. 8 pav.;
3. Apskaičiuojame dabartinio taško tikrąją reikškę taikydami 3 formulę. Gauname tokį vaizdą - žr. 9 pav.

Po šių žingsnių turime matricą, kuriose keičiasi spalva. dabar išimti tuos taškus, kurie yra, pavyzdžiui, ta pati balta spalva, tik kitokio atspalvio? O gi imame ir susižinome matricos vidurkį ir atmetame visus taškus, kurie yra mažesni už vidurkį. Kitaip tariant, išimame mažo skirtumo taškus. Taip gauname, jog lieka tik tie taškai, kurie jau turėtų priklausyti gesto kontūrai.

Ieškant kontūro taip pat atmetame ir tuos kontūrus, kurie, pavyzdžiui, neužima daugiau nei 5% viso ploto ir nubrėžiame kontūrą.

Kitas žingsnis paversti kadrai į kompiuteriui suprantamą kalbą ir kuo paprastesnę. Turėdami kontūrus, galime kadrai paversti į dvispalvį kadrai, kur viskas, kas yra kontūre bus baltos spalvos, o viskas kas už kontūro - juodos. Tam reiktų pasidaryti kaukę (*angl. mask*), kurioje viskas, kas už kontūro ribų bus juodos spalvos (žr. 10 pav.), o paskui ištrinti tai, kas yra kontūre ir padaryti baltos spalvos (žr. 11 pav.).

Po šių žingsnių turime masyvą, kuris yra sudarytas iš reikšmių, esančių tarp 0 ir 255, kur 0 - juoda spalva, o 255 - balta. Taip, pritaikę savo kadrai gautą kaukę (*angl. mask*) gauname kadrai, kuriame yra likę tik kontūrai (žr. 9 pav.)

Kitas žingsnis yra remiantis šia kauke paversti paveikslėlį į juodai baltą paveikslėlį (žr. 11 pav.).

3.2. Kadro atidavimas savaime apsimokančioms sistemoms

Savaime apsimokančios sistemos, tokios kaip *sklearn* konvertuoja jau apdorotą kadrai į matricą, kur juoda - 0, o balta - 255. Tuomet ši matrica yra „suplokštinama“ ir gauname masyvą. Galiausiai šiuos duomenis sudeda į failą, kad vėliau jais būtų galima pasinaudoti.

4. Gestų kalbos statinių ženklų atpažinimas

4.1. Atpažinimas naudojant gestų nuotraukas

Norint, jog sistema apsimokytų kuo tiksliau, reikia jai duoti kuo daugiau duomenų. Sakykime, kad:

- a_{ik} - i -toji k -tosios abėcėlės raidės nuotrauka
- $a_{1k}, a_{2k}, \dots, a_{nk}$ - nuotraukų rinkinys, sudarytas iš n k -tosios raidės nuotraukų

Tuomet, turėdami $n * k$ nuotraukų, kuriuose vaizduojami gestų abėcėlės gestai, turime, jog tiek eilučių duomenų turėsime gestams atpažinti. Kuo gestai yra įvairesni prie skirtingų apšvietimų, skirtingų rankų ir pan., tuo tiksliau sistema pati galės vėliau atpažinti gestus.

4.2. Atpažinimas naudojant internetinę kamerą

Norint atpažinti statinius gestus naudojant internetinę kamerą, vienas iš to būdų yra naudotis *BackgroundSubtractorMOG2* klase. Ši klasė remiasi Gauso maišos priekinio plano/fono atskyrimo algoritmu

5. Gestų kalbos dinaminių ženklų atpažinimas

6. Pagrindinė tiriamoji dalis

Pagrindinėje tiriamojoje dalyje aptariama ir pagrindžiama tyrimo metodika; pagal atitinkamas darbo dalis, nuosekliai, panaudojant lyginamosios analizės, klasifikacijos, sisteminimo metodus bei apibendrinimus, dėstoma sukaupta ir išanalizuota medžiaga.

6.1. Poskyris

Citavimo pavyzdžiai: cituojamas vienas šaltinis [4]; cituojami keli šaltiniai [**PvzElPubLt**, 5, 6, 7, 8, 9, 10, 11, 12].

6.1.1. Skirsnis

6.1.1.1. Straipsnis

6.1.2. Skirsnis

7. Skyrius

7.1. Poskyris

7.2. Poskyris

Išvados

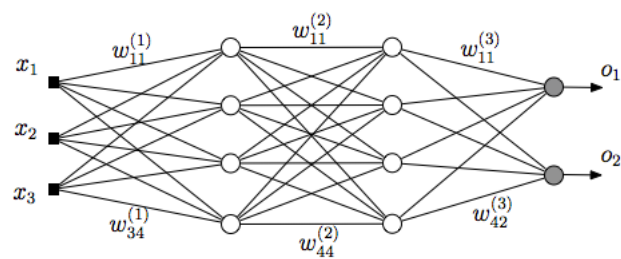
Išvadose ir pasiūlymuose, nekartojant atskirų dalių apibendrinimų, suformuluojamos svarbiausios darbo išvados, rekomendacijos bei pasiūlymai.

Literatūra

- [1] Christopher Olah. *Understanding LSTM Networks*. <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>. 2015.
- [2] Simon Haykin. *Neural Networks and Learning Machines*. 1 psl. Upper Saddle River, New Jersey 07458: Pearson Education inc., 2009.
- [3] World Health Organization. *Deafness and hearing loss*. <http://www.who.int/mediacentre/factsheets/fs300/en/>. 2017-02.
- [4] A. Pavardenis, B. Pavardonis ir C. Pavardauskas. „Straipsnio pavadinimas“. *Žurnalo pavadinimas* IV (2001), p.p. 8–17.
- [5] A. Pavardenis, B. Pavardonis ir C. Pavardauskas. *Knygos pavadinimas*. 172 psl. Miestas, šalis: Leidykla, 2003.
- [6] A. Pavardenis ir k.t. „Straipsnio pavadinimas“. *Rinkinio pavadinimas*. Miestas, šalis: Leidykla, 2002, p.p. 3–15.
- [7] A. Pavardonis. „Magistrinio darbo pavadinimas“. Magistrinis darbas. Universiteto pavadinimas, 2005.
- [8] A. Surname. “Title fo PhD thesis”. PhD thesis. Title of university, 2005.
- [9] A. Surname, B. Tsurname, and C. Usurname. “Article Title”. In: *Journal Title* IV (2001), pp. 3–15.
- [10] A. Surname, B. Tsurname, and C. Usurname. *Book title*. 172 p. City, country: Publisher, 2003.
- [11] A. Surname, B. Tsurname, and C. Usurname. *Online Publication Title*. <http://example.com/path/to/the/article.pdf>. 45 KB, accessed 2015-02-01. 2004.
- [12] A. Surname et al. “Article title”. In: *Conference book title*. City, country: Publisher, 2002, pp. 3–15.

Priedas Nr. 1

Niauroninio tinklo struktūra



12 pav. Paveikslėlio pavyzdys

Priedas Nr. 2

Eksperimentinio palyginimo rezultatai

1 lentelė. Lentelės pavyzdys

Algoritmas	\bar{x}	σ^2
Algoritmas A	1.6335	0.5584
Algoritmas B	1.7395	0.5647