

# Στερεοσκοπική όραση με χρήση τεχνητού νευρωνικού δικτύου

Βασίλης Γκολέμης

10/11/2017

## 1 Περιγραφή προβλήματος

Η στερεοσκοπική υπολογιστική όραση προσομοιάζει την μέθοδο με την οποία το μεγαλύτερο μέρος των ζωντανών οργανισμών (αναμεσά τους και ο άνθρωπος) αντιλαμβάνονται το τρισδιάστατο περιβάλλον. Ως πληροφορία εισόδου δεχόμαστε δύο λήψεις σε στερεοσκοπική διάταξη, όπως φαίνεται στην εικόνα 1α'. Τα σημεία που ανήκουν εντός του οπτικού πεδίου και των δύο λήψεων, ακολουθούν την σχέση

$$z = \frac{fB}{x_R - x_L} = \frac{fB}{d} \quad (1)$$

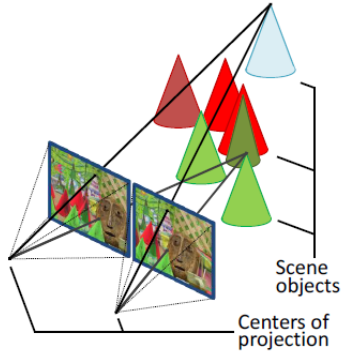
Τα μεγέθη της σχέσης 1 αποτυπώνονται στην εικόνα 1β'. Συμπερασματικά, από το πόσο μετατοπισμένο κατά τον οριζόντιο άξονα (μέγεθος  $d = x_R - x_L$ ) εμφανίζεται κάθε σημείο της εικόνας αναφοράς αντιλαμβανόμαστε πόσο κοντά/μακριά είναι από το πέτασμα της κάμερας (αντιστρόφως ανάλογα). Για παράδειγμα, ο ουρανός που βρίσκεται πάρα πολύ μακριά από τις κάμερες ( $z \rightarrow \infty$ ) θα αποτυπωθεί στο ίδιο σημείο (ίδια τετμημένη) και στις δύο λήψεις ( $d = x_R - x_L \rightarrow 0$ ). Τα μεγέθη  $f$ ,  $b$  είναι σταθερές που περιγράφουν τη στερεοσκοπική διάταξη.

Επομένως, για την επίλυση του προβλήματος οφείλουμε να αντιστοιχίσουμε κάθε σημείο της εικόνας αναφοράς στο «αντίστοιχο» σημείο που αποτυπώθηκε στην έτερη λήψη. Η ακρίβεια με την οποία θα επιτευχθεί αυτή η «αντιστοίχιση» προσδιορίζει την ποιότητα της λύσης.

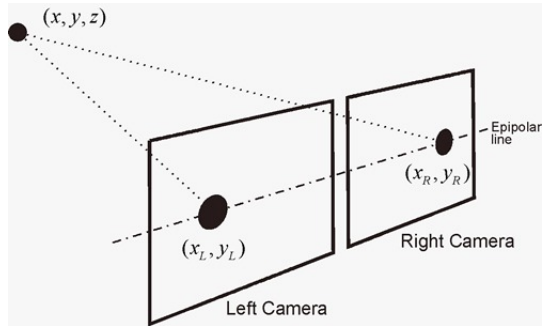
## 2 Επίλυση του προβλήματος

Βασισμένοι σε πρόσφατη βιβλιογραφία, προσεγγίσαμε το πρόβλημα της «αντιστοίχισης» με χρήση τεχνητού συνελκτικού νευρωνικού δικτύου, έναν εξειδικευμένο αλγόριθμο μηχανικής μάθησης σε εικόνες. Στόχος του αλγορίθμου είναι η εκμάθηση από παραδείγματα της σύγκρισης δύο σημείων κι η επιστροφή μιας τιμής (σκορ) ομοιότητας. Αν τα δύο σημεία είναι «αντίστοιχα» (αποτυπώνουν το ίδιο τρισδιάστατο σημείο στην κάθε λήψη) επιθυμούμε το σκορ να μεγιστοποιείται. Αντιθέτως, για δύο τυχαία (αναντίστοιχα) σημεία επιθυμούμε το σκορ να είναι μικρό. Η σύγκριση δύο σημείων γίνεται συγκρίνοντας τις τετράγωνες περιοχές που τα περιβάλλουν (γειτονιές).

Για την εκμάθηση αυτής της διαδικασίας χρησιμοποιήθηκαν οι πιο γνωστές συλλογές εικόνων KITTI 2012, KITTI 2015 και Middlebury. Οι συλλογές αυτές



(α) Παράδειγμα στερεοσκοπικής λήψης



(β') Στερεοσκοπική γεωμετρία

<b>KITTI 2012</b>	
Μέσο σφάλμα	5.787%
Μέσο σφάλμα απόστασης	1.357px
<b>KITTI 2015</b>	
Μέσο σφάλμα	6.545%
Μέσο σφάλμα απόστασης	1.577px
<b>Middlebury</b>	
Μέσο σφάλμα	9.475%
Μέσο σφάλμα απόστασης	2.144px

Πίνακας 1: Συνοπτικός πίνακας αποτελεσμάτων.

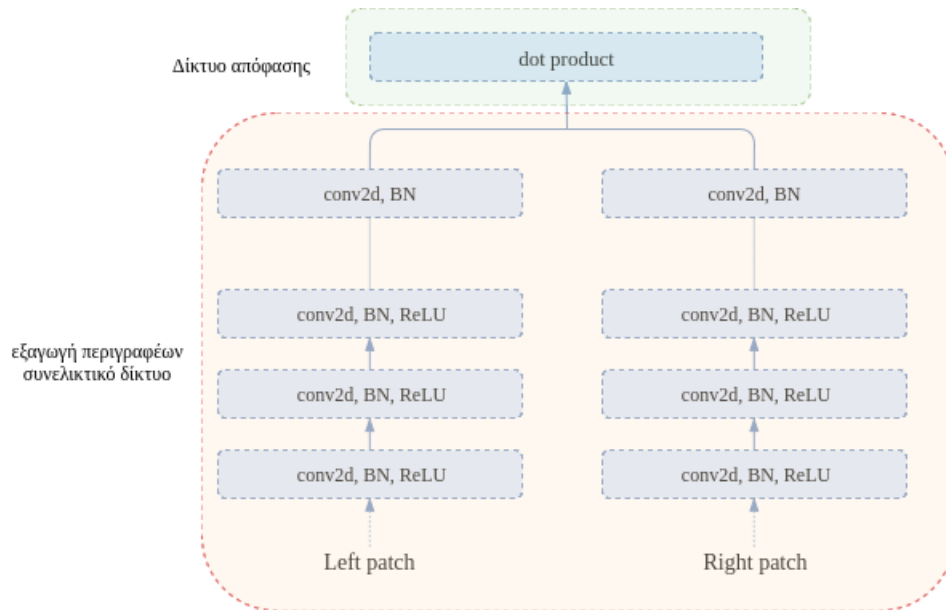
περιέχουν την πραγματική πληροφορία μετατόπισης κάθε σημείου μετρημένη με κατάλληλα εργαλεία, όπως για παράδειγμα το lidar.

Η αρχιτεκτονική του νευρωνικού δικτύου που δημιουργήσαμε φαίνεται στο σχήμα 2. Υπάρχουν 9 συνεκτικτικά επίπεδα, υπεύθυνα για την εξαγωγή του κατάλληλου διανύσματος χαρακτηριστικών (feature vector) κάθε γειτονιάς και μια πράξη εσωτερικού γινομένου υπολογίζει την τιμή ομοιότητας.

Η σύγκριση κάθε σημείου της εικόνας αναφοράς με κάθε πιθανή θέση που μπορεί να αποτυπώθηκε στην έτερη λήψη δημιουργεί τον πίνακα ομοιότητας  $C_{init}(x, y, d)$ . Ακολούθως εφαρμόζουμε τις τεχνικές εξομάλυνσης και βαλτίωσης του αποτελέσματος cross-based cost aggregation, semi-global matching και left-right consistency check. Τα βήματα αυτά δημιουργούν έναν βελτιωμένο πίνακα ομοιότητας  $C_{refined}(x, y, d)$ . Από αυτόν τον πίνακα επιλέγεται για κάθε σημείο  $(x, y)$  η θέση με την μέγιστη ομοιότητα.

### 3 Αποτελέσματα

Συγκεντρικά αποτελέσματα της μεθόδου μας φαίνονται στον πίνακα 1. Ακολούθως φαίνεται ένα ενδεικτικό παράδειγμα εφαρμογής της μεθόδου.



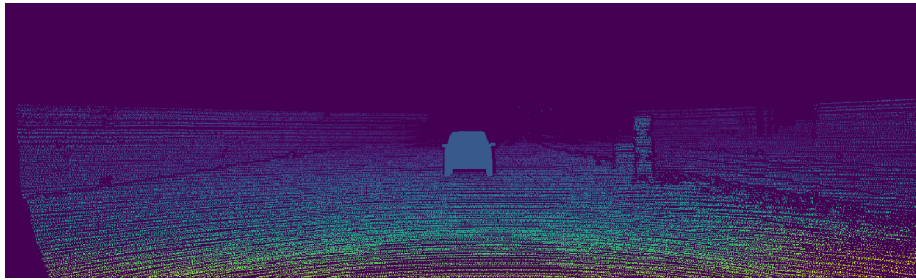
Σχήμα 2: Αρχιτεκτονική νευρωνικού δικτύου

## References

- [1] Spyros Gidaris and Nikos Komodakis. Detect, replace, refine: Deep structured prediction for pixel wise labeling. *arXiv preprint arXiv:1612.04770*, 2016.
- [2] Alex Kendall, Hayk Martirosyan, Saumitro Dasgupta, Peter Henry, Ryan Kennedy, Abraham Bachrach, and Adam Bry. End-to-end learning of geometry and context for deep stereo regression. In *Proceedings of the IEEE International Conference on Computer Vision*, 2017.
- [3] Wenjie Luo, Alexander G. Schwing, and Raquel Urtasun. Efficient deep learning for stereo matching. In *CVPR*, pages 5695–5703. IEEE Computer Society, 2016.
- [4] Jure Zbontar and Yann LeCun. Stereo matching by training a convolutional neural network to compare image patches. *Journal of Machine Learning Research*, 17(1-32):2, 2016.



(α') Εικόνα αναφοράς.



(β') Πραγματική πληροφορία βάθους, όπως μετρήθηκε μέσω lidar.



(γ') Πρόβλεψη πληροφορίας βάθους από την μέθοδό μας.

Σχήμα 3: Ενδεικτικό παράδειγμα.