

ՀՏԴ 621.337.1

ԱՎՏՈՄԱՏԱՑՈՒՄ ԵՎ
ԿԱՌԱՎԱՐՄԱՆ ՀԱՄԱԿԱՐԳԵՐ

Տ.Բ. ԽԱՉԱՏՐՅԱՆ

**ՊԱՏԿԵՐԻ ԽՈՐՈՒԹՅԱՆ ԳՆԱՀԱՏՈՒՄԸ ԶՎԵՐԱՀԱՄԱԿՈՂ ՎԱՐԺԵՑՄԱՄԲ
ՆՆՅՐԱԿԱՅԻՆ ՑԱՆՑԻ ՄԻՋՈՑՈՎ**

Ներկայացված է պատկերում խորությունների գնահատման մեթոդ՝ խոր ուսուցման և նեյրոնային ցանցերի միջոցով: Առաջարկվում է պատկերում խորությունների գնահատումը կատարել՝ համատեղ կիրառելով երկու մեթոդներ՝ խորության գնահատումը երկկապիճային անհամատեղելիության միջոցով և խորության գնահատումը շարժվող տեսանկարահանող սարքի միջոցով: Այդ նպատակով կառուցվել է նեյրոնային ցանց, որը վարժեցվել է չվերահսկվող վարժեցման եղանակով՝ երկկապիճային տեսանկյունների միջոցով:

Առանցքային բառեր խոր ուսուցում, նեյրոնային ցանց, պատկերի խորություն, չվերահսկվող վարժեցում:

Ներածություն: Պատկերում խորության գնահատումը տեսարանի եռաչափ կառուցվածքի նկարագրման հիմնական բաղադրիչներից մեկն է: Պատկերի խորության գնահատումը կամ ճանաչումը տեսարանի երկչափ պատկերից տեսարանում առկա օբյեկտների հեռավորությունները բնութագրող խորության քարտեզի ստացումն է:

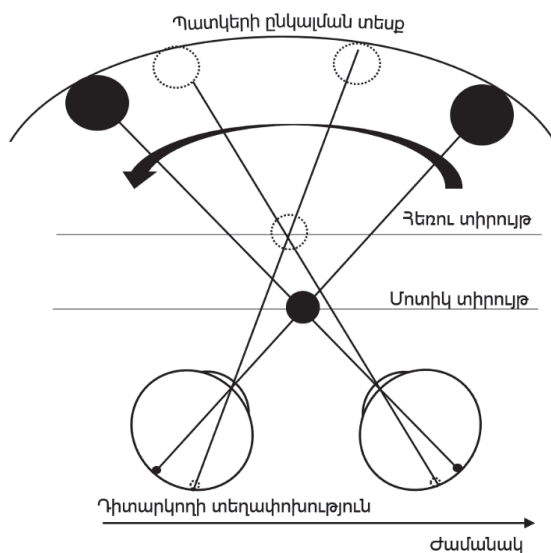
Ինքնավար շարժական համակարգերում շատ կարևոր է ինքնավար համակարգի շարժումից բխող խորության փոփոխության գնահատումը: Այդ առումով կարևոր է կենտրոնանալ պատկերում շարժումից բխող խորության գնահատման մեթոդների վրա:

Որպես նման խնդիրների լուծման լավագույն միջոց ընտրվել է վերջին ժամանակներում մեծ հաջողությունների հասած, խոր ուսուցման և նեյրոնային ցանցերի վրա հիմնված մեթոդը՝ փաթույթային նեյրոնային ցանցը:

Չվերահսկվող վարժեցմամբ մեթոդ. Ներկայումս մեծ հետաքրքրություն են ներկայացնում չվերահսկվող եղանակով վարժեցված մոդելները, քանի որ այս դեպքում նեյրոնային ցանցը վարժեցվում է միայն պատկերների միջոցով, և այլ զգայունակներից ստացված տվյալներ անհրաժեշտ չեն: Կարելի է առանձնացնել պատկերի խորության գնահատման նեյրոնային ցանցերի չվերահսկվող վարժեցման երկու հիմնական եղանակ՝ երկկապիճային անհամատեղելիությունների եղանակը և միակապիճային՝ շարժումից պատկերի տեղափոխության եղանակը: Առավել լավ ուսումնասիրված և տարբեր աշխատանքներում կիրառված է երկկա-

պիճային անհամատեղելիությունների եղանակը [1-3]: Ըստ այս եղանակի առանձնահատկության՝ վարժեցման համար ներդրոնային ցանցին տրամադրվում է երկու նկարահանող սարքերից տեղեկություն՝ պատկեր՝ նկարահանված միևնույն պահին տարբեր անկյուններից: Այս հանգամանքի պատճառով ներդրոնային ցանցը չի պարունակում տեղափոխության վերաբերյալ վարժեցման տվյալներ, ինչի հետևանքով հետագա օգտագործման ժամանակ լավ արդյունքներ են գրանցվում միայն այն պատկերներում, որտեղ համեմատաբար քիչ շարժվող տարրեր կան: Ինքնավար շարժական համակարգերի համար սա կարելի է դիտել որպես թերություն:

Խորության գնահատումը՝ շարժումից պատկերի տեղափոխության եղանակով. Դիտարկող համակարգի շարժումից բխող դիտարկվող օբյեկտների տեղափոխությունը հանդիսանում է խորության հուշիչ [4]: Շարժման հետևանքով պատկերում ավելի շատ տեղափոխություն են կրում ավելի մոտ օբյեկտները (նկ. 1):



Նկ. 1. Խորության ընկալումը շարժումից պատկերի տեղափոխության միջոցով

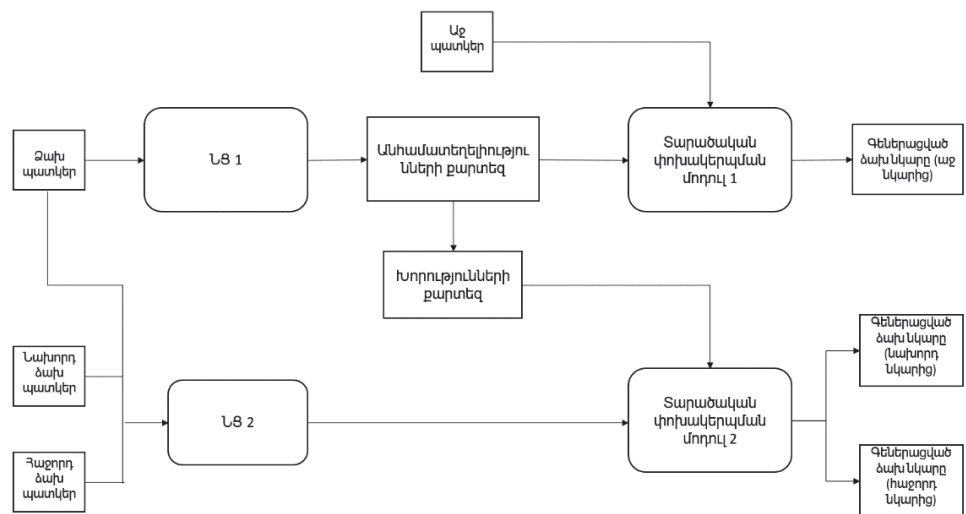
Ունենալով շարժվող տեսանկարահանող սարքի գրանցած տեսարանի հաջորդական պատկերները, կարող ենք դատողություններ անել տեսարանում առկա օբյեկտների հեռավորությունների մասին:

Խորության գնահատման առաջարկվող եղանակը: Խորության գնահատման երկկապիճային անհամատեղելիությունների և շարժումից պատկերի տեղափոխության եղանակների համատեղումը թույլ կտա տարբեր պայմաններում նույնականացնել խորությունը: Նման մոտեցումը կերկարացնի ներդրոնային ցանցի վարժեցման գործընթացը, քանի որ մուտքային տվյալների քանակն ավելանում է:

Բայց այս հանգամանքը, բնականաբար, չի վերաբերում ցանցի առերեսմանը, ինչը համակարգի առումով թիրախային գործոն է:

Այս մեթոդի իրականացման համար կառուցվել է նոր մոդել, որը բաղկացած է երկու փաթույթային նեյրոնային ցանցերից: Դրանցից առաջինը նախատեսված է ձախ և աջ պատկերների երկկապիճային անհամատեղելիությունների, իսկ երկրորդը՝ շարժումից պատկերի տեղափոխության հայտնաբերման համար: Երկու ցանցերը միավորված են ընդհանուր նպատակային ֆունկցիայով:

Նկ. 2-ում պատկերված է կառուցված մոդելի ընդհանուր գծապատկերը:



Նկ.2. Առաջարկվող մոդելի ընդհանուր կառուցվածքը

Առաջին նեյրոնային ցանցն ունի ավտոկոդավորիչի կառուցվածք, որը բաղկացած է երկու հիմնական մասերից՝ կոդավորիչից և վերծանիչից: Կոդավորիչը ներկայացնում է փաթույթային շերտերի հաջորդականություն, որը մուտքին ստանալով ձախ նկարը՝ այն հաջորդաբար անցկացնում է փաթույթային շերտերի միջով՝ կոդավորելով պատկերում եղած հիմնական ինֆորմացիան և փոքրացնելով նկարի չափերը: Կոդավորված ինֆորմացիայի մեջ են մտնում նաև ձախ և աջ նկարների անհամատեղելիությունները, որոնցից կարող ենք ստանալ խորությունների քարտեզը՝ օգտվելով հետևյալ հավասարությունից՝

$$d = x - x' = \frac{Bf}{z} \quad (1)$$

Նեյրոնային ցանցի մյուս բաղադրիչը՝ վերծանիչը, փորձում է վերծանել կոդավորված նկարը՝ գտնել ձախ և աջ նկարների անհամատեղելիությունները, ինչպես նաև վերականգնել նկարի սկզբնական չափերը:

Երկրորդ նեյրոնային ցանցը մուտքին ստանում է ձախ պատկերը, ինչպես նաև դրա նախորդ ու հաջորդ պատկերները և փորձում է գտնել այդ պատկերներում շարժումից բխող օբյեկտների հարաբերական տեղափոխությունները: Երկրորդ ցանցը կրկնվող նեյրոնային ցանց է, որի մեջ ներդրված են նաև փաթույթային շերտեր [5]: Նեյրոնային ցանցի այսպիսի կառուցվածքը հնարավորություն է տալիս օգտագործել տեսարանի հաջորդական պատկերներում պարունակվող տեղեկատվությունը: Այնուհետև, նախորդ և հաջորդ պատկերներից սինթեզվում է ձախ պատկերը, ինչի համար օգտագործվում են երկրորդ նեյրոնային ցանցի հաշվարկած օբյեկտների հարաբերական տեղափոխությունները և առաջին նեյրոնային ցանցի հաշվարկած ձախ պատկերի խորությունների քարտեզը: Պատկերների սինթեզը կատարվում է տարածական փոխակերպման մոդուլի միջոցով [6]:

Ցանցի վարժեցման համար որպես նպատակային ֆունկցիա օգտագործվել է վերակառուցված և իրական պատկերների տարբերությունների ստուգման ֆունկցիան: Նեյրոնային ցանցը պատկերների խորությունները գտնում է անուղղակի եղանակով՝ վերակառուցված և իրական պատկերների տարբերությունների չափից ելնելով՝ թարմացվում են նեյրոնային ցանցի պարամետրերը, ինչը ցանցին ստիպում է ստանալ ճշգրիտ խորությամբ քարտեզներ:

Վարժեցման համար օգտագործված կորստի ֆունկցիաները: Վարժեցման կորստի ֆունկցիան բաղկացած է երկու բաղադրիչից՝

$$C = C_1 + \alpha C_2, \quad (2)$$

որտեղ C_1 -ն ընդհանուր կորստի ֆունկցիայի երկկապիճային անհամատեղելիությունների եղանակի բաղադրիչն է [2], C_2 -ը՝ շարժումից պատկերի տեղափոխության եղանակի բաղադրիչը, իսկ α -ն շարժման կարևորությունը բնութագրող գործակից է:

Կորստի ֆունկցիայի C_2 բաղադրիչը որոշվում է հետևյալ բանաձևով՝

$$C_2 = \alpha_{ap}(C_{ap}^p + C_{ap}^n) + \alpha_{ms}(C_{ms}^p + C_{ms}^n), \quad (3)$$

որտեղ C_{ap}^p -ն և C_{ap}^n -ն, համապատասխանաբար նախորդ և հաջորդ նկարներից ձախ նկարի վերակառուցման կորստի ֆունկցիաներն են, C_{ms}^p -ն և C_{ms}^n -ն՝ համապատասխանաբար նախորդ և հաջորդ նկարներում շարժումից բխող, օբյեկտների հարաբերական տեղափոխությունների հարթեցման կորստի ֆունկցիաները, իսկ α_{ap} -ն և α_{ms} -ն համապատասխան կորստի ֆունկցիաների կարևորությունը բնութագրող գործակիցներ են: Կորստի ֆունկցիայի վերևում թվարկված բաղադրիչները որոշվում են հետևյալ բանաձևերով (պարզության համար նշված են միայն նախորդ պատկերի բաղադրիչները)՝

$$C_{ap}^p = \frac{1}{N} \sum_{i,j} \alpha \frac{1-SSIM(I_{ij}^l, \tilde{I}_{ij}^l)}{2} + (1-\alpha) \|I_{ij}^l - \tilde{I}_{ij}^l\|, \quad (4)$$

որտեղ N -ը նկարի կետերի քանակն է, SSIM-ը՝ իրական և նեյրոնային ցանցի միջոցով սինթեզված պատկերների կառուցվածքային նմանությունների ինդեքսը,

$$C_{ms}^p = \frac{1}{N} \sum_{i,j} |\partial_x d_{ij}^l| e^{-\|\partial_x I_{ij}^l\|} + |\partial_y d_{ij}^l| e^{-\|\partial_y I_{ij}^l\|} : \quad (5)$$

Եզրակացություն: Ինչպես և պետք էր ակնկալել տվյալների քանակի ավելացումից և նպատակային ֆունկցիայի բարդացումից, այս մեթոդի վարժեցման փուլն ավելի երկար է՝ համեմատած միայն երկապիճային անհամատեղելիության վարժեցման տևողությանը: Այն կազմել է մոտավորապես 90 ժամ՝ KITTI տվյալների հավաքածուի և մեկ գրաֆիկական մշակող միավորի օգտագործման դեպքում: Բայց առերեսման ժամանակահատվածի խիստ տարբերություններ չեն բացահայտվել: Թեև այս մեթոդի ամենամեծ առանձնահատկությունը պետք է լինի շարժական տեսանկարահանող սարքերի շարժման արագությունից կախված նկարահանած պատկերներում խորության գնահատման ճշտությունը, սակայն այդ գործոնը այս աշխատանքի ընթացքում դեռևս չի դիտարկվել:

ԳՐԱԿԱՆՈՒԹՅԱՆ ՑԱՆԿ

1. **Garg R., BG V.K., Carneiro G., Reid I.** Unsupervised CNN for Single View Depth Estimation: Geometry to the Rescue // ECCV.- 2016. - P. 740-756.
2. **Godard C., Mac Aodha O., Brostow G. J.** Unsupervised Monocular Depth Estimation with Left-Right Consistency // Computer Vision and Pattern Recognition.-2017. – P. 6602-6611.
3. **Liu G., Jiang G., Xiong R., Ou Y.** Binocular Depth Estimation Using Convolutional Neural Network With Siamese Branches // IEEE International Conference on Robotics and Biomimetics (ROBIO).- 2019. - P. 1717-1722.
4. **Kim H.R., Angelaki D.E., DeAngelis G.C.** The neural basis of depth perception from motion parallax // Philosophical Transactions of The Royal Society B Biological Sciences.- 2016.- P. 1697-1708.
5. **Wang R., Pizer S.M., Frahm J.M.** Recurrent Neural Network for (Un-)supervised Learning of Monocular Video, Visual Odometry and Depth // Computer Vision and Pattern Recognition.- 2019. - P. 5555-5564.
6. **Jaderberg M., Simonyan K., Zisserman A.** Spatial Transformer Networks // Advances in Neural Information Processing Systems.- 2015. - P. 2017-2025.

Հայաստանի ազգային պոլիտեխնիկական համալսարան: Նյութը ներկայացվել է խմբագրություն 26.02.2020:

Т.Б. ХАЧАТРЯН

**ОЦЕНКА ГЛУБИНЫ ИЗОБРАЖЕНИЯ С ПОМОЩЬЮ
БЕСКОНТРОЛЬНО ТРЕНИРОВАННОЙ НЕЙРОННОЙ СЕТИ**

Представлен метод оценки глубин изображения путем глубокого обучения и использования нейронных сетей. Для оценки глубин изображения предлагается использовать комбинацию двух методов: оценка глубин с помощью стереоизображений и оценка глубины с помощью видеокамеры. С этой целью была устроена нейронная сеть, которая была тренирована бесконтрольным методом с помощью стереовидеороликов. При применении нейронной сети глубина изображения оценивается с помощью одной картины.

Ключевые слова: глубокое обучение, нейронная сеть, глубина изображения, бесконтрольная тренировка.

T.B. KHACHATRYAN

**ESTIMATING THE IMAGE DEPTH BY AN UNSUPERVISED TRAINED
NEURAL NETWORK**

A method is presented for estimating the depth of images using deep learning and neural networks. The combination of two methods is proposed to estimate the image depths: depth estimation with stereo images and depth estimation with a moving camera. A neural network was built for that purpose which was trained in an unsupervised manner using stereo videos. At inference time, the depth is estimated by one image.

Keywords: deep learning, neural network, image depth, unsupervised training.