

0.1 Σχετική βιβλιογραφία

Σε αυτή την ενότητα, παρουσιάζουμε ορισμένες από τις πιο συναφείς μεθόδους για την εργασία μας και άλλες που μελετήθηκαν για τον σχεδιασμό αυτής της προσέγγισης. Οι μέθοδοι αυτές χωρίζονται σε δύο ενότητες *Αναγνώριση Δραστηριότητας* και *Εντοπισμός Δραστηριότητας*. Το πρώτο μέρος αναφέρεται σε κλασικές μεθόδους ταξινόμησης δράσης που εισήχθησαν μέχρι πρόσφατα και το δεύτερο μέρος, αντίστοιχα, σε πρόσφατες μεθόδους εντοπισμού της δράσης.

0.1.1 Αναγνώριση Δραστηριότητας

Οι πρώτες προσεγγίσεις για την κατάταξη της δράσης αποτελούνταν από δύο βήματα α) αρχικά υπολογισμός σύνθετων ‘χειροποίητων’ χαρακτηριστικών από ακατέργαστα καρέ βίντεο και β) εκπαίδευση ενός ταξινομητή με βάση αυτά τα χαρακτηριστικά. Αυτά τα χαρακτηριστικά μπορούν να διαχωριστούν σε 3 κατηγορίες: 1) προσεγγίσεις χωροχρονικού όγκου (space-time volume), 2) τροχιές (trajectories) και 3) χωροχρονικά χαρακτηριστικά. Για τις μεθόδους χωροχρονικού όγκου, η προσέγγιση είναι η εξής: Με βάση τα training βίντεο, το σύστημα συνάπτει ένα μοντέλο τρισδιάστατου χωροχρόνου, συνενώνοντας δυσδιάστατες εικόνες (διάσταση $x-y$) κατά τη διάρκεια του χρόνου (διάσταση t ή z), για την αναπαράσταση κάθε δράσης. Όταν το σύστημα δέχεται ένα βίντεο που δεν έχει ετικέτα, κατασκευάζει μια τρισδιάστατη χωροχρονική αναπαράσταση που αντιστοιχεί σε αυτό το βίντεο. Αυτό η νέα τρισδιάστατη αναπαράσταση, στη συνέχεια, συγκρίνεται με κάθε μοντέλο 3D χωροχρόνου, συγκρίνοντας την ομοιότητα στο σχήμα και την εμφάνιση μεταξύ αυτών των δύο χωροχρονικών όγκων. Το σύστημα εξάγει την κατηγορία του άγνωστου βίντεο, αντιστοιχώντας την με αυτήν της δράσης με την υψηλότερη ομοιότητα. Επιπλέον, υπάρχουν διάφορες παραλλαγές των χωροχρονικών αναπαράστασεων. Αντί της αναπαράστασης space-time volume, το σύστημα μπορεί να αναπαριστά τη κάθε δράση ως τροχιές σε χωροχρονικές διαστάσεις ή ακόμη περισσότερο, η ενέργεια μπορεί να αναπαρασταθεί ως ένα σύνολο χαρακτηριστικών που εξάγονται από τον χωροχρονικό όγκο ή τις τροχιές. Οι ‘καθαρές’ χωροχρονικές αναπαραστάσεις περιλαμβάνουν μεθόδους σύγκρισης των περιοχών προσκήνιου ενός ατόμου (δηλ. σιλουέτες) όπως Bobick and Davis 2001, συγκρίνοντας όγκους σε σχέση με επιφάνεια τους όπως οι Shechtman and Irani 2005. Η μέθοδος Ke, Sukthankar, and Hebert 2007 χρησιμοποιεί oversegmented όγκους, αυτομάτως υπολογίζοντας ένα σύνολο τμημάτων τρισδιάστατου όγκου XYT που αντιστοιχεί σε έναν κινούμενο άνθρωπο. Rodriguez, Ahmed, and Shah 2008 πρότειναν φίλτρα για να αποτυπώνουν τα χαρακτηριστικά του χωροχρονικού όγκου, προκειμένου να τα ταιριάζουν πιο αξιόπιστα και αποδοτικά. Από την άλλη πλευρά, οι προσεγγίσεις με βάση την τροχιά περιλαμβάνουν την αναπαράσταση μιας ενέργειας ως σύνολο 13 κοινών διαδρομών (Sheikh, Sheikh, and Shah 2005) ή τη χρήση ενός συνόλου XYZT-διαστάσεων κοινών τροχιών που λαμβάνονται από κινούμενες κάμερες (Yilmaz and Shah 2005). Τέλος, διάφορες μέθοδοι χρησιμοποιούν τοπικά χαρακτηριστικά που εξάγονται από χωροχρονικούς όγκους τριών διαστάσεων, όπως η εξαγωγή τοπικών χαρακτηριστικών σε κάθε καρέ του βίντεο και η ένωση του χρονικά (Chomat and Crowley 1999; Zelnik-Manor and Irani 2001; Blank

et al. 2005, η εξαγωγή αραιών χωροχρονικών τοπικών σημείων ενδιαφέροντος από τρισδιάστατους όγκους (Laptev and Lindeberg 2003; Dollar et al. 2005; Niebles, Wang, and Li 2006; Alper Yilmaz and Mubarak Shah 2005; Ryoo and Aggarwal 2006) Οι προσεγγίσεις αυτές κατέστησαν την επιλογή των χαρακτηριστικών σημαντικό παράγοντα για την απόδοση του δικτύου. Αυτό συμβαίνει επειδή οι διαφορετικές κατηγορίες δράσεων μπορεί να διαφέρουν δραματικά από την άποψη της εμφάνισής τους και των μοτίβων κίνησης. Ένα άλλο πρόβλημα ήταν ότι οι περισσότερες από αυτές τις προσεγγίσεις κάνουν υποθέσεις, υπό τις οποίες το βίντεο λήφθηκε λόγω προβλημάτων όπως το γεμάτο φόντο, γωνιές κάμερας κλπ. Μια ανασκόπηση των τεχνικών, που χρησιμοποιούνταν μέχρι το 2011, παρουσιάζεται στο Aggarwal and Ryoo 2011.

Τα πρόσφατα αποτελέσματα σε βαθιές αρχιτεκτονικές και ειδικά στον τομέα της ταξινόμησης εικόνας έδωσε κίνητρο στους ερευνητές να εκπαιδεύσουν δίκτυα CNN για το πρόβλημα της αναγνώρισης δράσης. Η πρώτη σημαντική απόπειρα έγινε από τους Karpathy et al. 2014. Σχεδίασαν την αρχιτεκτονική τους με βάση το καλύτερο CNN στον ανταγωνισμό της ImageNet.

0.1.2 Εντοπισμός Δραστηριότητας

Βιβλιογραφία

- [1] Ο. Ήοματ και Θ. Α. Ήρωλεψ. ‘Προβαβιλιστις ρεζογνιτιον οφ αστινιψ υσινγ λοσαλ αππεαφανζε’. Στο: *Προσεεδινγς. 1999 IEEE δμπτυερ Σοσιετιψ δνφερενζε ον δμπτυερ ίσιον ανδ Παττερν Ρεζογνιτιον (ατ. Νο ΠΡ00149)*. Τόμ. 2. 1999, 104–109 όλ. 2. Δοι: 10.1109/ΉΉΠΡ.1999.784616.
- [2] Ααρον Βοβικ και Θ.Ω. Δαις. ‘Τηε ρεζογνιτιον οφ ημμαν μοεμεντ υσινγ τεμποραλ τεμπλατες’. Στο: *Παττερν Αναλψσις ανδ Μασηινε Ιντελλιγεγκες, IEEE Τρανσαστιονς ον 23* (Απρ. 2001), σσ. 257 –267. Δοι: 10.1109/34.910878.
- [3] Α. Ζελνικ-Μανορ και Μ. Ιρανι. ‘Εεντ-βασεδ αναλψσις οφ ιδεο’. Στο: *Προσεεδινγς οφ τηε 2001 IEEE δμπτυερ Σοσιετιψ δνφερενζε ον δμπτυερ ίσιον ανδ Παττερν Ρεζογνιτιον. ΉΉΠ 2001*. Τόμ. 2. 2001, σσ. Π–Π. Δοι: 10.1109/ΉΉΠΡ.2001.990935.
- [4] Λαπτε και Λινδεβεργ. ‘Σπασε-τιμε ιντερεστ ποιητς’. Στο: *Προσεεδινγς Νιντη IEEE Ιντερνατιοναλ δνφερενζε ον δμπτυερ ίσιον*. 2003, 432–439 όλ.1. Δοι: 10.1109/ΙΉΉΉ.2003.1238378.
- [5] Αλπερ Ψιλμαζ και Μυβαρακ Σηαη. ‘Αστιονς σκετςη: α νοελ αστιον ρεπρεσε-ντατιον’. Στο: *2005 IEEE δμπτυερ Σοσιετιψ δνφερενζε ον δμπτυερ ίσιον ανδ Παττερν Ρεζογνιτιον (ΉΉΠ’05)*. Τόμ. 1. 2005, 984–989 όλ. 1. Δοι: 10.1109/ΉΉΠΡ.2005.58.
- [6] Μ. Βλανκ κ.ά. ‘Αστιονς ας σπασε-τιμε σηαπεσ’. Στο: *Τεντη IEEE Ιντερνατιοναλ δνφερενζε ον δμπτυερ ίσιον (ΙΉΉΉ’05) όλυμει 1*. Τόμ. 2. 2005, 1395–1402 όλ. 2. Δοι: 10.1109/ΙΉΉΉ.2005.28.
- [7] Π. Δολλαφ κ.ά. ‘Βεηαιορ ρεζογνιτιον ια σπαρσε σπατιο-τεμποραλ ψεατυρεσ’. Στο: *2005 IEEE Ιντερνατιοναλ Ωορκσηοπ ον ίσυαλ Συρειλλαηζε ανδ Περφορμανζε Εαλυατιον οφ Τραςκινγ ανδ Συρειλλαηζε*. 2005, σσ. 65–72. Δοι: 10.1109/ΉΣΠΕΤΣ.2005.1570899.
- [8] Ε. Σηεζεητμαν και Μ. Ιρανι. ‘Σπασε-τιμε βεηαιορ βασεδ ζορρελατιον’. Στο: *2005 IEEE δμπτυερ Σοσιετιψ δνφερενζε ον δμπτυερ ίσιον ανδ Παττερν Ρεζογνιτιον (ΉΉΠ’05)*. Τόμ. 1. 2005, 405–412 όλ. 1. Δοι: 10.1109/ΉΉΠΡ.2005.328.

- [9] Ψ. Σηεικη, Μ. Σηεικη και Μ. Σηαη. ‘Εξπλορινγ τηε σπαξε οφ α ηυμαν αςτιον’. Στο: *Τεντη IEEE Ιντερνατιοναλ δνφερενςε ον δμπττερ ι̇σιον (I[™]05) δλυμε 1*. Τόμ. 1. 2005, 144–149 ὅλ. 1. ΔΟΙ: 10.1109/I[™].2005.90.
- [10] Α. Ψιλμαζ και Μ. Σηαη. ‘Ρεζογνιζινγ ηυμαν αςτιονς ιν ιδεος αςχυιρεδ βψ υνςαλιβρατεδ μοινγ ζαμερας’. Στο: *Τεντη IEEE Ιντερνατιοναλ δνφερενςε ον δμπττερ ι̇σιον (I[™]05) δλυμε 1*. Τόμ. 1. 2005, 150–157 ὅλ. 1. ΔΟΙ: 10.1109/I[™].2005.201.
- [11] Θυαν ἄρλος Νιεβλες, Ηονγςηενγ Ωανγ και Φει Φει Λι. ‘Υνσυπερισεδ Λε-αρνινγ οφ Ηυμαν Αςτιον ἄτεγοριες Υ̇σινγ Σπατιαλ-Τεμποραλ Ωορδς.’ Στο: τόμ. 79. Σεπτ. 2006, σσ. 1249–1258.
- [12] Μ.Σ. Ρψοο και Θ.Κ. Αγγαρωαλ. ‘Σεμαντις Υ̇νδερστανδινγ οφ δντινυεδ ανδ Ρεζυρσιε Ηυμαν Αςτιυτιες’. Στο: Ιαν. 2006, σσ. 379 –378. ΔΟΙ: 10.1109/I[™]PP.2006.1043.
- [13] Ψ. Κε, Ρ. Συκτθανκαρ και Μ. Ηεβερετ. ‘Σπατιο-τεμποραλ Σηαπε ανδ Φλωω δρρελατιον φορ Αςτιον Ρεζογνιτιον’. Στο: *2007 IEEE δνφερενςε ον δμπττερ ι̇σιον ανδ Παττερν Ρεζογνιτιον*. 2007, σσ. 1–8. ΔΟΙ: 10.1109/I[™]PP.2007.383512.
- [14] Μ. Δ. Ροδριγεζ, Θ. Αημεδ και Μ. Σηαη. ‘Αςτιον ΜΑ[™]Η α σπατιο-τεμποραλ Μαξιμου Αεραγε δρρελατιον Ηειγητ φιλτερ φορ αςτιον ρεζογνιτιον’. Στο: *2008 IEEE δνφερενςε ον δμπττερ ι̇σιον ανδ Παττερν Ρεζογνιτιον*. 2008, σσ. 1–8. ΔΟΙ: 10.1109/I[™]PP.2008.4587727.
- [15] Θ.Κ. Αγγαρωαλ και Μ.Σ. Ρψοο. ‘Ηυμαν Αςτιυτψ Αναλψσις: Α Ρειεω’. Στο: *Α[™]Μ δμπτ. Συρ. 43.3* (Απρ. 2011), 16:1–16:43. ΙΣΣΝ: 0360-0300. ΔΟΙ: 10.1145/1922649.1922653. ΥΡΛ: <http://doi.acm.org/10.1145/1922649.1922653>.
- [16] Α. Καρπατηψ κ.ά. ‘Λαργε-Σζαλε ι̇δεο [™]λασσιφικατιον ωιτη δνολυτιοναλ Νευραλ Νετωορκς’. Στο: *2014 IEEE δνφερενςε ον δμπττερ ι̇σιον ανδ Παττερν Ρεζογνιτιον*. 2014, σσ. 1725–1732. ΔΟΙ: 10.1109/I[™]PP.2014.223.