(19) Οργανισμός Βιομηχανικής Ιδιοκτησίας (ΟΒΙ)

(21) Αριθμός αίτησης: **GR 20230100727**

(12) # ΑΙΤΗΣΗ ΔΙΠΛΩΜΑΤΟΣ ΕΥΡΕΣΙΤΕΧΝΙΑΣ (Α)

(41) Ημ/νία Δημοσιοποίησης: **24.08.2024**

(11) Αριθμός Χορήγησης:

(22) Ημ/νία Κατάθεσης: **11.09.2023**

(30) Προτεραιότητα (ες):
24.02.2023 US 63447972

(43) Ημ/νία Δημοσίευσης της Αίτησης:
06.09.2024 ΕΔΒΙ 8/2024

(71) Αρχικός (οί) Καταθέτης (ες):
MOVERSE I.K.E.; Στέλιου Καζαντζίδη 47, 55535 ΠΥΛΑΙΑ (ΘΕΣΣΑΛΟΝΙΚΗΣ) - GR.

(72) Εφευρέτης (ες):
ΖΙΟΥΛΗΣ ΝΙΚΟΛΑΟΣ; , GR. **ΘΕΡΜΟΣ ΣΠΥΡΙΔΩΝ**; , GR. **ΧΑΤΖΗΤΟΦΗΣ ΑΝΑΡΓΥΡΟΣ**; , GR. **ΑΛΜΠΑΝΗΣ ΓΕΩΡΓΙΟΣ**; , GR.

(73) Δικαιούχος (οι):
MOVERSE I.K.E.; Στέλιου Καζαντζίδη 47, 55535 ΠΥΛΑΙΑ (ΘΕΣΣΑΛΟΝΙΚΗΣ) - GR.

(54) Τίτλος (Ελληνικά)
ΣΥΣΤΗΜΑ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΓΙΑ ΚΑΤΑΓΡΑΦΗ ΚΙΝΗΣΗΣ ΜΕ ΧΡΗΣΗ ΔΕΙΚΤΩΝ

(54) Τίτλος (Αγγλικά)
MACHINE LEARNING SYSTEM FOR MARKER-BASED MOTION CAPTURE

(57) Περίληψη
Η εφεύρεση περιλαμβάνει ένα οπτικό σύστημα βασισμένο στη μηχανική μάθηση και μεθόδους για την εκτέλεση καταγραφής κίνησης. Το σύστημα περιλαμβάνει μια διάταξη οπτικών αισθητήρων που παρέχουν ανεπεξέργαστες, μη δομημένες εισόδους ορόσημων και βασίζεται σε ένα νευρωνικό δίκτυο, εκπαιδευμένο για ισορροπημένη παλινδρόμηση, για να τις αποθορυβοποιήσει, να τις επισημάνει και να τις συμπληρώσει, εκτιμώντας ένα πλουσιότερο σύνολο επεξεργασμένων θέσεων ορόσημων. Το σύστημα και οι μέθοδοι βαθμονομούν αυτόματα το υποκείμενο/ αντικείμενο της καταγραφής και βελτιώνουν αυτόματα τις κινήσεις που καταγράφονται, βελτιώνοντας τη φυσικότητα, την ομαλότητα και διορθώνοντας τις ατελείς της καταγραφής. Το σύστημα και οι μέθοδοι επωφελούνται από ένα ισορροπημένο σύστημα εκπαίδευσης παλινδρόμησης για τη βελτίωση της απόδοσης του εν λόγω νευρωνικού δικτύου σε σύνολα δεδομένων κίνησης με κατανομή μακράς ουράς.

GR 20230100727

ΠΕΡΙΛΗΨΗ

Η εφεύρεση περιλαμβάνει ένα οπτικό σύστημα βασισμένο στη μηχανική μάθηση και μεθόδους για την εκτέλεση καταγραφής κίνησης. Το σύστημα περιλαμβάνει μια διάταξη οπτικών αισθητήρων που παρέχουν ανεπεξέργαστες, μη δομημένες εισόδους ορόσημων και βασίζεται σε ένα νευρωνικό δίκτυο, εκπαιδευμένο για ισορροπημένη παλινδρόμηση, για να τις αποθορυβοποιήσει, να τις επισημάνει και να τις συμπληρώσει, εκτιμώντας ένα πλουσιότερο σύνολο επεξεργασμένων θέσεων ορόσημων. Το σύστημα και οι μέθοδοι βαθμονομούν αυτόματα το υποκείμενο/αντικείμενο της καταγραφής και βελτιώνουν αυτόματα τις κινήσεις που καταγράφονται, βελτιώνοντας τη φυσικότητα, την ομαλότητα και διορθώνοντας τις ατέλεις της καταγραφής. Το σύστημα και οι μέθοδοι επωφελούνται από ένα ισορροπημένο σύστημα εκπαίδευσης παλινδρόμησης για τη βελτίωση της απόδοσης του εν λόγω νευρωνικού δικτύου σε σύνολα δεδομένων κίνησης με κατανομή μακράς ουράς.

ΤΙΤΛΟΣ ΤΗΣ ΕΦΕΥΡΕΣΗΣ

[0001] Σύστημα Μηχανικής Μάθησης για Καταγραφή Κίνησης με χρήση Δεικτών

ΠΕΔΙΟ ΤΗΣ ΕΦΕΥΡΕΣΗΣ

[0002] Η παρούσα εφεύρεση αφορά γενικά τον τομέα των τεχνολογιών καταγραφής κίνησης και ειδικότερα τα συστήματα καταγραφής κίνησης σε πραγματικό χρόνο με χρήση δεικτών. Συγκεκριμένα, το σύστημα επεξεργάζεται δεδομένα εισόδου τύπου εικόνας, για να εκτιμήσει σε πραγματικό χρόνο προκαθορισμένα σημεία αναφοράς (ορόσημα) του καταγραφόμενου σώματος, που ανήκουν είτε στην επιφάνεια του είτε στο εσωτερικό του. Το σύστημα εκτιμά πρώτα το σχήμα του σώματος των καταγραφόμενων υποκειμένων χρησιμοποιώντας τα εκτιμώμενα ορόσημα και στη συνέχεια το χρησιμοποιεί για να υπολογίσει την άρθρωση του υποκειμένου στον χρόνο, παρέχοντας μια πλήρη 4Δ χρονικά μεταβαλλόμενη καταγραφή, η οποία γενικά σχετίζεται με τον ευρύτερο τομέα της καταγραφής ερμηνειών. Ενώ το σύστημα βασίζεται στη μηχανική μάθηση για την εκτίμηση των ορόσημων του σώματος από τις μετρήσεις θέσης των δεικτών εισόδου, δεν απαιτεί δεδομένα θέσης των δεικτών για την εκπαίδευσή του. Αντ' αυτού, αξιοποιεί μια παραμετρική αναπαράσταση του μοντέλου σώματος για την προσομοίωση των δεδομένων θέσης του δείκτη και καταγράφει τις απαραίτητες παραμέτρους του μοντέλου σώματος εκπαίδευσης χρησιμοποιώντας μια μέθοδο καταγραφής κίνησης χωρίς δείκτες. Περαιτέρω, με τη χρήση γεννητικών μοντέλων, το σύστημα χρησιμοποιεί μια νέα τεχνική για την εξισορρόπηση της εκπαίδευσης του μοντέλου σε σχέση με τη μεροληψία του και την υποκείμενη κατανομή δεδομένων. Τελικά, το παραμετρικό μοντέλο σώματος τυποποιεί τις εξόδους του συστήματος σε σχέση με την αναπαράστασή του, επιτρέποντας τα αποτελέσματα σε πραγματικό χρόνο να βελ-

τιώνονται αυτόματα μέσω μιας διαδικασίας μετα-επεξεργασίας που βασίζεται σε μια ανθεκτική και προσαρμοζόμενη στον θόρυβο μέθοδο δέσμης βελτίωσης.

ΥΠΟΒΑΘΡΟ ΤΗΣ ΕΦΕΥΡΕΣΗΣ

[0003] Η καταγραφή κίνησης (Motion Capture, ή εν συντομία, MoCap) είναι η τεχνολογία που ψηφιοποιεί την κίνηση άκαμπτων ή αρθρωτών φυσικών οντοτήτων ή αντικειμένων. Αυτά μπορεί να είναι άκαμπτα (π.χ. αντικείμενα) ή αρθρωτά (π.χ. άνθρωποι, ζώα). Η τεχνολογία βρίσκει πολυάριθμες εφαρμογές, με παραδείγματα όπως τη μεταφορά της ψηφιοποιημένης πληροφορίας κίνησης σε εικονικά περιβάλλοντα για προσομοίωση οδήγησης (όπως συνθετικοί προσχηματισμένοι κόσμοι για την απόδοση δεδομένων για αυτόνομη οδήγηση), ή την αύξηση του ρεαλισμού του απεικονιζόμενου περιεχομένου (όπως σε βιντεοπαιχνίδια), ή εφέ σύνθεσης περιεχομένου (όπως εικονικά εφέ στη βιομηχανία του κινηματογράφου), ή και την ανάλυση των στατιστικών ιδιοτήτων της ανθρώπινης κίνησης και των χαρακτηριστικών που απορρέουν από αυτήν (όπως στην έρευνα για την υγεία ή την ανθρώπινη κινησιολογία) και άλλα παρεμφερή.

[0004] Υπάρχουν διάφοροι τρόποι υλοποίησης μιας τεχνολογίας καταγραφής κίνησης και συνήθως κατηγοριοποιούνται σε σχέση με τον εξοπλισμό (π.χ. αισθητήρες, εξαρτήματα) που αναπτύσσεται και χρησιμοποιείται για την καταγραφή της κίνησης των υποκειμένων (ή των αντικειμένων). Από πλευράς αισθητήρων, χρησιμοποιούνται μονάδες αδρανειακής μέτρησης και κάμερες, με τις πρώτες να τοποθετούνται πάνω στο σώμα, συνήθως προσαρτημένες σε μια στολή που φοράει το υποκείμενο, και τις δεύτερες να στοχεύουν συνήθως σε μια στολή που έχει προσαρτημένους δείκτες αντί για αισθητήρες. Οι δείκτες μπορεί να είναι είτε ενεργοί, εκπέμποντας σήματα αυτοπροσδιορισμού που αποκωδικοποιούνται από τις κάμερες, είτε παθητικοί, βασιζόμενοι σε φυσικές ιδιότητες (π.χ. ρετρο-ανακλαστικό υλικό) για να βελτιώσουν τη δυνατότητα καταγραφής τους από τις κάμερες. Για να επιτευχθεί αυτό, απαιτούνται κάμερες με

υπέρυθρους προβολείς, των οποίων το φάσμα λήψης είναι περιορίσμένο σε συγκεκριμένο έυρος, ώστε να ταιριάζει με τις ιδιότητες του εκπεμπόμενου φωτός. Τα συστήματα που βασίζονται σε δείκτες αναφέρονται ως συστήματα MoCap με χρήση δεικτών, ενώ υπάρχουν και συστήματα με βάση τις κάμερες που δεν απαιτούν από τα υποκείμενα να φορούν στολές και δείκτες και λειτουργούν απλώς με τα δεδομένα εικόνας που παρέχουν οι κάμερες. Τα συστήματα αυτά αναφέρονται ως συστήματα χωρίς δείκτες και εκμεταλλεύονται ενδιάμεσες αναπαραστάσεις πληροφορίας που εκτιμώνται από αλγορίθμους που λειτουργούν σε ανεπεξέργαστα δεδομένα εικόνας.

[0005] Η τεχνολογία MoCap έχει επωφεληθεί από τις εξελίξεις της τελευταίας δεκαετίας που βασίζονται στην εκμετάλλευση των δεδομένων, κυρίως λόγω της σημαντικής έρευνας για την ανθρωποκεντρική κατανόηση με χρήση οπτικής πληροφορίας που επικεντρώνεται στην απερίσπαστη καταγραφή με τη χρήση ανεπεξέργαστων εισόδων έγχρωμων εικονών (δηλαδή χωρίς δείκτες). Ο χρυσός κανόνας της τεχνολογίας MoCap με χρήση "οπτικής" πληροφορίας εξακολουθεί να χρησιμοποιεί δείκτες προσαρτημένους σε στολή για καταγραφή κίνησης με ακρίβεια και με ανοχή στο θόρυβο, και έχει λάβει ελάχιστη προσοχή στη βιβλιογραφία.

[0006] Στο υπόλοιπο αυτής της ενότητας θα συζητηθούν σύχρονες τεχνολογίες MoCap με χρήση δεικτών και μηχανικής μάθησης με σκοπό την ευκολότερη κατανόηση των πλεονεκτημάτων της εφεύρεσης από τον αναγνώστη.

[0007] Τα συστήματα MoCap που βασίζονται στη μηχανική μάθηση και χρησιμοποιούν δείκτες επικεντρώνονται κυρίως στην επεξεργασία (ανεπεξέργαστων) αρχειακών δεδομένων MoCap με σκοπό την άμεση επισήμανσή τους [*Ghorbani N. et al., 2019*] ή την επισήμανσή τους μέσω παλινδρόμησης [*Han et al., 2018*], την επίλυση των αρθρώσεων του σκελετού [*Chen et al., 2021*] ή των μετασχηματισμών των αρθρώσεων [*Holden, 2018*], καθώς και τη χρήση χαμηλού κόστους αισθητήρων που εισάγουν υψηλότερα επίπεδα θορύβου [*Chatzitofis et al., 2021*].

[0008] Καθώς ακόμη και τα συστήματα υψηλής τεχνολογίας παράγουν αποτελέ-

σματα με ποικίλα επίπεδα θορύβου, είτε πρόκειται για θόρυβο στο επίπεδο της ίδιας της πληροφορίας των δεικτών (εναλλαγές, αποκρύψεις και λανθασμένες εμφανίσεις), είτε στις μετρήσεις (τρεμούλιασμα, μετατοπίσεις θέσης), όλες οι προαναφερθείσες προσεγγίσεις εκμεταλλεύονται την απλή φύση της αναπαράστασης των ανεπεξέργαστων δεικτών για να προσθέσουν συνθετικό θόρυβο κατά τη διάρκεια της εκπαίδευσης. Ακόμα, για τα συστήματα που βασίζονται σε δεδομένα, η μεταβλητότητα της τοποθέτησης των δεικτών (δηλ. των διατάξεων) αποτελεί μια άλλη πρόκληση που πρέπει να ξεπεραστεί.

[0009] Είναι δυνατόν να αντιμετωπιστεί αυτό εμμέσως, βασιζόμενοι στη διαδικασία μάθησης, ή να αντιμετωπιστεί σχεδόν εμμέσως, θεωρώντας τα ως είσοδο στο μοντέλο. Ένας άλλος τρόπος αντιμετώπισης αυτού του προβλήματος περιλαμβάνει την προσαρμογή των ανεπεξέργαστων δεδομένων σε ένα παραμετρικό μοντέλο μετά από χειροκίνητη [*Han et al., 2018*, *Loper et al., 2014*] ή αυτόματη [*Ghorbani N. et al., 2019*] επισήμανση ή/και επισημείωση αντιστοιχιών, τυποποιώντας την υποκείμενη αναπαράσταση.

[0010] Η επίλυση των θέσεων των αρθρώσεων από τα δεδομένα των δεικτών αποτελείται από μία σειρά (ενίοτε προαιρετικών) βημάτων. Οι δείκτες πρέπει να επισημανθούν (ή αλλιώς ταυτοποιηθούν), οι καταγεγραμμένοι δείκτες που δεν αντιστοιχούν σε κάποια γνωστή επισήμανση (ή αλλιώς δεν ταυτοποιούνται, αποκαλούμενοι δείκτες-φαντάσματα) πρέπει να εντοπιστούν, οι δείκτες οποίοι δεν καταγράφηκαν (είτε γιατί δεν εκτός της θέασης των καμερών, είτε γιατί τους αποκρύπτει το σώμα) πρέπει να προβλεφθούν, και στη συνέχεια πρέπει να προσαρμοστεί μια σκελετική δομή ή μια δομή άκαμπτων σωμάτων στα παρατηρούμενα δεδομένα των δεικτών. Διάφορες δουλειές αντιμετωπίζουν σφάλματα σε διάφορα στάδια της επίλυσης του MoCap, με τις πιο σύγχρονες να βασίζονται σε περιορισμούς που σχετίζονται με την ομαλότητα και τα χαρακτηριστικά των οστών (γωνίες, μετατοπίσεις και μήκη).

[0011] Άλλες προσεγγίσεις καταφεύγουν σε λύσεις που βασίζονται σε προ-υπάρχοντα

δεδομένα για την αρχικοποίηση [*Schubert et al., 2015*] ή τον καθαρισμό των δεικτών [*Aristidou et al., 2018*]. Το MoSh [*Loper et al., 2014*], αντίθετα, χρησιμοποίησε ένα παραμετρικό μοντέλο του ανθρώπινου σώματος για την επίλυση ανεπεξέργαστων δεδομένων δεικτών και την εκτίμηση της άρθρωσης της πόζας και των θέσεων των αρθρώσεων, λαμβάνοντας υπόψη ακόμη και ασυνέπειες στη διάταξη των δεικτών και/ή την κίνηση του μαλακού ιστού.

[0012] Παρ' όλα αυτά, η έλευση των σύγχρονων τεχνολογιών που βασίζονται σε δεδομένα (βαθιά μηχανική μάθηση) έχει δώσει ώθηση σε νέες προσεγγίσεις για την επίλυση των αρθρώσεων στό MoCap. Μια προσέγγιση επισήμανσης μέσω παλινδρόμησης [*Han et al, 2018*] χρησιμοποίησε ένα βαθύ μοντέλο για να εκτιμήσει τις θέσεις των δεικτών και στη συνέχεια, λύνοντας ένα πρόβλημα βέλτιστης (μέγιστης) τοποθέτησης να επιτύχει την επισήμανση των δεικτών εισόδου. Η επισήμανση έχει επίσης εκφραστεί και ως μάθηση μεταλλαγών [*Ghorbani S. et al., 2019*], αν και στη συγκεκριμένη περίπτωση χρησιμοποιήθηκαν περιορισμοί στο σήμα εισόδου, οι οποίοι μπορούν να χαλαρώσουν με την προσθήκη μιας επιπλέον κατηγορίας δεικτών-φαντασμάτων προς επισήμανση [*Ghorbani N. et al., 2019*], [*Black et al., 2023*].

[0013] Ωστόσο, η επισήμανση προϋποθέτει ότι τα ανεπεξέργαστα δεδομένα είναι ορισμένης ποιότητας, καθώς στη συνέχεια χρησιμοποιούνται για την επίλυση των μετασχηματισμών των αρθρώσεων ή αλλιώς απαιτούνται επιπλέον βήματα επεξεργασίας για την αποθορυβοποίηση της εισόδου. Ως αποτέλεσμα, οι ανεπάλληλες προσεγγίσεις (μηχανικής μάθησης) με βάση τα δεδομένα αποτέλεσαν μια παράλληλη γραμμή έρευνας, στην οποία μελετήθηκε ταυτόχρονη αποθορυβοποίηση και επίλυση.

[0014] Παρόλο που είναι δυνατόν να εκτελεστεί διαδοχικός καθαρισμός και επίλυση με τη χρήση ενός μόνο νευρωνικού δικτύου πρόσω τροφοδότησης [*Holden, 2018*], η διαδικασία επωφελείται φυσικά από τη χρήση δύο νευρωνικών δικτύων τύπου αυτοκωδικοποιητών [*Pavllo et al., 2019*], με τον πρώτο να λειτουργεί σε δεδομένα δεικτών και να τα καθαρίζει για το επόμενο δίκτυο που εκτελεί την παλινδρόμηση των θέ-

σεων των αρθρώσεων. Η σταδιοποίηση αυτή από τους δείκτες στις αρθρώσεις έχει αποδειχθεί ότι βελτιώνει την ακρίβεια ακόμη και όταν χρησιμοποιούνται συνελικτικά νευρωνικά δίκτυα, αλλά αυτό έχει ως κόστος τις μειωμένες επιδόσεις όσον αφορά τον χρόνο εκτέλεσης [*Chatzitofis et al., 2021*]. Τα συνελικτικά δίκτυα γράφων είναι μία άλλη υποψήφια προσέγγιση [*Chen et al., 2021*], καθώς επιτρέπουν την κωδικοποίηση τοπολογικών χαρακτηριστικών όπως η ρητή διάταξη των δεικτών και η ιεραρχία του σκελετού, δύο σημαντικές παραλλαγές που αντιμετωπίζονταν μόνο έμμεσα στις προαναφερθέντες ανεπάλληλες προσεγγίσεις.

[0015] Η εκμάθηση της επίλυσης του MoCap με τη χρήση δεδομένων δεικτών απαιτεί επίβλεψη που παρέχεται από τη συλλογή δεδομένων με επαγγελματικά συστήματα MoCap υψηλής τεχνολογίας. Τα δεδομένα δεικτών μπορούν εύκολα να (επανα)συντίθενται όταν είναι διαθέσιμες πληροφορίες υψηλότερου επιπέδου (π.χ. μετατοπίσεις δεικτών προς αρθρώσεις, τρισδιάσταταν μοντέλα με τη μορφή πλέγματος), αλλά είναι επίσης δυνατό να χρησιμοποιηθούν συνθετικές προσαρμογές παραμετρικών μοντέλων σε πραγματικά δεδομένα, ακόμη και όταν χρησιμοποιούνται αισθητήρες χαμηλότερης ποιότητας, για την απόκτηση και στη συνέχεια τη δημιουργία συνθετικών δεδομένων δεικτών και αρθρώσεων.

[0016] Τα στατιστικά παραμετρικά μοντέλα [*Loper et al., 2015*] (SMPL), [*Osman et al., 2020*] (STAR), [*Osman et al., 2022*] (SUPR), [*Wang et al., 2021*] (PanoMan), [*Xu et al., 2020*] (GHUM), [*Yan et al., 2021*] (DeepDaz) είναι πιο εκφραστικές εναλλακτικές λύσεις από τα απλά δυναμικά 3Δ πλέγματα, καθώς εκτός από πιο ρεαλιστικές παραλλαγές σχήματος, μπορούν να αξιοποιηθούν διορθωτικοί παράγοντες παραμόρφωσης για τη βελτίωση του αποτελέσματος. Έχουν χρησιμοποιηθεί στο παρελθόν για τη σύνθεση δεδομένων εκπαίδευσης αλγορίθμων μηχανικής μάθησης για φωτορεαλιστική απόδοση [*Varol et al., 2017*] και προσομοίωση μετρήσεων από αισθητήρες [*Kaufmann et al., 2021*]. Όμως, και αυτά βασίζονται σημαντικά στην απόκτηση MoCap δεδομένων υψηλής ποιότητας.

[0017] Αυτές οι παραμετρικές αναπαραστάσεις μοντέλων χαμηλής διάστασης μπορούν να προσαρμοστούν σε δεδομένα δεικτών, με αξιοσημείωτα παραδείγματα το σύνολο δεδομένων AMASS [*Mahmood et al., 2019*] που ενοποιεί μια ποικιλία συνόλων δεδομένων MoCap χρησιμοποιώντας την οικογένεια μοντέλων SMPL και τη βάση δεδομένων Fit3D [*Fieraru et al., 2021*] χρησιμοποιώντας το μοντέλο GHUM. Η δυνατότητα απόκτησης τέτοιων δεδομένων με τη χρήση λιγότερο δαπανηρών λύσεων καταγραφής είναι πολύ σημαντική, εφόσον είναι εφικτή η εκπαίδευση μοντέλων υψηλής ποιότητας. Παρόλα αυτά, πρέπει επίσης να ληφθεί υπόψη η φύση των δεδομένων ανθρώπινης ερμηνείας και οι διαδικασίες συλλογής τους. Όπως φαίνεται τόσο στη βάση δεδομένων AMASS όσο και στη Fit3D, αυτά περιέχουν σημαντικούς πλεονασμούς και πλήττονται από κατανομές μακράς ουράς.

[0018] Η πρόβλεψη σπάνιων θέσεων αποτελεί πρόκληση για τα μοντέλα παλινδρόμησης, μια πρόκληση που απορρέει κυρίως από τη συνδυασμένη επίδραση των επιλεγμένων εκτιμητών και της στοχαστικής βελτιστοποίησης με μικρές ομάδες/παρτίδες δεδομένων. Διάφορες λύσεις έχουν εμφανιστεί στη βιβλιογραφία, ορισμένες προσαρμοσμένες στη φύση του προβλήματος [*Rong et al., 2022*], αξιοποιώντας έναν ταξινομητή πρωτοτύπων για την αρχικοποίηση μιάς διαδικασίας επαναληπτικής βελτίωσης μέσω μηχανικής μάθησης, αλλά και άλλες, οποίες προσαρμόζουν εργασίες από την μη ισορροπημένη ταξινόμηση στον τομέα της παλινδρόμησης.

[0019] Οι παραδοσιακές προσεγγίσεις ανήκουν είτε στην κατηγορία της επαναδειγματοληψίας είτε στην κατηγορία της επαναστάθμισης, με την πρώτη να επικεντρώνεται στην εξισορρόπηση της συχνότητας των δειγμάτων και τη δεύτερη στη σωστή προσαρμογή της διαδικασίας βελτιστοποίησης των παραμέτρων.

[0020] Οι στρατηγικές επαναδειγματοληψίας περιλαμβάνουν την υποδειγματοληψία των πιο κοινών δειγμάτων [*Torgo et al., 2015*], την υπερδειγματοληψία των πιο σπάνιων δειγμάτων χρησιμοιώντας τη σύνθεση νέων σπάνιων δειγμάτων μέσω παρεμβολής [*Torgo et al., 2013*], την επαναδειγματοληψία τους μετά από την εισαγωγή θορύβου, αλλά και

υβριδικές προσεγγίσεις που κάνουν ταυτόχρονα υπο- και υπερδειγματοληψία [*Branco et al., 2017*]. Ωστόσο, η παρεμβολή δειγμάτων υψηλών διαστάσεων, όπως η ανθρώπινη πόζα, είναι προβληματική όπως και ο ίδιος ο καθορισμός των σπάνιων δειγμάτων που πρέπει να επαναδειγματοληφθούν.

[0021] Η παλινδρόμηση με βάση τη χρησιμότητα [*Torgo et al., 2007*] - ή αλλιώς, με ευαισθησία στο κόστος - αποδίδει διαφορετικά βάρη - ή σημασία - σε διαφορετικά δείγματα. Ο ορισμός μιας συνάρτησης χρησιμότητας είναι επίσης σημαντικός για τις στρατηγικές επαναληπτικής δειγματοληψίας για την παλινδρόμηση [*Torgo et al., 2015*]. Πρόσφατες προσεγγίσεις χρησιμοποιούν την εκτίμηση πυκνότητας πυρήνα [*Steininger et al., 2021*], είτε προσαρμόζουν τις μετρικές αξιολόγησης ως συναρτήσεις κόστους [*Silva et al., 2022*] ή καταφεύγουν στην εξομάλυνση κατηγοριών/χαρακτηριστικών και στην ομαδοποίηση [*Yang et al., 2021*].

[0022] Μια άλλη οικογένεια μεθόδων που διερευνώνται τώρα μπορούν να κατηγοριοποιηθούν ως αντιθετικές, με τους [*Gong et al., 2022*] να κανονικοποιούν την εκπαίδευση των αλγορίθμων μηχανικής μάθησης για να επιβάλλουν την εγγύτητα των ενδιάμεσων χαρακτηριστικών και του χώρου χαρακτηριστικών. Μια ισορροπημένη εκδοχή του μέσου τετραγωνικού σφάλματος (ΜΤΣ) [*Ren et al., 2022*] είναι επίσης ένας στόχος που μοιάζει με αντιθετικό και χρησιμοποιεί την ταξινόμηση δειγμάτων ελάχιστου σφάλματος εντός κάθε μικρής ομάδας/παρτίδας χρησιμοποιώντας έναν όρο διασταυρούμενης εντροπίας που αντιστοιχεί σε ένα σφάλμα L2 σύμφωνα με την θεωρία πιθανοτήτων.

[0023] Ωστόσο, οι περισσότερες προσεγγίσεις βασίζονται στην ταξινόμηση του χώρου εξόδου, η οποία, όσον αφορά ανθρώπινη πόζα, είναι ένα δύσκολο έργο.

[0024] Επιπλέον, η καταγραφή κίνησης ανά καρέ είναι επιρρεπής σε θορυβώδεις εκτιμήσεις ή προσαρμογή σε θορυβώδεις παρατηρήσεις. Ενώ το χρονικό φιλτράρισμα είναι μια επιλογή για την επίλυση του χρονικού τρέμουλου και της μη συνέπειας των καταγραφών ανά στιγμιότυπα [*Ingwersen et al., 2023*], είναι πολύ δύσκολο να σχεδια-

στούν φίλτρα για να χειριστούν τις αποκρύψεις και τις ακραίες τιμές. Οι περισσότερες προσεγγίσεις ενσωματώνουν χρονικούς περιορισμούς, όπως η ομαλότητα στις αρθρώσεις [*Arnab et al., 2019*], [*Ye et al., 2023*], ή η σταθερότητα της ταχύτητας [*Loper et al., 2014*], [*Zanfir et al., 2018*], [*Mahmood et al., 2019*], κατά την επίλυση της πόζας και της κίνησης, συμπληρωματικά με την εκμετάλλευση της αιτιότητας με την ταυτόχρονη επίλυση ενός αριθμού καρέ. Ένας όρος ομαλότητας DCT έχει επίσης χρησιμοποιηθεί κατά την προσαρμογή του σώματος SMPL σε πολλαπλές εικόνες [*Huang et al., 2017*]. Πιο πρόσφατες προσεγγίσεις βασίζονται σε μοντέλα προγενέστερης γνώσης που εκπαιδεύτηκαν με δεδομένα, όπως ο χώρος χαρακτηριστικών ενός αυτόματου κωδικοποιητή ταχύτητας [*Huang et al., 2022*], ο λανθάνων χώρος των πόζων [*Huang et al., 2021*], της κίνησης [*Saini et al., 2023*], ή ο λανθάνων χώρος και η κατάσταση μετάβασης μιας αυτοπαλινδρομικής παρελθοντικής κίνησης [*Rempe et al., 2021*]. Ένα κοινό στοιχείο όλων των προηγούμενων προσεγγίσεων είναι ότι η ομαλότητα της κίνησης επιτυγχάνεται με την προσθήκη ενός ή περισσότερων στόχων στο πρόβλημα βελτιστοποίησης. Αυτό συνεπάγεται το κόστος των πρόσθετων παραμέτρων που χρειάζονται ρύθμιση, μια περίπτωση που είναι ιδιαίτερα περίπλοκη όταν λαμβάνεται υπόψη ο μεγάλος αριθμός στόχων που εμπλέκονται κατά την προσαρμογή ενός παραμετρικού μοντέλου σώματος σε εικόνες.

[0025] Η προσθήκη τέτοιων χρονικών περιορισμών είναι πιο αποτελεσματική όταν επιλύεται ταυτόχρονα για μια ομάδα καρέ. Το μοντέλο προγενέστερης γνώσης βασισμένο σε Διακριτό Μετασχηματισμό Συνημιτόνου (ΔΜΣ) που χρησιμοποιήθηκε από τους [Huang et al., 2017] προστέθηκε στο δεύτερο στάδιο της βελτιστοποίηση όταν επιλύθηκε ταυτόχρονα μια ομάδα 30 καρέ. Ομοίως, επιβλήθηκε ομαλότητα στις αρθρώσεις σε μια ομάδα καρέ που επιλύθηκαν ταυτόχρονα [*Saini et al., 2023*], [*Ye et al., 2023*]. Εκτός από την επίλυση μιας μικρής ομάδας, έχει επίσης αποδειχθεί ότι είναι δυνατή η ταυτόχρονη επίλυση ολόκληρων βίντεο [*Arnab et al., 2019*], αξιοποιώντας ολόκληρο το χρονικό πλαίσιο, ακόμη και με πολύ σύνθετες συναρτήσεις σκοπού [*Huang et al.,*

2022] που περιλαμβάνουν αλληλεπιδράσεις ανθρώπου-αντικειμένου. Παρόλα αυτά, αυτό απαιτεί καλή αρχικοποίηση, η οποία μπορεί να επιτευχθεί χρησιμοποιώντας ένα μοντέλο παλινδρόμησης [*Arnab et al., 2019*] ή ένα στάδιο προσαρμογής σώματος ανά καρέ [*Huang et al., 2022*]. Αυτός ο τύπος σταδιακής βελτιστοποίησης είναι συνηθισμένος σε περιπτώσεις δεδομένων από μία κάμερα [*Pavlakos et al., 2019*], αλλά βρίσκει επίσης ευρεία χρήση σε ρυθμίσεις πολλαπλών καμερών [*Huang et al., 2017*], [*Saini et al., 2023*]. Όσον αφορά στα βίντεο, χρησιμοποιούνται πολλαπλά στάδια για την αρχικοποίηση των εκτιμήσεων για κάθε ξεχωριστό καρέ, και στη συνέχεια επιλύονται δέσμες χρησιμοποιώντας περιορισμούς ομαλότητας [*Huang et al., 2017*], [*Saini et al., 2023*], [*Ye et al., 2023*] , [*Arnab et al., 2019*], είτε μέσω άμεσων μοντέλων παλινδρόμησης είτε μέσω επίλυσης ενός προβλήματος βελτιστοποίησης για κάθε καρέ.

[0026] Ωστόσο, αυτό είναι μάλλον υπολογιστικά δαπανηρό, καθώς απαιτούνται πολλαπλά περάσματα πάνω από τα βίντεο. Επιπλέον, η χρονική φύση των δεδομένων λαμβάνεται υπόψη μόνο σε επίπεδο περιορισμών, αλλά όχι σε επίπεδο παραμέτρων, με την κατάσταση κάθε καρέ να επιλύεται ξεχωριστά στη δέσμη.

ΣΥΝΟΨΗ ΤΗΣ ΕΦΕΥΡΕΣΗΣ

[0027] Η παρούσα εφεύρεση αποκαλύπτει ένα σύστημα καταγραφής κίνησης με χρήση δεικτών σε πραγματικό χρόνο που εκτιμά γνωστά ορόσημα ενός αντικειμένου ή υποκειμένου και επιπλέον την πόζα και το σχήμα του χρησιμοποιώντας αισθητήρα(-ες) που βασίζονται σε πληροφορία εικόνας, παρέχοντας ουσιαστικά μια τετραδιάστατη (4Δ) αναπαράσταση της κίνησης του αντικειμένου ή/και της απόδοσης του υποκειμένου, δηλαδή την τρισδιάστατη (3Δ) χρονικά μεταβαλλόμενη κίνησή του στη διάσταση του χρόνου.

[0028] Το σύστημα λειτουργεί σε δύο καταστάσεις, είτε σε πραγματικό χρόνο, όπου η είσοδος επεξεργάζεται με διαδραστικούς ρυθμούς, προσφέροντας μια ζωντανή

απεικόνιση της εξόδου, είτε μετά την καταγραφή, όπου οι εκτιμήσεις που έγιναν σε πραγματικό χρόνο βελτιώνονται.

[0029] Η αποκαλυπτόμενη εφεύρεση χρησιμοποιεί μια βάση δεδομένων παραμέτρων νευρωνικών δικτύων και μια βάση δεδομένων μοντέλων σώματος και χρησιμοποιεί κατάλληλα ταιριαστούς συνδυασμούς τους για την καταγραφή του αντικειμένου/υποκειμένου.

[0030] Σε μία υλοποίηση, το σύστημα παράγει μόνο προκαθορισμένα ορόσημα σε πραγματικό χρόνο χρησιμοποιώντας την ανεπεξέργαστη πληροφορία της θέσης των δεικτών, όπως αποκτώνται και εκτιμώνται από τον αισθητήρα (ή τους αισθητήρες) με βάση την εικόνα. Το σύστημα επεξεργάζεται τα παρατηρούμενα μη δομημένα ορόσημα και παράγει υψηλότερης ποιότητας επισημασμένες εκτιμήσεις.

[0031] Σε μια άλλη υλοποίηση, το σύστημα επεξεργάζεται ανεπεξέργαστα μη δομημένα ορόσημα που προέρχονται από μετρήσεις θέσης των δεικτών και εκτιμά τα αντίστοιχα ορόσημα στην επιφάνεια του σώματος, ουσιαστικά, συμπληρώνοντας και βελτιώνοντας τις εκτιμήσεις των καταγεγραμμένων δεικτών για να παράγει εξόδους που δεν μπορούν εύκολα να μετρηθούν με την απαραίτητη ακρίβεια.

[0032] Σε μια περαιτέρω υλοποίηση, επιλύει επιπροσθέτως μη παρατηρούμενα ορόσημα, όπως εσωτερικά ορόσημα (π.χ. ανθρώπινες αρθρώσεις) ή ορόσημα στην επιφάνεια του σώματος, χρησιμοποιώντας μόνο μετρήσεις θέσης εξωθημένων δεικτών ως είσοδο.

[0033] Σε μια προτιμώμενη υλοποίηση, εκτιμά επίσης την πόζα και το σχήμα του υποκειμένου, προσαρμόζοντάς τα στα εκτιμώμενα ορόσημα μέσω βελτιστοποίησης.

[0034] Σε μια περαιτέρω προτιμώμενη υλοποίηση, το σχήμα του υποκειμένου βαθμονομείται αρχικά, με την προσαρμογή του παραμετρικού μοντέλου σώματος στα εκτιμώμενα ορόσημα μέσω μιας διαδικασίας βελτιστοποίησης. Το βαθμονομημένο σχήμα στη συνέχεια μένει σταθερό και χρησιμοποιείται κατά την επίλυση της πόζας του υποκειμένου για την παραγωγή της ψηφιακής αναπαράστασης 4Δ της καταγεγραμμένης

ερμηνείας.

[0035] Η αποκαλυπτόμενη εφεύρεση χρησιμοποιεί ένα μοντέλο μηχανικής μάθησης με τη μορφή νευρωνικού δικτύου για την επεξεργασία της μη δομημένης εισόδου θέσεων δεικτών και μπορεί να εκτιμήσει αποκρυπτόμενους δείκτες, να αφαιρέσει λανθασμένες εκτιμήσεις που δεν αντικατοπτρίζουν τους πραγματικούς δείκτες που έχουν τοποθετηθεί στο υποκείμενο (δείκτες-φαντάσματα) και να αποθορυβήσει τους δείκτες εισόδου (εκλεπτυσμένη τοποθέτηση). Αυτό καθιστά το σύστημα ανθεκτικό στις ατέλειες των σύγχρονων συστημάτων MoCap.

[0036] Ένα άλλο πλεονεκτικό χαρακτηριστικό αυτής της εφεύρεσης είναι ότι το ίδιο μοντέλο μηχανικής μάθησης νευρωνικού δικτύου μπορεί να εκπαιδευτεί για την επίλυση εσωτερικών ορόσημων που αντιπροσωπεύουν την εσωτερική (σκελετική) δομή του αντικειμένου ή του υποκειμένου που καταγράφεται.

[0037] Ένα ακόμη πλεονεκτικό χαρακτηριστικό της παρούσας εφεύρεσης είναι ότι το ίδιο μοντέλο μηχανικής μάθησης νευρωνικού δικτύου μπορεί να εκπαιδευτεί για να λύσει τα ορόσημα στην επιφάνεια του σώματος που αντιπροσωπεύουν την επιφάνεια του αντικειμένου ή του υποκειμένου που καταγράφεται.

[0038] Σε μια πιο πλεονεκτική υλοποίηση, η παρούσα εφεύρεση χρησιμοποιεί ένα μοντέλο μηχανικής μάθησης νευρωνικού δικτύου το οποίο εκπαιδεύεται για την επίλυση των ορόσημων στην επιφάνεια του σώματος και των ορόσημων της επιφάνειας του σώματος που αντιπροσωπεύουν ταυτόχρονα τη σκελετική δομή και την επιφάνεια του σώματος του αντικειμένου ή του υποκειμένου που καταγράφεται.

[0039] Η εφεύρεση αποκαλύπτει μια μέθοδο για την εκπαίδευση του μοντέλου μηχανικής μάθησης που χρησιμοποιείται από το σύστημα χρησιμοποιώντας δεδομένα από διάφορες πηγές. Αυτό είναι ένα πλεονέκτημα της εφεύρεσης, καθώς το σύστημα μπορεί να τροφοδοτείται με δεδομένα που αποκτώνται από πιο πρακτικές και προσιτές λύσεις χωρίς να θυσιάζεται η ποιότητα ή η απόδοση.

[0040] Σε μία υλοποίηση το σύστημα μπορεί να χρησιμοποιεί δεδομένα από σύγ-

χρονα συστήματα MoCap.

[0041] Σε μια προτιμώμενη ενσωμάτωση, το σύστημα μπορεί να χρησιμοποιεί δεδομένα που αποκτήθηκαν από (μια) κάμερα(ες) χαμηλού κόστους με χρήση τεχνολογίας MoCap χωρίς δείκτες.

[0042] Σε μια περαιτέρω προτιμώμενη υλοποίηση, το σύστημα μπορεί να χρησιμοποιήσει δεδομένα που αποκτήθηκαν από διαφορετικές πηγές, συμπεριλαμβανομένων των σύγχρονων συστημάτων MoCap και των προσιτών συστημάτων χωρίς δείκτες. Αυτό καθιστά το σύστημα επεκτάσιμο σε σχέση με τα δεδομένα που θα χρησιμοποιούνται για τη συνεχή βελτίωσή του. Κάνει επίσης το σύστημα ευέλικτο, καθώς μπορεί να εξειδικευτεί σε νεοαποκτηθέντα υποσύνολα δεδομένων που συλλέγονται από διαφορετικά συστήματα από τα δεδομένα που είχαν συλλεχθεί προηγουμένως.

[0043] Ένα ακόμα πλεονέκτημα που αποκαλύπτεται στην παρούσα εφεύρεση είναι μια μέθοδος για την καλύτερη αξιοποίηση των διαθέσιμων δεδομένων για τη βελτίωση της εκπαίδευσης του μοντέλου μηχανικής μάθησης με τη μείωση του συστηματικού σφάλματος (τάση/πόλωση) των συλλεχθέντων δεδομένων. Η μέθοδος προσαρμόζει την κατανομή των δεδομένων εκπαίδευσης για να μειώσει την επίδραση της μακρόστενης φύσης της, καθώς και να την καταστήσει ανθεκτική στη σημαντική μεροληψία που απεικονίζεται στις ληφθείσες πόζες. Πλεονεκτικά αυτό επιτυγχάνεται χωρίς πρόσθετα δεδομένα και αυτόματα.

[0044] Ακόμα ένα πλεονέκτημα που αποκαλύπτεται στην παρούσα εφεύρεση είναι μια μέθοδος που χρησιμοποιείται από το σύστημα για την προσαρμογή του σχήματος του σώματος και της πόζας σε θορυβώδεις εκτιμήσεις, μια κατάσταση που απουσιάζει από τη βιβλιογραφία, καθώς αυτή βασιζόταν συμβατικά σε εκτιμήσεις υψηλής ποιότητας από ακριβά συστήματα υψηλής τεχνολογίας. Αυτό καθιστά την παρούσα εφεύρεση προσαρμόσιμη σε διάφορες επιλογές αισθητήρων και, το σημαντικότερο, σε πιο προσιτές επιλογές.

[0045] Μια περαιτέρω προτιμώμενη υλοποίηση περιλαμβάνει το εύρωστο εκπαιδευ-

μένο νευρωνικό δίκτυο για την εκτίμηση των ορόσημων και τη διαδικασία προσαρμογής με επίγνωση του θορύβου για την παραγωγή υψηλής ποιότητας εκτιμήσεων ορόσημου, πόζας και σχήματος, ανακατασκευάζοντας μια τετραδιάστατη προσωποποιημένη ψηφιακή αναπαράσταση της καταγεγραμμένης ερμηνείας.

[0046] Πλεονεκτικά, η παρούσα εφεύρεση βασίζεται σε παραμέτρους ενός μοντέλου σώματος. Αυτό επιτρέπει την απόκτηση δεδομένων ανθρώπινης κίνησης με τη χρήση τεχνολογίας σύλληψης χωρίς δείκτες, επιλύοντας έτσι το πρόβλημα της αρχικής εκκίνησης της υλοποίησης και παράδοσης του εν λόγω συστήματος. Κάποιος θα χρειαζόταν ένα συστήμα καταγραφής κίνησης για την απόκτηση δεδομένων εκπαίδευσης των μοντέλων, εμποδίζοντας την επεκτασιμότητα, αυξάνοντας το κόστος αλλά και το χρόνο υλοποίησης, και απαιτώντας πολύπλοκες διαδικασίες για την τυποποίηση των δεδομένων. Με την καταγραφή χωρίς δείκτες και την προσαρμογή των παραμέτρων του μοντέλου σώματος σε ορόσημα από δεδομένα εικόνας, η απόκτηση δεδομένων απλοποιείται σημαντικά, είναι κλιμακούμενη και προεπιλεγμένη σε μια τυποποιημένη αναπαράσταση, μειώνοντας την προσπάθεια μεταγενέστερης επεξεργασίας.

[0047] Ακόμα ένα πλεονέκτημα της παρούσας εφεύρεσης είναι ότι εφαρμόζεται σε συστήματα που παρέχουν ελάχιστα έτοιμα οπτικά δεδομένα και λειτουργούν σε περιβάλλοντα με ελάχιστους περιορισμούς.

[0048] Ακόμα ένα πλεονέκτημα της αποκαλυπτόμενης εφεύρεσης είναι η μέθοδος που μοντελοποιεί ρητά την αβεβαιότητα της μέτρησης και/ή της συναγόμενης αβεβαιότητας κατά την προσαρμογή των παραμέτρων του ανθρώπινου σώματος σε τρισδιάστατα ορόσημα, επιτρέποντας αυξημένη ανθεκτικότητα στο θόρυβο των ορόσημων εισόδου.

[0049] Ένα άλλο πλεονεκτικό χαρακτηριστικό της αποκαλυπτόμενης εφεύρεσης είναι ο τρόπος με τον οποίο το νευρωνικό δίκτυο σε συνδυασμό με το παραμετρικό μοντέλο σώματος τυποποιεί την αναπαράσταση των ορόσημων και, κατ' επέκταση, τα δεδομένα εξόδου του MoCap.

[0050] Ένα τελευταίο πλεονέκτημα της εφεύρεσης είναι ότι εκμεταλλεύεται την τυποποιημένη έξοδο για να εκτελέσει μια διαδικασία γεννητικής βελτίωσης μετά την καταγραφή, η οποία χρησιμοποιεί παρελθοντικές και μελλοντικές πληροφορίες για να διορθώσει αποτελεσματικά τις λανθασμένες εκτιμήσεις και να παράγει ομαλή κίνηση, ακόμη και σε ιδιαίτερα απαιτητικά περιβάλλοντα ανάπτυξης.

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ ΤΩΝ ΣΧΕΔΙΩΝ

[0051] Το ΣΧΗΜΑ 1 περιέχει την απεικόνιση ενός συστήματος για την καταγραφή κίνησης με χρήση τεχνολογιών βασισμένων σε δεδομένα που επιλύει θορυβώδη ορόσημα χρησιμοποιώντας ένα μοντέλο μηχανικής μάθησης, οπτικούς αισθητήρες εικόνας, ένα μοντέλο σώματος χαμηλής διάστασης και μια μέθοδο προσαρμογής με επίγνωση του θορύβου.

[0052] Το ΣΧΗΜΑ 2 απεικονίζει μια διαδικασία εκπαίδευσης του νευρωνικού δικτύου που παρουσιάζεται ως ένα τμήμα στο σύστημα του ΣΧΗΜΑΤΟΣ 1.

[0053] Το ΣΧΗΜΑ 3 απεικονίζει μια διαδικασία καταγραφής, επεξεργασίας και αποθήκευσης πραγματικών δεδομένων για την εκπαίδευση του νευρωνικού δικτύου που παρουσιάζεται ως τμήμα στο σύστημα του ΣΧΗΜΑΤΟΣ 1. Τα πραγματικά δεδομένα χρησιμοποιούνται ως είσοδος στη διαδικασία εκπαίδευσης του ΣΧΗΜΑΤΟΣ 2.

[0054] Το ΣΧΗΜΑ 4 είναι μια σχηματική αναπαράσταση του αυτοκωδικοποιητικού γεννητικού μοντέλου που εκπαιδεύεται για τη δημιουργία συνθετικών δεδομένων. Τα συνθετικά δεδομένα χρησιμοποιούνται ως είσοδος στη διαδικασία εκπαίδευσης του ΣΧΗΜΑΤΟΣ 2.

[0055] Το ΣΧΗΜΑ 5 απεικονίζει μια διαδικασία επιλογής λανθανόντων δειγμάτων από έναν τοπολογικό χώρο, τα οποία μετασχηματίζονται ώστε να καταλήξουν σε νέα λανθάνοντα δείγματα στο ίδιο πολύπτυχο μόρφωμα, ανακατασκευάζονται με τη χρήση του αποκωδικοποιητή του γεννητικού μοντέλου και μετέπειτα τα ανακατασκευασμένα

δείγματα αποθηκεύονται στη βάση δεδομένων συνθετικών δεδομένων.

[0056] Το ΣΧΗΜΑ 6 απεικονίζει την εφαρμογή μιας εκθετικής συνάρτησης σημασίας που στηρίζεται στο σφάλμα ανακατασκευής. Οι τιμές σφάλματος χρωματίζονται με τη χρήση δυαδικού χρωματικού χάρτη, όπου το μαύρο χρώμα υποδηλώνει το μέγιστο σφάλμα, ενώ τα δεδομένα εισόδου απεικονίζονται με διαφανές γκρι χρώμα για λόγους σύγκρισης..

[0057] Το ΣΧΗΜΑ 7 απεικονίζει τη διαδικασία βελτίωσης μιας καταγραφής που παράγεται από το σύστημα καταγραφής κίνησης σε πραγματικό χρόνο με τη χρήση ενός γεννητικού μοντέλου για την από κοινού βελτιστοποίηση μιας δέσμης καρέ.

[0058] Το ΣΧΗΜΑ 8 είναι ένα μπλοκ διάγραμμα ενός υποδειγματικού υπολογιστικού περιβάλλοντος εντός του οποίου μπορούν να υλοποιηθούν διάφορες υλοποιήσεις της εφεύρεσης και στο οποίο μπορούν να χρησιμοποιηθούν διάφορες υλοποιήσεις της εφεύρεσης.

ΛΕΠΤΟΜΕΡΗΣ ΠΕΡΙΓΡΑΦΗ ΤΗΣ ΕΦΕΥΡΕΣΗΣ ΚΑΙ ΤΩΝ ΣΧΕΔΙΩΝ

Ακολουθεί περιγραφή με αναφορά στα επισυναπτόμενα σχήματα 1-8. Θα πρέπει να γίνει κατανοητό ότι τα εν λόγω σχήματα έχουν υποδειγματικό χαρακτήρα και σε καμία περίπτωση δεν χρησιμεύουν για τον περιορισμό του πεδίου εφαρμογής της εφεύρεσης.

[0059] Η αποκαλυπτόμενη εφεύρεση χρησιμοποιεί ένα παραμετρικό μοντέλο $\mathcal{B}$, του οποίου οι παράμετροι ανακτώνται από μια βάση δεδομένων **108**, η οποία κωδικοποιεί από κοινού μια αναπαράσταση των αρθρώσεων και του σχήματος του σώματος ως την MoCap αναπαράστασή του. Μετά την ανακατασκευή του σχήματος του μοντέλου, ένα σύνολο προκαθορισμένων αρθρώσεων στο εσωτερικό του σχήματος χρησιμοποιείται μαζί με ένα σύνολο παραμέτρων άρθρωσης για την παραμόρφωση του σχήματος και την αλλάγή της πόζας χρησιμοποιώντας μια μέθοδο τοποθέτησης πλέγματος μέσω

κινηματικής αρθρώσεων (Skinning). Συνήθως, οι ατέλειες του skinning λαμβάνονται υπόψη μετά την παραμόρφωση με τη χρήση μετατοπίσεων που υπολογίζονται από τις διαθέσιμες παραμέτρους (σχήματος και πόζας). Υπάρχουν διαφορετικές παραλλαγές, όλες βασισμένες σε δεδομένα, άλλες βασίζονται σε στοχαστικές αναπαραστάσεις [*Xu et al., 2020*], άλλες σε σαφείς [*Loper et al., 2015*], [*Osman et al., 2020*], με μια αξιοσημείωτη εξαίρεση που χρησιμοποιεί μια παραλλαγή κατασκευασμένη από καλλιτέχνες [*Yan et al., 2021*] και όλες συνήθως χρησιμοποιούν γραμμικούς διορθωτικούς παράγοντες για τις ατέλειες του skinning που στηρίζονται στην πληροφορία της πόζας [*Loper et al., 2015*], [*Xu et al., 2020*]. Η παρούσα εφεύρεση δεν δεσμεύεται από μια συγκεκριμένη υλοποίηση του $\mathcal{B}$ και μπορεί να υλοποιηθεί με οποιαδήποτε από τις διαθέσιμες παραλλαγές, αρκεί η χαμηλής διάστασης αναπαράσταση να μπορεί να μετατραπεί σε μια τρισδιάστατη αναπαράσταση που ορίζει την επιφάνεια του σώματος (π.χ. πλέγμα, νέφος σημείων). Καθώς οι διορθωτικές παραμορφώσεις της πόζας μπορούν να θεωρηθούν προαιρετικές, η παρούσα εφεύρεση μπορεί να υλοποιηθεί με μία αναπαράσταση συντελεστών blendshape για το σχήμα, σχεδιασμένων από έναν καλλιτέχνη, και ένα συσχετισμένο κινηματικό δέντρο - που προκύπτει είτε χειροκίνητα είτε αυτόματα [*Baran et al., 2007*] - συνοδευόμενο από βάρη δι-αρμονικού Skinning [*Jacobson et al., 2011*].

[0060] Στην παρούσα εφεύρεση, το $\mathcal{B}$ ορίζεται γενικά ως μια συνάρτηση $(\mathbf{v},\mathbf{f}) = \mathcal{B}(\beta,\theta,\mathbf{T})$, όπου $(\mathbf{v},\mathbf{f})$ είναι οι κορυφές $\mathbf{v} \in \mathbb{R}^{V\times 3}$ και οι πλευρές $\mathbf{f} \in \mathbb{N}^{F\times 3}$ μιας τριγωνικής επιφάνειας πλέγματος που ορίζεται από S συντελεστές blendshape $\beta \in \mathbb{R}^S$, τοποθετείται/ποζάρεται από P παραμέτρους πόζας $\theta \in \mathbb{SO}(3)^P$ και τοποθετείται στο χώρο από τον μετασχηματισμό $\mathbf{T} \in \mathbb{SE}(3)$.

[0061] Μετά την ανακατασκευή ενός τοποθετημένου και αρθρωμένου σχήματος, που αναπαρίσταται ως πλέγμα, είναι δυνατόν να χρησιμοποιηθούν γραμμικές συναρτήσεις r εκφρασμένες ως πίνακες $\mathcal{R}$ για την εξαγωγή των ορόσημων του σώματος $\ell := r(\mathbf{v}) = \mathcal{R} \times \mathbf{v}$, με $\ell \in \mathbb{R}^{L\times 3}$ και $\mathcal{R} \in \mathbb{R}^{L\times V}$. Με αυτόν τον τρόπο, τα σημεία της επιφάνειας του ℓ^v μπορούν να εξαχθούν χρησιμοποιώντας συναρτήσεις δέλτα (επιλογή κορυφών) ή

βαρυκεντρικές (παρεμβολή πλευράς τριγώνου), και οι αρθρώσεις ℓ^j χρησιμοποιώντας συναρτήσεις σταθμισμένου μέσου όρου. Δεδομένου ότι οι δείκτες εξέχουν του σώματος μίας απόστασης d, αντιστοιχούν σε $\ell^m = \ell^v + d(\mathcal{R} \times \mathbf{n})$, με $\mathbf{n}$ τον προσανατολισμό της επιφάνειας των κορυφών (δηλαδή τη διεύθυνση του ορθοκανονικού διανύσματος).

[0062] Αν και η ακόλουθη περιγραφή και οι εικόνες θα επικεντρωθούν σε παραμετρικά μοντέλα ανθρώπινου σώματος, ένας ειδικός καταλαβαίνει ότι το σύστημα και οι μέθοδοι της παρούσας εφεύρεσης μπορούν εφαρμόζονται σε οποιοδήποτε παραμετρικό ή αρθρωτό μοντέλο που αναπαριστά, για παράδειγμα, ζώα [*Zuffi et al., 2017*], ή αρθρωτά αντικείμενα [*Xue et al., 2021*].

[0063] Το ΣΧΗΜΑ 1 απεικονίζει ένα σύστημα μηχανικής μάθησης **100** που χρησιμοποιεί αισθητήρες **101** με βάση την εικόνα για την απόκτηση ακατέργαστων μη δομημένων θέσεων δεικτών **102** ως είσοδο σε ένα νευρωνικό δίκτυο **103** και εκτιμά τις παραμέτρους που σχετίζονται με την παραμετρική αναπαράσταση του μοντέλου σώματος **104**, είτε πρόκειται για σαφή **104** είτε για έμμεση **106**. Η σαφής αναπαράσταση αναφέρεται στο ορόσημο **104** ή στην αναπαράσταση με βάση το πλέγμα που περιλαμβάνει κορυφές **v** και **f**, ή αναπαραστάσεις που εξάγονται από την αναπαράσταση με βάση το πλέγμα, όπως ορόσημα δείκτη και κοινά ορόσημα **104**, ενώ η έμμεση αναπαράσταση αναφέρεται στην παραμετρική αναπαράσταση **106** που περιλαμβάνει τις παραμέτρους σχήματος β, πόζας θ και χωρικού μετασχηματισμού **T**.

[0064] Ενώ σε αυτή τη συγκεκριμένη απεικόνιση **100** η ενδιάμεση αναπαράσταση που εξάγεται από το νευρωνικό δίκτυο **103** είναι σαφής (π.χ. ορόσημα **104**) και η τελική βελτιστοποιημένη αναπαράσταση είναι έμμεση (π.χ. παράμετροι του μοντέλου σώματος **106**), αυτές μπορούν να εναλλάσσονται. Ένα άλλο παράδειγμα θα ήταν η έξοδος μιας έμμεσης αναπαράστασης από το νευρωνικό δίκτυο **103**, η βελτιστοποίησή της για την αύξηση της ακρίβειάς της, και στη συνέχεια η εξαγωγή της σαφής αναπαράστασης από τη βελτιωμένη έμμεση. Γενικά, η εφεύρεση αναφέρεται σε εκτιμήσεις εισόδου μη δομημένων ορόσημων **102** που προέρχονται από δεδομένα εικόνας/οπτικής

ανίχνευσης **101**, σε ένα μοντέλο μηχανικής μάθησης **103** που τα λαμβάνει ως είσοδο και του οποίου οι παράμετροι επιλέγονται από μια βάση δεδομένων παραμέτρων **107**, σε ένα παραμετρικό μοντέλο σώματος ενός αρθρωμένου αντικειμένου/υποκειμένου του οποίου οι παράμετροι επιλέγονται από μια βάση δεδομένων παραμέτρων **108**, και σε μια διαδικασία βελτιστοποίησης **105** που βελτιώνει την εκτιμώμενη αναπαράσταση του μοντέλου σώματος που προέρχεται από το νευρωνικό δίκτυο, ώστε να παράγει βελτιωμένες προσαρμοσμένες παραμέτρους που περιγράφουν καλύτερα τις παρατηρήσεις. Αυτές διοχετεύονται σε ένα σύστημα αποθήκευσης δεδομένων **111** (π.χ. στη μνήμη, σε μια βάση δεδομένων ή σε ένα αρχείο).

[0065] Επιπλέον, η διαδικασία προσαρμογής **105** μπορεί να επωφεληθεί από ένα προ-βαθμονομημένο σχήμα β **109**, καθώς η βελτιστοποίηση **105** θα πρέπει να επιλύει μόνο τις παραμέτρους της πόζας θ^t και τον μετασχηματισμό $\mathbf{T}^t$ **106** για κάθε χρονική στιγμή t, σε σύγκριση με την επίλυση και για το σχήμα, δεδομένου ότι το σχήμα είναι σε μεγάλο βαθμό σταθερό κατά τη διάρκεια μιας ερμηνείας.

[0066] Περαιτέρω, η διαδικασία προσαρμογής **105** μπορεί να εκμεταλλευτεί υψηλότερου επιπέδου πρότερη γνώση για την πόζα, όπως εκφράζεται από την προγενέστερη γνώση που κωδικοποιείται από τα γεννητικά μοντέλα **110**, τα οποία κωδικοποιούν υψηλών διαστάσεων κατανομές. Πρόκειται για ένα βήμα πέρα από τους σταθερούς διακριτούς περιορισμούς (π.χ. κοινά όρια) προς πιο ομαλούς περιορισμούς κατανομής (π.χ. η κανονική κατανομή που κωδικοποιείται από μεταβλητούς αυτοκωδικοποιητές ή μοντέλα κανονικοποιημένων ροών).

[0067] Οι αποθηκευμένες καταγεγραμμένες ερμηνείες **111** μπορούν να βελτιωθούν **600** μέσω της χρήσης γεννητικών μοντέλων **110**, αξιοποιώντας την πληρότητα της χρονικής πληροφορίας, παράγοντας κινήσεις ρεαλιστικές, ομαλές και χωρίς ατέλειες.

[0068] Το ΣΧΗΜΑ 2 απεικονίζει τη διαδικασία εκπαίδευσης του νευρωνικού δικτύου **103** που εκτιμά τις παραμέτρους του και τις προσθέτει στη βάση δεδομένων **107**. Το νευρωνικό δίκτυο **103** που χρησιμοποιείται στην παρούσα εφεύρεση λαμβάνει

ως δεδομένα εισόδου εκπαίδευσης τις παραμέτρους ενός μοντέλου σώματος $\mathcal{B}$. Αυτές μπορεί να είναι είτε παράμετροι που λαμβάνονται από καταγραφές του πραγματικού κόσμου **201** ή από σύνθεση (δηλ. παραγμένες από υπολογιστή) **202**. Είτε πρόκειται για λήψεις από τον πραγματικό κόσμο είτε για συνθετικές παραμέτρους, η αναπαράσταση που προκύπτει έχει συνθετικό χαρακτήρα και, ως εκ τούτου, μπορεί να ενισχυθεί **203** και/ή να αλλοιωθεί **205** με τεχνιτές ατέλειες και θόρυβο, διαταράσσοντας ουσιαστικά τα δεδομένα εκπαίδευσης ώστε να αυξηθεί η ποικιλομορφία τους και να προσομοιωθούν ρεαλιστικές συνθήκες καταγραφής. Η επαύξηση ή/και η αλλοίωση των δεδομένων μπορεί να λάβει χώρα και στις δύο αναπαραστάσεις του μοντέλου σώματος $\mathcal{B}$, σαφείς $(\mathbf{v}, \mathbf{f})$ ή έμμεσες $(\beta, \theta, \mathbf{T})$.

[0069] Στο ΣΧΗΜΑ 2 παρουσιάζεται η διαδικασία εκπαίδευσης που αρχίζει με τη φόρτωση δειγμάτων από μια βάση δεδομένων με παραμέτρους του μοντέλου σώματος, τις επαυξάνει ή/και τις αλλοιώνει, εξάγει την αναπαράσταση που θα χρησιμοποιηθεί για να τροφοδοτήσει το νευρωνικό δίκτυο ως δείγματα εκπαίδευσης **204**, εκτελεί μια στοχαστική επανάληψη βελτιστοποίησης για να βελτιστοποιήσει τις παραμέτρους του και επαναλαμβάνει αυτή τη διαδικασία για έναν προκαθορισμένο αριθμό επαναλήψεων ή έως ότου κριθεί σκόπιμο να σταματήσει **206**. Οι παράμετροι του νευρωνικού δικτύου **103** που προκύπτουν αποθηκεύονται στη συνέχεια σε μια βάση δεδομένων παραμέτρων του μοντέλου **107** μαζί με κατάλληλα μεταδεδομένα που περιγράφουν το μοντέλο και τις λεπτομέρειες της εκπαίδευσης. Τα τελευταία περιλαμβάνουν την εξαγόμενη αναπαράσταση **204** που επιλέχθηκε ως είσοδος στο νευρωνικό δίκτυο.

[0070] Σε αυτή τη συγκεκριμένη απεικόνιση, εξετάζεται ένα σύνολο από ορόσημα της επιφάνειας του σώματος, τα οποία εξέχουν κατά μια προκαθορισμένη απόσταση d, για να προσομοιωθούν οι θέσεις των δεικτών. Ωστόσο, δεν υπάρχει συγκεκριμένος περιορισμός σε αυτή την αναπαράσταση, καθώς μπορεί να συνοδεύεται από οποιονδήποτε άλλο μετασχηματισμό που δρα στη σαφή ή έμμεση αναπαράσταση του σώματος. Ένα παράδειγμα θα ήταν μια διαφορετική παραμετροποίηση της περιστροφής των

παραμέτρων πόζας που χρησιμοποιούνται για τον καθορισμό της άρθρωσης του σώματος.

[0071] Εκτός από την περιγραφή του μοντέλου μηχανικής μάθησης **103** που θα ξεπεράσει τις προκλήσεις που σχετίζονται με την αδόμητη είσοδο των θέσεων των δεικτών και τα ελαττώματα της καταγραφής με βάση τους δείκτες (π.χ. ανταλλαγές δεικτών, αποκρύψεις, δείκτες-φαντάσματα), το αποκαλυπτόμενο σύστημα επωφελείται από μια μέθοδο εκπαίδευσης **200** που επιτρέπει ισχυρή απόδοση εκτιμήσεων, ξεπερνώντας την κατανομή μακράς ουράς των δεδομένων εκπαίδευσης, και επιπλέον μια μέθοδο προσαρμογής μοντέλου που είναι ανθεκτική στα ποικίλα επίπεδα θορύβου εισόδου **105**. Περαιτέρω, καθώς το σύστημα χρησιμοποιεί μια τυποποιημένη αναπαράσταση με τη μορφή του μοντέλου σώματος $\mathcal{B}$, επωφελείται από δεδομένα εκπαίδευσης που αποκτήθηκαν από πολλαπλές πηγές και διαφορετικές τεχνικές. Επομένως, εκτός από την απόκτηση δεδομένων με τη χρήση ακριβών συστημάτων MoCap, το νευρωνικό δίκτυο μπορεί να εκπαιδευτεί με δεδομένα **201**, **300**, που αποκτήθηκαν από προσιτά συστήματα **301** με χρήση τεχνολογίας MoCap χωρίς δείκτες **302**.

[0072] Συνολικά, η παρούσα εφεύρεση είναι ένα οικονομικό σύστημα MoCap, καθώς μπορεί να αναπτυχθεί με τη χρήση χαμηλότερης ποιότητας αλλά προσιτών σε θέμα κόστους αισθητήρων **101** και μπορεί να αξιοποιήσει δεδομένα που αποκτώνται από τεχνολογία MoCap χαμηλότερου κόστους **300**, **201** για τη συνεχή βελτίωση της απόδοσής του μέσω της εκπαίδευσης βελτιωμένων παραλλαγών νευρωνικών δικτύων **200**. Όλες οι μέθοδοι και τα σχήματα εκπαίδευσης νευρωνικών δικτύων μπορούν να υλοποιηθούν μέσω ενός υπολογιστή που διαθέτει επεξεργαστή και αποθηκευτικό χώρο, όπως περιγράφεται παρακάτω. Στο κείμενο που ακολουθεί θα περιγραφεί λεπτομερώς το σχήμα ανθεκτικής εκπαίδευσης του μοντέλου μηχανικής μάθησης του συστήματος και η προσέγγιση βελτιστοποίησης προσαρμογής του μοντέλου στα εκτιμώμενα ορόσημα.

Balanced Regression.

[0073] Οι συναρτήσεις συνάφειας είναι οι οδηγοί της παλινδρόμησης χρησιμό-

τητας [*Branco et al., 2017*], [*Torgo et al., 2015*], [*Torgo et al., 2007*], [*Torgo et al., 2013*] και καθοδηγούν επίσης την επιλογή/δημιουργία δειγμάτων εκ έπανα-/ύπερ-/δία-δείγματοληψιας. Αντί του ορισμού της συνάφειας ή της επιλογής δείγματος με βάση έναν σαφή τύπο ή ένα σύνολο κανόνων, η παρούσα εφεύρεση χρησιμοποιεί τη μάθηση αναπαράστασης για να τη μάθει απευθείας από τα δεδομένα.

[0074] Τα μοντέλα μηχανικής μάθησης **207** [*Kingma et al., 2014*], [*Rezende et al., 2015*] που προσανατολίζονται στη σύνθεση με αυτό-κωδικοποίηση μαθαίνουν από κοινού ένα μοντέλο ανακατασκευής **401** καθώς και έναν παραγωγικό δειγματολήπτη **202**:

$$\theta^{\ddagger} = \mathcal{G}(\mathbf{E}(\theta)), \qquad \theta^{\star} = \mathcal{G}(\mathcal{S}(\cdot)), \tag{1}$$

με ποικίλους περιορισμούς για την είσοδο θ και τους λανθάνοντες χώρους $\mathbf{z} = \mathbf{E}(\theta), \mathbf{z} \in \mathbb{R}^L$, ένας τυπικός περιορισμός είναι μια πρότερη κανονική κατανομή (Gaussian) που επιβάλλεται στον λανθάνοντα χώρο. Ένας κωδικοποιητής $\mathbf{E}(\theta)$ απεικονίζει την είσοδο θ σε έναν λανθάνοντα χώρο $\mathbf{z}$, ο οποίος ανακατασκευάζεται σε $\theta^{\ddagger}$ από μια γεννήτρια / αποκδωδικοποιητή $\mathcal{G}(\mathbf{z})$. Χρησιμοποιώντας μια συνάρτηση δειγματοληψίας $\mathcal{S}$ για τη δειγματοληψία του λανθάνοντος χώρου είναι επίσης δυνατό να παραχθούν νέα δείγματα εξόδου $\theta^{\star}$ **202**.

[0075] Η παρούσα εφεύρεση εκμεταλλεύεται την υβριδική φύση τέτοιων μοντέλων **207** ως λύση της μη ισορροπημένης παλινδρόμησης, η οποία ταυτόχρονα υπερδειγματοληπτεί την κατανομή στην ουρά **202** και ρυθμίζει τη βελτιστοποίηση των παραμέτρων του νευρωνικού με την εκ νέου στάθμιση των σπανιότερων δειγμάτων **208**.

Συνάρτηση σημασίας.

[0076] Τα μοντέλα αυτόκωδικοποίησης αναμένεται να αντικατοπτρίζουν την τάση / πόλωση των δεδομένων εκπαίδευσής τους, με τα εκ νέου περιττά/σπάνια δείγματα να είναι ευκολότερο/δυσκολότερο να ανακατασκευαστούν σωστά αντίστοιχα. Αυτή η μεροληψία στην ανακατασκευή μέσω αυτόκωδικοποίησης μπορεί να χρησιμοποιηθεί

για την απόδοση σημασίας σε κάθε δείγμα, καθώς εκείνα που είναι πιο δύσκολο να αναδομηθούν σωστά είναι πιο πιθανό να είναι δείγματα από τη μακρά ουρά της κατανομής. Ορίζεται μια συνάρτηση σημασίας ρ:

$$\rho(\theta) = exp(\frac{\varepsilon}{\sigma}), \tag{2}$$

χρησιμοποιώντας ένα σφάλμα ανακατασκευής ε:

$$\varepsilon = \sqrt{\frac{1}{J}\sum_{i=1}^{J} ||\bar{\ell^j}_i - \bar{\ell^j}_i^{\ddagger}||_2}, \tag{3}$$

με το ($^-$) να δηλώνει την κανονικοποίηση στη μονάδα χρησιμοποιώντας τη διαγώνιο του πλαισίου οριοθέτησης των αρθρώσεων εισόδου, το ε η κανονικοποιημένη ρίζα του μέσου τετραγωνικού σφάλματος στα ανακατασκευασμένα ορόσημα / αρθρώσεις με τις αρχικές τους θέσεις και σ έναν παράγοντα κλιμάκωσης που ελέγχει τη συνάφεια ρ.

[0077] Η χρήση θέσεων ορόσημων διατηρεί ερμηνεύσιμη σημασιολογία στα ρ και σ, καθώς είναι μονοκατευθυντικά εναλλάξιμες (γραμμική αντιστοίχιση) με την πόζα θ δεδομένου σταθερού σχήματος β. Στο ΣΧΗΜΑ 6 παρουσιάζονται υποδειγματικά ζεύγη εισόδου-ανακατασκευής όπως βαθμολογούνται από τη συνάρτηση σημασίας της εξίσωσης (2).

[0078] Η συνάρτηση σημασίας Εξ. (2) **208** χρησιμοποιείται ως βάρος για κάθε δείγμα κατά τη διάρκεια του υπολογισμού της επιβλεπόμενης απώλειας. Σε σύγκριση με την παραδοσιακή παλινδρόμηση με βάση τη χρησιμότητα [*Torgo et al., 2015*], [*Torgo et al., 2007*], [*Torgo et al., 2013*] των οποίων οι συναρτήσεις σημασίας υποβαθμίζουν τα κοινά δείγματα, η παρούσα εφεύρεση αντ' αυτού ενισχύει το σφάλμα - και κατ' επέκταση τις κλίσεις για την εκμάθηση - που προέρχονται από κάθε δείγμα που δεν ανακατασκευάζεται πιστά από το μοντέλο σύνθεσης αυτοκωδικοποίησης της Εξίσωσης (1) **207**.

Ελεγχόμενη σύνθεση.

[0079] Παρόλο που τα δείγματα της ουράς δεν ανακατασκευάζονται πιστά, η παραγωγική φύση των σύγχρονων μοντέλων σύνθεσης διαμορφώνει πολύπτυχα μορφώματα που αντιστοιχούν τις εισόδους στους υποκείμενους παράγοντες της μεταβολής των δεδομένων, αντιστοιχίζοντας αποτελεσματικά παρόμοιες πόζες σε κοντινούς λανθάνοντες κώδικες που μπορούν να διατρέξουν τις διαστάσεις του λανθάνοντα χώρου. Λαμβάνοντας υπόψη αυτό, η παρούσα εφεύρεση ορίζει ένα ελεγχόμενο σχήμα δειγματοληψίας για τη σύνθεση νέων δειγμάτων ουράς **500** (βλ. Σχ. 5).

[0080] Χρησιμοποιώντας τη συνάρτηση σημασίας από την εξίσωση (2) **208** τα δείγματα ουράς $\theta^{\dagger}$ μπορούν να εντοπιστούν αυτόματα μέσω στατιστικής κατωφλίωσης και να δημιουργήσουν ένα σύνολο σταθερών λανθανόντων κωδίκων $\mathcal{A} = \{ \mathbf{z}^{\dagger} \mid \mathbf{z}^{\dagger} = \mathbf{E}(\theta^{\dagger}) \}$. Στη συνέχεια χρησιμοποιείται η ακόλουθη συνάρτηση δειγματοληψίας:

$$\mathcal{S}_{i,j}(\cdot) = \varsigma(\mathcal{N}(\mathbf{a}_i, \mathbf{s}), \mathcal{N}(\mathbf{a}_j, \mathbf{s}), b), \quad \mathbf{a}_{i,j} \in_R \mathcal{A}. \tag{4}$$

[0081] Η εξίσωση (4) **501** παίρνει δείγματα από μια κανονική κατανομή με κέντρο δύο τυχαίους σταθερούς κώδικες i, j από το σύνολο $\mathcal{A}$ χρησιμοποιώντας μια τυπική απόκλιση $\mathbf{s}$ και τις αναμειγνύει **502** χρησιμοποιώντας σφαιρική γραμμική παρεμβολή ς με έναν ομοιόμορφα δειγματοληπτικό παράγοντα ανάμειξης $b \in \mathcal{U}(0, b^u)$. Η μη γραμμική παρεμβολή μεταξύ των δειγμάτων αποφεύγει τις νεκρές περιοχές του πολύπτυχου μορφώματος, καθώς δεν οδηγούν όλες οι κατευθύνσεις σε ρεαλιστικά δείγματα, και ως εκ τούτου αυξάνει την αληθοφάνεια των δημιουργηθέντων δειγμάτων. Αυτή είναι μια συγκεκριμένη υλοποίηση της χρήσης ενός γεννητικού μοντέλου αυτόκωδικοποίησης για την εξισορρόπηση της εκπαίδευσης του νευρωνικού δικτύου.

[0082] Πλεονεκτικά, μια διαδικασία που περιλαμβάνει μια επιλογή λανθανόντων δειγμάτων **501** από S δείγματα και στη συνέχεια μια συνάρτηση για το μετασχηματισμό

τους **502** αντιπροσωπεύει μια οικογένεια υλοποιήσεων της αποκαλυπτόμενης μεθόδου. Η συνάρτηση δειγματοληψίας μπορεί να είναι ένας τυχαίος δειγματολήπτης από μια κανονική κατανομή $\mathcal{N}(\mathbf{0},\mathbf{I}) \in \mathbb{R}^L$, ή ένας διακριτός δειγματολήπτης σταθερών κωδίκων, ή ο προαναφερθείς τυχαίος δειγματολήπτης γύρω από σταθερούς κώδικες, ενώ η συνάρτηση μετασχηματισμού μπορεί να είναι οποιαδήποτε συνάρτηση παρεμβολής ή ακόμη και η συνάρτηση ταυτότητας όταν $S = 1$. Γενικά, η συνάρτηση δειγματοληψίας **501** παράγει $S \geq 1$ δείγματα κωδίκων λανθάνοντος χώρου και η συνάρτηση μετασχηματισμού **502** παράγει ένα μόνο δείγμα από τα S δείγματα.

[0083] Αυτή η ελεγχόμενη διαδικασία δειγματοληψίας ουράς **500** απεικονίζεται σχηματικά στο ΣΧΗΜΑ 5 και ωφελεί την παρούσα εφεύρεση με την υπερδειγματοληψία της κατανομής δεδομένων εκπαίδευσης με δείγματα που συντίθενται. Αυτό παρουσιάζει πολλά πλεονεκτήματα, καθώς η ποικιλομορφία των παραγόμενων δειγμάτων είναι επωφελής για την εκπαίδευση του νευρωνικού δικτύου, αλλά και η δυνατότητα ελέγχου της αποσυμφόρησης της κατανομής των δειγμάτων δεδομένων καταγραφής κίνησης με μακρά ουρά.

[0084] Πλεονεκτικά, η παρούσα εφεύρεση έχει σχεδιαστεί για να λειτουργεί με οποιοδήποτε αυτοκωδικοποιούμενο γεννητικό μοντέλο **207**, ή ακόμη και με διαδοχικά αυτοκωδικοποιούμενα γεννητικά μοντέλα, ικανά να παράγουν και να ανακατασκευάζουν δεδομένα. Για την επικύρωση της μεθόδου μας, χρησιμοποιείται ένας μεταβλητός αυτοκωδικοποιητής (VAE), ο οποίος εκπαιδεύεται με την ακόλουθη επίβλεψη εκμάθησης:

$$\mathcal{L}_{VAE} = \lambda_1 \mathcal{L}_{KL} + \lambda_2 \mathcal{L}_{rec} + \lambda_3 \mathcal{L}_{orth} \tag{5}$$

$$\mathcal{L}_{KL} = \Psi(D_{KL}(q_\theta(z|\mathbf{R})||\mathcal{N}(0,I))) \tag{6}$$

$$\mathcal{L}_{rec} = (V - \hat{V})^2, \tag{7}$$

$$\mathcal{L}_{orth} = \frac{Trace(\mathbf{R}^T\hat{\mathbf{R}}) - 1}{2}, \tag{8}$$

όπου $z \in R^{32}$ είναι ο 32-διάστατος λανθάνων κώδικας, $\mathbf{R} \in SO(3)$ είναι ο πίνακας περιστροφής 3×3 για κάθε άρθρωση, ενώ $\hat{\mathbf{R}}$ είναι η έξοδος 3×3 του αποκωδικοποιητή. Τα $V, \hat{V}$ αντιστοιχούν στις προβλεπόμενες κορυφές και στις αντιστοιχιζόμενες κορυφές, υποδεικνύοντας ότι ο όρος ανακατασκευής ενσωματώνει τόσο τα γωνιακά όσο και τα σφάλματα θέσης. Η σχηματοποίηση του πολύπτυχου μόρφωματος κανονικοποιείται με τη χρήση της απόκλισης Kullback-Leibler (KL) με μια συνάρτηση σφάλματος Charbonnier Ψ, με $\Psi(x) = \sqrt{1+x^2} - 1$, για να αποφευχθεί η εκ των υστέρων κατάρρευση και να αποδομηθούν αποτελεσματικότερα οι παράγοντες διακύμανσης. Οι εξισώσεις (6,7) ακολουθούν το σχήμα εκπαίδευσης VAE - π.χ., διαπραγμάτευση της ποιότητας ανακατασκευής με την εκμάθηση ενός πολύπτυχου μόρφωματος που προσεγγίζει μια πολυδιάστατη κανονική κατανομή (Gaussian), ενώ οι εξισώσεις (7,8) αναγκάζουν το μοντέλο να κατασκευάσει έναν έγκυρο λανθάνοντα χώρο περιστροφών. Το σχήμα εκπαίδευσης VAE συμπληρώνεται από τη μέθοδο βελτιστοποίησης Adam με μείωση βάρους που τιμωρεί τα μεγάλα βάρη και αποτρέπει την υπερπροσαρμογή. Δύο μετρικές χρησιμοποιούνται για την επιλογή του μοντέλου με τις καλύτερες επιδόσεις για τη δημιουργία ουράς και την κανονικοποίηση της παλινδρόμησης, (i) την απόσταση Frechet inception για την ποσοτικοποίηση της ικανότητας του παραγωγικού μοντέλου να συνθέτει δεδομένα υψηλής ποιότητας και (ii) μια μετρική ποικιλομορφίας για την αξιολόγηση της ικανότητας του μοντέλου να παράγει ποικίλα δεδομένα. Το μοντέλο με την καλύτερη αντιστάθμιση ποικιλομορφίας-ποιότητας έχει επιλεγεί για **500** και **200**.

Table 1

	THuman2.0			
	RMSE ↓	PCK1(%) ↑	PCK3(%) ↑	PCK7(%) ↑
Τιμή βάσης	0.0214	28.69	92.08	98.60
BMSE	0.0220	25.51	91.90	98.62
Εξισορροπημένο	**0.0191**	**32.38**	**93.55**	**99.11**

[0085] Ο Πίνακας 1 απεικονίζει ένα παράδειγμα και συνοψίζει τα αποτελέσματα της παλινδρόμησης στο THuman2.0 σύνολο δεδομένων. Το σύνολο δεδομένων THuman2.0

είναι μια συλλογή διαφορετικών ποζών που έχουν ληφθεί από 5 άτομα. Ο Πίνακας 1 αναφέρει το μέσο τετραγωνικό σφάλμα (RMSE) που μετρά την ευκλείδεια απόσταση μεταξύ των προβλεπόμενων σημείων αναφοράς και των σημείων αναφοράς, και το ποσοστο σωστών σημείων αναφοράς (PCK) που μετρά το ποσοστό των προβλεπόμενων σημείων αναφοράς που βρίσκονται εντός της ακτίνας γ σε σχέση με τη θέσης αναφοράς: α) $\gamma = 0.01m$ που συμβολίζεται ως PCK1, β) $\gamma = 0.03m$ που συμβολίζεται ως PCK3 και γ) $\gamma = 0.07m$ που συμβολίζεται ως PCK7. Ο Πίνακας 1 παρουσιάζει μια σύγκριση επιδόσεων μεταξύ i) **103** εκπαιδεύμένο με μια συνάρτηση σφάλματος παλινδρόμησης και το οποίο συμβολίζεται ως "Τιμή βάσης", ii) **103** εκπαιδευμένο με τη συνάρτηση σφάλματος Balanced Mean Squared Error (BMSE) [*Ren et al., 2022*] και το οποίο συμβολίζεται ως "BMSE", και (iii) **103** που εκπαιδεύτηκε με σφάλμα παλινδρόμησης, δείγματα με χρήση **500**, κανονικοποίηση με **208**, και συμβολίζεται ως "Εξισορροπημένο". Το μοντέλο 'Εξισορροπημένο' αποδίδει καλύτερα από τα άλλα δύο σε όλες τις μετρικές, υποδεικνύοντας ότι ο συνδυασμός της ελεγχόμενης σύνθεσης και της κανονικοποίησης με μια συνάρτηση συνάφειας μπορεί να παλινδρομήσει ποικίλες πόζες από σύνολα δεδομένων που δεν παρουσιάζουν έντονη μεροληψία πόζας αλλά δεν περιλαμβάνουν αποκλειστικά σπάνιες πόζες.

Table 2

Προσαρμοσμένα σετ ουράς (μέσος όρος)				
	RMSE (↓)	PCK1 (%)	PCK3 (%)	PCK7 (%)
Τιμή βάσης	0.0358	22.04	80.27	94.31
BMSE	0.0329	**27.66**	81.98	94.92
Εξισορροπημένο	**0.0293**	23.42	**84.70**	**97.24**

[0086] Ο Πίνακας 2 απεικονίζει ένα άλλο παράδειγμα και συνοψίζει τα αποτελέσματα της παλινδρόμησης σε ένα προσαρμοσμένο σύνολο σπάνιων στάσεων που περιλαμβάνουν "ψηλές κλωτσιές", "σταυρωμένα πόδια", "σταυρωμένα χέρια" και "σκύψιμο". Τα αποτελέσματα είναι ο μέσος όρος των επιδόσεων του μοντέλου και στα 4 προσαρμοσμένα σύνολα ουρών. Ο Πίνακας 2 αναφέρει το RMSE και το PCK για τις

3 διαφορετικές τιμές ακτίνας όπως παρουσιάστηκαν παραπάνω. Ο Πίνακας 1 παρουσιάζει μια σύγκριση των επιδόσεων μεταξύ (i) **103** που εκπαιδεύτηκε με μια συνάρτηση σφάλματος παλινδρόμησης και συμβολίζεται ως "Τιμή βάσης", (ii) **103** που εκπαιδεύτηκε με τη συνάρτηση σφάλματος BMSE και συμβολίζεται ως "BMSE" και (iii) **103** που εκπαιδεύτηκε με μια συνάρτηση σφάλματος παλινδρόμησης, δείγματα με χρήση **500**, κανονικοποίηση με **208** και συμβολίζεται ως "Εξισορροπημένο". Το μοντέλο 'Εξισορροπημένο' έχει καλύτερες επιδόσεις από το 'Τιμή βάσης' σε όλες τις μετρικές, ενώ επιτυγχάνει καλύτερες επιδόσεις RMSE, PCK3 και PCK7 σε σύγκριση με το 'BMSE', γεγονός που υποδεικνύει ότι μαθαίνει να παλινδρομεί τη συνολική πόζα εύρωστα, αν και χάνει κάποια ακρίβεια σε συγκεκριμένα μέρη του σώματος.

[0087] Ένας ειδικευμένος επαγγελματίας θα αντιληφθεί ότι οποιοδήποτε νευρωνικό δίκτυο που ικανοποιεί τις ιδιότητες της αυτοκωδικοποίησης (δηλαδή της ανακατασκευής δειγμάτων δεδομένων) και της παραγωγής (δηλαδή της δειγματοληψίας δεδομένων) μπορεί να χρησιμοποιηθεί για την υλοποίηση του προαναφερθέντος σχήματος εκπαίδευσης ισορροπημένης παλινδρόμησης. Επιπροσθέτως, το εν λόγω πρόσωπο θα κατανοήσει επίσης ότι το αυτοκωδικοποιούμενο γεννητικό μοντέλο όταν χρησιμοποιείται για την εξισορρόπηση της εκπαίδευσης ενός νευρωνικού δικτύου παλινδρόμησης μπορεί να χρησιμοποιηθεί μόνο είτε για την ανάθεση σημασίας είτε για τη δειγματοληψία συνθετικών δεδομένων.

Εκτίμηση ορόσημου σε πραγματικό χρόνο.

[0088] Σε σύγκριση με τις καθαρές προσεγγίσεις επισήμανσης [*Ghorbani N. et al., 2021*], [*Black et al., 2023*], [*Ghorbani S. et al., 2019*] ή καθαρής επίλυσης [*Holden, 2018*], [*Chen et al., 2021*] το μοντέλο νευρωνικών δικτύων **103** της παρούσας εφεύρεσης ταυτόχρονα αποθορυβοποιεί, επιλύει και συμπληρώνει τα ορόσημα. Με τις ανεπεξέργαστες θέσεις των σημάτων ως είσοδο **102**, είναι διαθέσιμες διαφορετικές επιλογές σε σχέση με το νευρωνικό δίκτυο που θα χρησιμοποιηθεί. Τα νευρωνικά μοντέλα που βασίζονται σε μετασχηματιστές μπορούν να λειτουργήσουν καθαρά με την αδόμητη

ανεπέξεργαστη είσοδο θέσης δεικτών, ενώ τα συνελικτικά νευρωνικά δίκτυα πρέπει πρώτα να τη μετατρέψουν σε δομημένη αναπαράσταση. Το πλεονέκτημα των πρώτων είναι ο άπειρος βαθμός λεπτομέρειας εισόδου, ενώ ο αντίστοιχος περιορισμός που επιβάλλεται κατά τη μετατροπή της ανεπέξεργαστης πληροφορίας σε ένα δομημένο πλέγμα ελέγχεται από τη χρήση καλά αναπτυγμένων αναπαραστάσεων εξόδου, όπως οι θερμικοί χάρτες (heatmaps). Τα heatmaps ως αναπαράσταση επιτρέπουν πλουσιότερες τεχνικές εποπτείας και αποκωδικοποίησης συντεταγμένων από την εναλλακτική λύση της καθαρής παλινδρόμησης που προσφέρουν οι μετασχηματιστές. Ωστόσο, η παρούσα εφεύρεση μπορεί να επωφεληθεί και από τις δύο αρχιτεκτονικές και δεν περιορίζεται μόνο σε αυτές, καθώς το καθεστώς εκπαίδευσης **200** είναι ανεξάρτητο από την εκάστοτε αρχιτεκτονική νευρωνικού δικτύου.

[0089] Ο στόχος είναι να μετατραπούν οι μη δομημένες, θορυβώδεις και ενδεχομένως μη-/υπερ-πλήρεις θέσεις δεικτών σε προκαθορισμένα - κατά τη φάση εκπαίδευσης του μοντέλου - σημεία αναφοράς (ορόσημα) του σώματος. Η ελεγχόμενη δειγματοληψία **500** και η συνάρτησης σημασίας **208** εφαρμόζονται πριν και μετά το πέρασμα του νευρωνικού δικτύου αντίστοιχα, πράγμα που σημαίνει ότι οποιοδήποτε νευρωνικό δίκτυο **103**, ώς ένα μαύρου κουτί, που ικανοποιεί αυτούς τους χαλαρούς περιορισμούς - ακολουθούμενο από την απόδοση σε πραγματικό χρόνο εκτέλεσης - είναι ένας πιθανός υποψήφιος υλοποίησης.

[0090] Ενώ ορισμένες μέθοδοι χρησιμοποιούν τεχνικές παλινδρόμησης για την εκτίμηση επιφανειακών δεικτών και άλλων κοινών ορόσημων, η παρούσα εφεύρεση επωφελείται από την από κοινού παλινδρόμηση επιφανειακών (υποκορυφή), εξωτερικών (εξωθημένοι δείκτες) και εσωτερικών (αρθρώσεις) ορόσημων, καθώς αυτό ενισχύει την απόδοση και προσθέτει χρησιμότητα όσον αφορά την προσαρμογή του σώματος και την επίλυση των περιστροφών των αρθρώσεων.

[0091] Όπως φαίνεται στο ΣΧΗΜΑ 2, κατά τη διάρκεια της εκπαίδευσης οι μίνι-παρτίδες περιέχουν τυχαία συνθετικά δείγματα **202** που αποκτήθηκαν από την Εξ. (4)

χρησιμοποιώντας το μοντέλο που διατυπώνεται από την Εξ. (1) (δεξιά) **207** και κάθε δείγμα σταθμίζεται χρησιμοποιώντας την Εξ. (2) μετά την ανακατασκευή του μέσω της Εξ. (1) (αριστερά) **208**.

[0092] Πλεονεκτικά, η παρούσα εφεύρεση έχει σχεδιαστεί για να λειτουργεί με οποιοδήποτε νευρωνικό δίκτυο **103**, ή ακόμη και διαδοχικά νευρωνικά δίκτυα, ικανά να εκτιμούν δείκτες και αρθρώσεις από μια είσοδο νέφους σημείων των δεικτών. Για την επικύρωση του μοντέλου, χρησιμοποιήθηκε μία από τις πιο συχνά χρησιμοποιούμενες αρχιτεκτονικές που έχει αποδειχθεί ότι επιτυγχάνουν κορυφαία αποτελέσματα σε διάφορες εφαρμογές, και είναι επίσης πιο αποδοτική από άποψη επιδόσεων κατά τη διάρκεια εκτέλεσης και μας επιτρέπει την παλινδρόμηση heatmaps υψηλής ανάλυσης. Μια τροποποιημένη έκδοση της αρχιτεκτονικής του νευρωνικού δικτύου UNet [*Ronneberger et al., 2015*] υλοποιήθηκε για **103** για την ταυτόχρονη πρόβλεψη 53 δεικτών και 18 ορόσημων αρθρώσεων.

[0093] Το μοντέλο αποτελείται από πέντε μπλοκ, με κάθε μπλοκ να αποτελείται από 32, 64, 128, 256 και 512 χαρακτηριστικά αντίστοιχα. Κάθε μπλοκ κωδικοποιητή αποτελείται από δύο συνελικτικά μπλοκ, με μέγεθος πυρήνα 3, βήμα 1 και γέμισμα στα όρια 1, ακολουθούμενο από τις συναρτήσεις ενεργοποίησης ReLU και κανονικοποίησης παρτίδας, οι οποίες χρησιμοποιούνται συνήθως στα βαθιά νευρωνικά δίκτυα για την αντιμετώπιση των συνεχόμενα μειωμένων και εκτοδευόμενων κλίσεων. Η υποβιβαστική κλιμάκωση σε όλα τα μπλοκ επιτυγχάνεται μέσω του τεχνικών αντι-επικάλυψης και μέγιστης ομαδοποίησης [*Zhang et al., 2019*], το οποίο συμβάλλει στη μείωση της επικάλυψης και στην παραγωγή ομαλότερων αποτελεσμάτων. Το μπλοκ με τα περισσότερα χαρακτηριστικά (bottleneck) του μοντέλου αποτελείται από ένα μόνο συνελικτικό μπλοκ, το οποίο χρησιμοποιεί τις ίδιες παραμέτρους με τα μπλοκ του κωδικοποιητή. Ο αποκωδικοποιητής περιλαμβάνει τα ίδια συνελικτικά μπλοκ και η έξοδος κάθε μπλοκ συνδυάζεται με την έξοδο του αντίστοιχου κωδικοποιητή μετά από (δι-)γραμμική υπερδειγματοληψία. Τέλος, η τελικά πρόβλεψη αποτελείται από ένα συνελικτικό μπλοκ με

μέγεθος πυρήνα 1, βήμα 1 και γέμισμα 0, με συνάρτηση ενεργοποίησης ReLU. Αυτό το στρώμα παράγει την τελική έξοδο του μοντέλου, η οποία χρησιμοποιείται για την εκτίμηση των προβλεπόμενων ορόσημων.

[0094] Οι ακατέργαστες μη δομημένες θέσεις των δεικτών εισόδου $\tilde{\ell}_{gt}$ **102** κανονικοποιούνται στο $[0,1]$ και ψηφιοποιούνται με χρήση ορθογραφικής προβολής για την παραγωγή δύο ορθογώνιων κανονικοποιημένων χαρτών βάθους. Υποθέτοντας την εκ των προτέρων γνώση της κατεύθυνσης της βαρύτητας κατά μήκος του άξονα y, αυτές οι ορθογώνιες και ορθογραφικές προβολές που συμβολίζονται ως xy και yz μοιράζονται τον άξονα y. Στη συνέχεια, χρησιμοποιείται μια παραλλαγή της οριακής παλινδρόμησης μέσω heatmap [*Nibali et al., 2019*], [*Ye et al., 2022*]. Το μοντέλο εποπτεύεται χρησιμοποιώντας παλινδρόμηση κέντρου μάζας [*Nibali et al., 2018*], [*Sun et al., 2018*] για την εκτίμηση των κανονικοποιημένων θέσεων των ορόσημων $\tilde{\ell}_{est}$, λαμβάνοντας τη μέση προσδοκόμενη τιμή [*Nibali et al., 2019*], [*Ye et al., 2022*] για τον άξονα y.

[0095] Πλεονεκτικά, αντί να απαιτείται από το μοντέλο να προβλέψει την επιπλέον οπτική γωνία, όπως γίνεται για τις μονοφθαλμικές εικόνες, η τρισδιάστατη φύση των δεδομένων εισόδου σε συνδυασμό με το ορθογώνιο βάθος προβολής map rendering, επιτρέπει τη ρητή πρόβλεψη κάθε προβολής εισόδου που θα συγχωνευθεί στη συνέχεια, μειώνοντας το χάσμα πεδίου και μάθησης.

[0096] Το μοντέλο εποπτεύεται από:

$$\mathcal{L} = \lambda \mathcal{L}_{JS}(\mathbf{H}_{gt}, \mathbf{H}_{est}) + \mathcal{L}_{w}^{\nu}(\tilde{\ell}_{gt}, \tilde{\ell}_{est}), \tag{9}$$

όπου $\mathcal{L}_{JS}$ είναι η $\lambda-$σταθμισμένη απόκλιση Jensen-Shannon [*Menendez et al., 1997*] μεταξύ της κανονικοποιημένης τιμής αναφοράς και των soft-max κανονικοποιημένων προβλεπόμενων heatmap, ενώ $\mathcal{L}_{w}^{\nu}$ είναι η εύρωστη συνάρτηση σφάλματος Welsch [*Holland et al., 1977*], με την παράμετρο υποστήριξης ν, μεταξύ των ψηφιοποιημένων και κανονικοποιημένων συντεταγμένων ορόσημων $\tilde{\ell}$.

[0097] Συνολικά, το $\mathcal{L}_{JS}$ επιταχύνει την εκπαίδευση με μίνι παρτίδες **206**, ενώ το $\mathcal{L}_w^v$ διευκολύνει υψηλότερα επίπεδα ακρίβειας υπο-εικονοστοιχείων, παρόλο που τα heatmaps **H** ανακατασκευάζονται χρησιμοποιώντας τις κανονικοποιημένες - μη κβαντισμένες - συντεταγμένες [*Zhang et al., 2020*], οι ατέλεις/σφάλματα διακριτοποίησης δεν μπορούν ποτέ να αφαιρεθούν πλήρως. ιακριτοποίησης δεν μπορούν ποτέ να αφαιρεθούν πλήρως.

[0098] Για την εκπαίδευση του νευρωνικού δικτύου, η συνθετική φύση των παραμέτρων του μοντέλου σώματος χαμηλής διάστασης μας επιτρέπει να παράγουμε τα δεδομένα εκπαίδευσης ανέξοδα την ώρα της εκπαίδευσης. Αυτός η μέθοδος δημιουργίας δεδομένων στοχεύει στη στενή προσέγγιση των συνθηκών MoCap σε πραγματικό κόσμο, λαμβάνοντας υπόψη τις παραλλαγές του σχήματος του σώματος του υποκειμένου και τα ζητήματα που σχετίζονται με το σήμα εισόδου, όπως οι δείκτες-φαντάσματα, οι αποκρύψεις των δεικτών και ο θόρυβος των δεικτών. Ενώ στη συνέχεια θα περιγραφεί μία υποδειγματική μέθοδος, αυτή δεν περιορίζεται σε αυτές τις περιγραφόμενες λειτουργίες μόνο, καθώς είναι επιλέξιμες οποιεσδήποτε μετασχηματιστικές λειτουργίες για τη μείωση του χάσματος μεταξύ συνθετικών και πραγματικών δεδομένων και την παραγωγή σημαντικών παραλλαγών δεδομένων.

[0099] Αρχικά, οι μεταβολές των παραμέτρων του μοντέλου σώματος $\mathcal{B}$ **203** πραγματοποιούνται ξεκινώντας με μια τυχαία ελεγχόμενη μετατόπιση των συντελεστών σχήματος β:

$$\beta' = \beta + \mathcal{U}(-1,1), \tag{10}$$

χρησιμοποιώντας μια ομοιόμορφη κατανομή γύρω από την κανονική τυπική απόκλιση. Επιπλέον, ένα υποσύνολο των παραμέτρων σχήματος β αντικαθίσταται τυχαία:

$$\beta_i' = \begin{cases} \beta_i, & \text{if } i \notin S \\ \mathcal{N}(0,1), & \text{if } i \in S \end{cases} \tag{11}$$

όπου S είναι ένα σύνολο από n συντελεστές που λαμβάνονται ομοιόμορφα από το σύνολο των συντελεστών. Στη συνέχεια, εφαρμόζεται μια τυχαία αναστροφή των αριστερά και δεξιά συνιστοσών της πόζας για να ληφθεί υπόψη η επικρατούσα δεξιοχειρία.

[0100] Στη συνέχεια, εφαρμόζεται επίσης ένας αριθμός συνθετικών αλλοιώσεων **205**, με τα ακόλουθα να αποτελούν μη περιοριστικά, ενδεικτικά παραδείγματα. Για την προσομοίωση αποκρύψεων δεικτών, ακολουθούμε την παρακάτω διαδικασία. Στη συνέχεια, επιλέγουμε τυχαία ένα υποσύνολο δεικτών για απόκρυψη, προσδιορίζοντας τον αριθμό των δεικτών που πρόκειται να αποκρυφτούν, ο οποίος συμβολίζεται ως k. Λαμβάνουμε ένα τυχαίο δείγμα από μια διακριτή ομοιόμορφη κατανομή για να προσδιορίσουμε το k, $k \sim \mathcal{U}(m,n'), \qquad m \leq k \leq n' \leq n$, όπου $\mathcal{U}(m,n')$ είναι η ομοιόμορφη κατανομή στο εύρος των ακεραίων $\{m_1,m_2,\ldots,n'\}$, και n ορίζει τον μέγιστο αριθμό δεικτών που πρέπει να αποκρυβούν. Στη συνέχεια, λαμβάνουμε ένα άλλο τυχαίο δείγμα από μια ομοιόμορφη κατανομή για να καθορίσουμε τους δείκτες των δεικτών που θα αποκρυφτούν, δηλαδή, $\mathbf{m} = (m_1,m_2,\ldots,m_k) \sim \mathcal{U}(1,n), k \leq n$, όπου $\mathcal{U}(1,n)$ είναι η ομοιόμορφη κατανομή στο σύνολο των συντελεστών των δεικτών. Το διάνυσμα $\mathbf{m}$ που προκύπτει περιέχει τους συντελεστές των προς απόκρυψη δεικτών και χρησιμοποιείται για τον αποκλεισμό αυτών των δεικτών από το ℓ^m. Ως επόμενο βήμα, οι δείκτες-φαντάσματα προσομοιώνεται με την εξαγωγή δειγμάτων από μια κατανομή Gauss με τιμές μέσης τιμής και τυπικής απόκλισης ισοδύναμιες με τις αρχικές θέσεις των δεικτών. Αναλυτικότερα, υπολογίζεται πρώτα η διάμεση θέση για κάθε χωρική διάσταση των θέσεων των δεικτών μ_j, (δηλαδή η διάμεση τιμή για την j-οστή χωρική διάσταση των θέσεων των δεικτών), και ο πίνακας συνδιακύμανσης του δείγματος Σ. Στη συνέχεια, λαμβάνονται δείγματα $g \sim \mathcal{G}(\mu,\Sigma)$ και προσαρτώνται στις αρχικές θέσεις των δεικτών ℓ^m. Τέλος, για να προσομοιωθεί ο θόρυβος των δεικτών, επιλέγεται τυχαία ένα σύνολο δεικτών στους οποίους προστίθεται θόρυβος. Με τον αριθμό των δεικτών προς μετατόπιση και τη μέγιστη επιτρεπόμενη απόσταση μετατόπισης, παίρνουμε τυχαίο δείγμα από μια ομοιόμορφη κατανομή για να καθορίσουμε τους δείκτες των δεικτών στους οποί-

ους θα προστεθεί ο θόρυβος $I \sim \mathcal{U}(1,N)$. Για κάθε δείκτη $i_j \in I$, μια τυχαία μετατόπιση δημιουργείται το διάνυσμα $o \sim \mathcal{U}(-M,M)$ και προστίθεται αυτή η μετατόπιση στην αρχική θέση του δείκτη για να ληφθεί η θορυβώδης θέση $\tilde{\ell^m} = \ell^m + \mathbf{o}$. Ο προτεινόμενος αγωγός επαύξησης **203** και διαφθοράς **205** εφαρμόζεται τυχαία σε κάθε εποχή, με διαφορετικές πιθανότητες ενεργοποίησης κάθε συνάρτησης επαύξησης ή διαφθοράς για κάθε δείγμα μέσα σε μια μίνι-παρτίδα. Ελέγχοντας προσεκτικά τη δημιουργία συνθετικών δεδομένων MoCap, είναι δυνατή η δημιουργία ρεαλιστικών δεδομένων κατάλληλων για την εκπαίδευση αλγορίθμων μηχανικής μάθησης για την εκτέλεση MoCap. Στον Πίνακα 3 παρουσιάζονται τα ποσοτικά αποτελέσματα με τη χρήση τυποποιημένων μετρικών των διαφόρων εκπαιδευμένων παραλλαγών αυτού του μοντέλου. Κάθε σειρά αντιστοιχεί σε διαφορετικό σύνολο δεδομένων εκπαίδευσης και όλες οι δοκιμές πραγματοποιήθηκαν στο τμήμα ACCAD του συνόλου δεδομένων AMASS [*Mahmood et al., 2019*], το οποίο δεν χρησιμοποιήθηκε κατά την εκπαίδευση και αντιστοιχεί σε ένα αμερόληπτο σύνολο δοκιμών. Τα δεδομένα εκπαίδευσης των 3 πρώτων σειρών (συμβολίζονται ως MoCap) αντιστοιχούν σε δεδομένα που αποκτήθηκαν με συστήματα MoCap υψηλής ποιότητας, ενώ η τελευταία σειρά αντιστοιχεί σε δεδομένα που αποκτήθηκαν με συστήματα MoCap χωρίς δείκτες (αναφέρονται ως Real). Τα διαφορετικά σύνολα δεδομένων εξισώθηκαν όσον αφορά τα συνολικά λεπτά MoCap και επιλέχθηκαν διαφορετικά τμήματα του συνόλου δεδομένων AMASS για τις παραλλαγές MoCap, με την απαραίτητη χρονική υποδειγματοληψία σε αυτά, ώστε να προκύψει συνολική διάρκεια περίπου 9 λεπτών MoCap, η οποία είναι η ποσότητα δεδομένων MoCap που είναι διαθέσιμη στα δεδομένα Real. Οι τελευταίες χρησιμοποίησαν τα GeneBody [*Cheng et al., 2022*] και THuman2.0 [*Yu et al., 2021*], ενώ οι πρώτες χρησιμοποίησαν τρεις διαφορετικούς συνδυασμούς τμημάτων συνόλων δεδομένων. Το MoCapv1 χρησιμοποίησε τα EKUT, HumanEva, MoSh και Soma για εκπαίδευση. Το MoCapv2 χρησιμοποίησε το CNRS και το HumanEva και το MoCapv3 χρησιμοποίησε μόνο το HumanEva για εκπαίδευση. Τα αποτελέσματα καταδεικνύουν ότι το μοντέλο που εκπαιδεύτηκε σε

πραγματικά δεδομένα αποδίδει ισάξια ή και καλύτερα από τις παραλλαγές δεδομένων του MoCap σε ορισμένες περιπτώσεις. Πλεονεκτικά, αυτό επικυρώνει την ικανότητα της αποκαλυπτόμενης εφεύρεσης να παρέχει ένα σύστημα MoCap υψηλής ποιότητας χρησιμοποιώντας οικονομικά προσιτά μέσα απόκτησης δεδομένων MoCap χωρίς δείκτες, όταν βασίζεται σε ένα τυποποιημένη αναπαράσταση χαμηλής διάστασης για την απόκτηση και τη δημιουργία δεδομένων εκπαίδευσης.

Προσαρμογή μοντέλου σώματος.

[0101] Αφού εκπαιδευτεί το νευρωνικό δίκτυο **103** και αποθηκευτούν οι παράμετροι του στη βάση δεδομένων παραμέτρων του μοντέλου **107**, μπορούν να φορτωθούν κατά παραγγελία για να ικανοποιήσουν συγκεκριμένες συνθήκες καταγραφής. Ένα παράδειγμα είναι η εκπαίδευση μοντέλων με διαφορετικές διατάξεις δεικτών και, ανάλογα με τη διάταξη των δεικτών που είναι προσαρτημένη στη στολή του καταγραφόμενου υποκειμένου, μπορούν να φορτωθούν οι αντίστοιχες παράμετροι του νευρωνικού δικτύου. Στη συνέχεια, στο νευρωνικό δίκτυο **103** μπορούν να παρουσιαστούν τα ανεπεξέργαστα δεδομένα θέσης του δείκτη **102**, όπως εκτιμώνται από μια διάταξη αισθητήρων που περιλαμβάνει $K \geq 1$ οπτικούς αισθητήρες/κάμερες **101**. Το νευρωνικό δίκτυο **103** θα προβλέψει αντίστοιχα ορόσημα **104**, με το να αντιστοιχεί στο σχήμα εκπαίδευσης του αντίστοιχου μοντέλου. Ωστόσο, αυτό δεν περιορίζεται σε ένα μόνο μοντέλο, καθώς το σχήμα εκπαίδευσης που παρουσιάζεται στο ΣΧΗΜΑ 2 είναι γενικό, ώστε να υποστηρίζει την αλυσιδωτή παλινδρόμηση. Ένα μοντέλο μπορεί να προβλέψει πρώτα ένα σύνολο ορόσημων που θα χρησιμοποιηθεί ως είσοδος σε ένα άλλο μοντέλο που εκπαιδεύεται για να συμπεράνει ένα άλλο σύνολο ορόσημων. Για παράδειγμα, σε μια άλλη υλοποίηση, ένα πρωτογενές νευρωνικό δίκτυο μπορεί να προβλέψει αποθορυβοποιημένους δείκτες και ένα άλλο νευρωνικό δίκτυο μπορεί να επιλύσει αυτά τους προβλεπόμενους αποθορυβοποιημένους δείκτες σε θέσεις αρθρώσεων.

[0102] Στο τέλος αυτής της επεξεργασίας από διαδοχικά νευρωνικά δίκτυα που περιλαμβάνει $1,...,C$ βήματα, είναι διαθέσιμο ένα σύνολο αποθορυβοποιημένων και

ολοκληρωμένων γνωστών ορόσημων $\ell_{est} \in \mathbb{R}^{L\times 3}$. Η παρούσα εφεύρεση αποκαλύπτει επιπροσθέτως μια μέθοδο **105** για την προσαρμογή του σώματος σε αυτές τις εκτιμήσεις προκειμένου να ληφθούν οι παράμετροι $(\beta, \theta, \mathbf{T})$ **106** του μοντέλου σώματος που ανακατασκευάζουν μια σαφή αναπαράσταση με τη μορφή αρθρωτού σκελετού και επιφάνειας πλέγματος.

[0103] Η εκτίμηση των παραμέτρων του σώματος $(\beta, \theta, \mathbf{T})$ **106** είναι ένα μη γραμμικό πρόβλημα βελτιστοποίησης **105** με την τυπική λύση να είναι το MoSh [*Loper et al., 2014*] και η συνέχεια του MoSh++ [*Mahmood et al., 2019*] που αξιοποιεί επίσης ένα χαμηλής διάστασης παραμετρικό μοντέλο σώματος. Ωστόσο, το MoSh(++) επιλύει επιπρόσθετα και τη διάταξη των δεικτών, η οποία στην περίπτωσή μας είναι γνωστή εκ των προτέρων, καθώς το νευρωνικό δίκτυο **103** που εκτιμά τα ℓ_{est} **104** φορτώθηκε από μια βάση δεδομένων **107** που περιέχει πολλαπλά εκπαιδευμένα μοντέλα, το καθένα με μια προκαθορισμένη διαμόρφωση που αντιστοιχούσε στο τρέχον πλαίσιο καταγραφής. Επομένως, η επίλυση της διάταξης των δεικτών δεν είναι απαραίτητη για το σύστημά μας, διαφοροποιώντας το σημαντικά από τις παραλλαγές MoSh που επικεντρώνονται στην επεξεργασία αρχειακών δεδομένων MoCap.

[0104] Επιπλέον, το MoSh υποθέτει ότι οι εκτιμήσεις είναι υψηλής ποιότητας ή, διαφορετικά, με χαμηλό λόγο σήματος προς θόρυβο, που καταγράφονται από ακριβά συστήματα MoCap υψηλής τεχνολογίας. Η παρούσα εφεύρεση αποκαλύπτει ένα σύστημα MoCap το οποίο αντ' αυτού χαλαρώνει αυτή την υπόθεση για να υποστηρίξει πρόσθετες επιλογές αισθητήρων **101**, συμπεριλαμβανομένων προσιτών εναλλακτικών λύσεων χαμηλότερης ποιότητας που παράγουν δεδομένα εισόδου με διάφορα επίπεδα θορύβου. Η λύση για την προσαρμογή σε θορυβώδη δεδομένα είναι η εύρωστη βελτιστοποίηση, αλλά οι τυπικές προσεγγίσεις που περιλαμβάνουν εύρωστους πυρήνες/εκτιμητές απαιτούν σίγουρη γνώση για την υποκείμενη κατανομή δεδομένων. Αυτή δεν είναι εύκολα διαθέσιμη στην πράξη και επιπλέον, ποικίλλει ανάλογα με τις διάφορες επιλογές καταγραφής.

[0105] Πιο σημαντικό είναι ότι η παρούσα εφεύρεση παρουσιάζει ένα σύστημα που ενσωματώνει ένα μοντέλο μηχανικής μάθησης με βάση τα δεδομένα με τη μορφή ενός νευρωνικού δικτύου **103** που παράγει αυτές τις εκτιμήσεις **104**. Επομένως, αν και το νευρωνικό δίκτυο **103** εκπαιδεύεται για να αποθορυβεί τις εισόδους του **102**, εισάγει μια άλλη μορφή θορύβου πληροφορίας, η κατανομή του οποίου είναι πολύ δύσκολο να μοντελοποιηθεί.

[0106] Μια λύση σε αυτό είναι η χρήση μιας προσαρμοστικής παραλλαγής εύρωστης βελτιστοποίησης, με παράδειγμα την χρήση της συνάρτησης σφάλματος Barron [*Barron, 2019*], η οποία προσαρμόζεται στην υποκείμενη κατανομή και παρεμβάλλει/γενικεύει πολλές γνωστές παραλλαγές προσαρμόζοντας από κοινού το σχήμα και την κλίμακά τους. Παρόλα αυτά, η βελτιστοποίηση μιας συνάρτησης προσαρμογής με αυτόν τον στόχο δεν έχει αποδειχθεί ότι αποτελεί πρακτική εναλλακτική λύση, καθώς αυτός ο τύπος συνάρτησης σφάλματος έχει χρησιμοποιηθεί μόνο στο πλαίσιο της στοχαστικής βελτιστοποίησης για την εκπαίδευση νευρωνικών δικτύων.

[0107] Αντ' αυτού, η παρούσα εφεύρεση αποκαλύπτει μια διατύπωση βασισμένη στην πιθανοθεωρία για τη διατύπωση ενός στόχου προσαρμογής με επίγνωση του θορύβου, ο οποίος είναι προσαρμοστικός και βελτιστοποιεί την Γκαουσιανή περιοχή αβεβαιότητας $\boldsymbol{\sigma} \in \mathbb{R}^L$ από κοινού με τα δεδομένα και τους προηγούμενους όρους:

$$\underset{\boldsymbol{\theta}^*,\beta^*,\mathbf{T}^*,\boldsymbol{\sigma}^*}{\operatorname{argmin}} \; \mathcal{L}_{data} + \mathcal{L}_{prior}, \tag{12}$$

όπου L είναι ο αριθμός των ορόσημων που βελτιστοποιούνται. Χρησιμοποιείται επίσης ένα σύνολο όρων που κωδικοποιούν την εκ των προτέρων γνώση μας για τις παραμέτρους:

$$\mathcal{L}_{prior} = \lambda_{\beta} \sum ||\beta||_2 + \lambda_{\mathbf{z}} \sum ||\mathbf{z}||_2, \tag{13}$$

όπου και οι δύο είναι μια στατιστική ποινή $L2$ από το μέσο σχήμα και τη μέση πόζα,

δεδομένης της κανονικής μοντελοποίησής τους, λόγω της στατιστικής μοντελοποίησης (Ανάλυση Κυρίων Συνιστωσών - ΑΚΣ) για το πρώτο (σχήμα), και της κανονικοποίησης του λανθάνοντος χώρου για το δεύτερο (πόζα). Στην πράξη, δεδομένης της πρότερης γνώσης του γεννητικού μοντέλου αυτοκωδικοποίησης για την πόζα θ, η επίλυση μπορεί είτε να λάβει χώρα απευθείας στις παραμέτρους πόζας θ, με το σκοπό της πρότερης γνώσης να επιβάλλεται στη μετακωδικοποίηση των παραμέτρων πόζας, είτε η επίλυση μπορεί να χρησιμοποιήσει τον λανθάνοντα χώρο $\mathbf{z}$ που θα αποκωδικοποιηθεί στις παραμέτρους πόζας θ και ο σκοπός της πρότερης γνώσης να επιβάλλεται απευθείας στον επιλυμένο λανθάνοντα κώδικα. Ουσιαστικά, η πόζα θ και ο λανθάνων κώδικας $\mathbf{z}$ θεωρούνται εναλλάξιμες από τις συναρτήσεις της Εξίσωσης (1). Οι τιμές για τα λ_β και λ_θ μπορούν να οριστούν εμπειρικά, αλλά μπορούν επίσης να ανακαλυφθούν με πειράματα προσομοίωσης θορύβου. Ο όρος δεδομένων διατυπώνεται ως εξής:

$$\mathcal{E}_{data} = \sum_i^L \frac{1}{2\sigma_i^2} ||\ell_{est,i} - \ell_i^*||_2 + log\sigma_i, \tag{14}$$

και βελτιστοποιεί ταυτόχρονα την περιοχή αβεβαιότητας ως σ, καθώς και τον όρο προσαρμογής $L2$ των ορόσημων.

[0108] Πραγματοποιείται σταδιακή βελτιστοποίηση με ανόπτηση, αλλά με μόνο 2 στάδια, καθώς δεν υπάρχει βελτιστοποίηση της διάταξης των δεικτών. Το πρώτο στάδιο βελτιστοποιεί επί των $\beta^*, \theta^*, \mathbf{T}^*$, ενώ το δεύτερο στάδιο καθορίζει τα β και $\mathbf{T}$ και βελτιστοποιεί τα θ^*, σ^*. Αυτό εξασφαλίζει τη σωστή αρχικοποίηση για τη βελτιστοποίηση της περιοχής αβεβαιότητας, αποφεύγοντας αποτελεσματικά τα τοπικά ελάχιστα και τις αβάσιμες λύσεις. Για το δεύτερο στάδιο οι προηγούμενοι όροι χαλαρώνουν επίσης στο ήμισυ των τιμών τους, καθώς επιτυγχάνεται ένα καλό τοπικό βέλτιστο στο πρώτο στάδιο, όπου οι αυξημένοι προηγούμενοι όροι βοηθούν στο να γίνει η αντικειμενική συνάρτηση πιο καλά συμπεριφερόμενη (δηλαδή κυρτή).

[0109] Ο Πίνακας 4 παρουσιάζει αποτελέσματα για την αποκαλυπτόμενη μέ-

Table 3

	n_d	n_m	δείκτες	αρθρώσεις	RMSE_j (m) ↓	RMSE_m (m) ↓	PCK1 ↑	PCK3 ↑	PCK7 ↑
MoSh(++)			✓	✗	0.0301	0.0333	11.79%	66.85%	98.34%
Εύρωστο	✓	✗	✓	✗	0.0290	0.0320	14.75%	69.85%	98.18%
Εύρωστο $_\text{j}$			✓	✓	0.0234	0.0275	19.66%	81.06%	99.11%
MoSh(++)			✓	✗	0.0206	0.0225	28.71%	89.03%	99.05%
Εύρωστο	✗	✓	✓	✗	0.0187	0.0197	41.99%	90.95%	98.81%
Εύρωστο $_\text{j}$			✓	✓	0.0185	0.0204	42.18%	91.44%	98.56%
MoSh(++)			✓	✗	0.0238	0.0261	24.26%	85.63%	98.22%
Εύρωστο	✓	✓	✓	✗	0.0224	0.0242	36.01%	87.14%	97.53%
Εύρωστο $_\text{j}$			✓	✓	0.0219	0.0242	36.67%	88.09%	97.69%

θοδο προσαρμογής σε σύγκριση με την τυπική προσέγγιση που παρουσιάζεται στο MoSh(++), αλλά χωρίς το βήμα βελτιστοποίησης της διάταξης των δεικτών, το οποίο δεν είναι απαραίτητο. Τεχνητός θόρυβος **205** προστίθεται στις θέσεις δεικτών που παράγονται από την προαναφερθείσα προσέγγιση εξαγωγής δεδομένων **204** και επιπλέον, οι εισροές αυτές τροφοδοτούνται στο νευρωνικό δίκτυο για την εκτίμηση των απαιτούμενων θέσεων ορόσημων, προσθέτοντας τον θόρυβο του εκτιμητή του νευρωνικού δικτύου. Αυτά οι διαφορετικές πηγές θορύβου συνδυάζονται όπως υποδεικνύεται στον Πίνακα, και πλεονεκτικά, η αποκακλυπρόμενη εφευρετική μέθοδος διαχειρίζεται και τις δύο πηγές θορύβου, παρουσιάζοντας σαφείς βελτιώσεις σε σύγκριση με την τυπική προσέγγιση προσαρμογής.

[0110] Μπορούν να προστεθούν έξτρα περιορισμοί πρότερης γνώσης p με τα αντίστοιχα βάρη λ_p , όπως είναι οι περιορισμοί ορίου στις αρθρώσεις ή η πρότερη γνώση που προσφέρεται από κατανομές βασισμένες σε δεδομένα, όπως κωδικοποιούνται από ένα γεννητικό μοντέλο κανονικοποιημένων ροών [*Rezende et al., 2015*]. Αυτή είναι ενδεικτική πρότερη γνώση που θεωρείται κομμάτι της υλοποίησης μαζί με τον αλγόριθμο βελτιστοποίησης που μπορεί να περιλαμβάνει επίλυση quasi-Newton, ή αλγόριθμους Gauss-Newton ή Levenberg-Marquardt για επίλυση σε πραγματικό χρόνο. Κατά προτίμηση, για να μειωθεί η πολυπλοκότητα, η επίλυση θα ελαχιστοποιεί τα κοινά ορόσημα μαζί με ένα μικρό υποσύνολο επιφανειακών ορόσημων για την επίλυση των περιστροφών των άκρων.

[0111] Για να μειωθεί περαιτέρω η πολυπλοκότητα, μπορεί να χρησιμοποιηθεί μια διαδικασία βαθμονόμησης σχήματος ώστε να εκτελείται μόνο η κινηματική βελτιστοποίηση κατά την επίλυση, καθώς το σχήμα θα θεωρούνταν σταθερό και γνωστό εκ των προτέρων με τη λήψη των παραμέτρων του αντικειμένου/υποκειμένου από μια βάση δεδομένων ή με επίλυση μια και μόνο φορά πριν από την καταγραφή της ερμηνείας/κίνησης.

[0112] Η εξαγωγή του σχήματος μπορεί να είναι το αποτέλεσμα της βελτιστοποίησης της Εξίσωσης (12) για ένα ή περισσότερα καρέ. Τα πολλαπλά καρέ χρησιμοποιούνται για την απομάκρυνση των επιδράσεων του θορύβου στους βελτιστοποιημένους συντελεστές σχήματος. Στην περίπτωση αυτή, εκτός από τη βελτιστοποίηση ενός ενιαίου διανύσματος συντελεστών σχήματος, είναι επίσης δυνατό να βελτιστοποιηθούν όλα τα καρέ χωριστά και στη συνέχεια να συγκεντρωθούν οι προκύπτοντες συντελεστές σχήματος του κάθε καρέ σε ένα ενιαίο διάνυσμα παραμέτρων και να υπολογιστεί ο μέσος όρος τους.

[0113] Ακόμα και σε αυτή την περίπτωση, τυχόν σφάλματα που προέρχονται από τη σωστή προσαρμογή της πόζας θα επηρέαζαν επίσης την ποιότητα της λύσης σχήματος. Έτσι, μια πιο αποτελεσματική προσέγγιση είναι η επίλυση του σχήματος μόνο μέσω της διατύπωσης του προβλήματος με τρόπο που δεν επηρεάζει το τοπικό σύστημα συντεταγμένων, εξαλείφοντας πρακτικά την εξάρτηση από την εκτίμηση της περιστροφής κάθε άρθρωσης. Χρησιμοποιώντας τα εκτιμώμενα ορόσημα ℓ_{est} διατυπώνεται ένα πρόβλημα βελτιστοποίησης, χρησιμοποιώντας τον ακόλουθο στόχο:

$$\mathcal{E}_{data} = \sum_{p,c}^{P} \Big| \, ||\ell_{est,p} - \ell_{est,p}||_2 - ||\ell_p^* - \ell_c^*||_2 \, \Big|, \tag{15}$$

ουσιαστικά ελαχιστοποιώντας τη διαφορά της απόστασης μεταξύ ενός συνόλου ζευγών ορόσημων. Εφόσον το σύνολο των ζευγών P επιλέγεται έτσι ώστε ο γονέας p και το παιδί c να είναι άκαμπτα κινούμενα ορόσημα, αυτός ο στόχος είναι ανεξάρτητος από

την παράμετρο θ της πόζας. Ειδικά στην περίπτωση που τα ορόσημα ℓ_{est} περιλαμβάνουν τόσο δείκτες όσο και αρθρώσεις, είναι δυνατή η εκτίμηση σχήματος με υψηλής ακρίβεια.

Βελτίωση μετά τη καταγραφή.

[0114] Η παρούσα εφεύρεση αξιοποιεί την ενοποίηση της αναπαράστασης σώματος και κίνησης όπως παρέχεται από το παραμετρικό σώμα, για να αυτοματοποιήσει τη μεταεπεξεργασία των δεδομένων εφαρμογής του σώματος που παράγονται από το σύστημα σε πραγματικό χρόνο. Πλεονεκτικά, η μετα-επεξεργασία **600** προσπελαύνει μια εγγραφή/τμήμα επίδοσης και χρησιμοποιεί ένα προηγούμενο κίνησης όπως κωδικοποιείται από ένα γεννητικό μοντέλο **111** για να βελτιώσει **601** τις εξόδους πριν αποθηκεύσει τη νέα επεξεργασμένη κίνηση **602** (σε μνήμη, αρχείο ή σε ένα σύνολο δεδομένων). Στόχος της διαδικασίας βελτίωσης είναι η διόρθωση πιθανών εσφαλμένων αποτελεσμάτων του συστήματος πραγματικού χρόνου, η αφαίρεση του θορύβου και η παραγωγή ομαλότερων κινήσεων. Αυτό επιτυγχάνεται με την αξιοποίηση παρελθοντικών και μελλοντικών πληροφοριών ή τεχνικά με την επίλυση για ένα σετ καρέ ταυτόχρονα **601**.

[0115] Αναλυτικότερα, ορίζουμε το παρακάτω πρόβλημα βελτιστοποίησης:

$$\underset{\mathbf{z}^t,\beta^t,\mathbf{T}^t}{\operatorname{argmin}} \mathcal{E}_{data}^t + \mathcal{E}_{prior}^t, \tag{16}$$

με τη βελτιστοποίηση να εκτελείται σε έναν λανθάνοντα κώδικα $\mathbf{z} \in \mathbb{R}^L$ που εκφράζει μια εκ των προτέρων εκτίμηση των παραμέτρων θ, όπως εξηγείται στην ενότητα της προσαρμογής του σώματος. Αυτό μπορεί να επεκταθεί σε ένα χρονικό παράθυρο $\mathcal{T} := [0, \ldots, T]$, λύνοντας για μια δέσμη πλαισίων ταυτόχρονα, όπως ορίζεται παρακάτω:

$$\underset{\mathbf{z}^t,\mathbf{T}^t,t\in\mathcal{T}}{\operatorname{argmin}} \sum_{t=0}^{T} \mathcal{E}_{data} + \sum_{t=0}^{T} \mathcal{E}_{prior}, \tag{17}$$

με το σχήμα β να παραλείπεται από την βελτιστοποίηση και να θεωρείται σταθερό.

[0116] Προηγούμενες εργασίες επιλύουν τις παραμέτρους πόζας θ απευθείας χρησιμοποιώντας μεγάλα χρονικά παράθυρα ταυτόχρονα και προσθέτοντας και χρονικούς περιορισμούς, με τους [*Mahmood et al., 2019*] να επιλύουν 30 καρέ (χρονικά στιγμιότυπα) ταυτόχρονα χρησιμοποιώντας έναν ΔΜΣ (Διακριτό Μετασχηματισμό Συνημιτόνου) αφού πρώτα επιλύσουν κάθε καρέ ξεχωριστά, και τους [*Arnab et al., 2019*] επιλύουν για ολόκληρη την ακολουθία χρησιμοποιώντας έναν κοινό στόχο ομαλότητας στον χρόνο, έχοντας αρχικές εκτιμήσεις από ένα μοντέλο μηχανικής μάθησης. Άλλες εργασίες θεωρούν ομαλά μεταβαλλόμενες παρελθόντες κινήσεις και προσθέτουν έναν στόχο εξομάλυνσης πεπερασμένων διαφορών των λανθανόντων κωδίκων πόζας [*Huang et al., 2021*], ή λειτουργούν απευθείας σε τροχιές θέσης [*Hang, et al, 2020*]. Ωστόσο, αυτές δεν είναι αποτελεσματικές, καθώς απαιτούν επιπλέον στάδια επίλυσης, δεν είναι επεκτάσιμες όταν χρησιμοποιούν πεπερασμένες διαφορές και απαιτούν την ρύθμιση μιας πρόσθετης υπερπαραμέτρου για τον στόχο ομαλότητας.

[0117] Αυτή η εφεύρεση βασίζεται στην επίλυση σε ένα κλιμακούμενο χρονικό παράθυρο με τη χρήση ανακατασκευής λανθανόντων κωδίκων πόζας μέσω παρεμβολής και ενός συνόλου καρέ-κλειδιών, που καθορίζουν τα σημεία έναρξης και λήξης ενός τμήματος ομαλής μετάβασης της πόζας. Αυτό μειώνει σημαντικά τον χώρο παραμέτρων που απαιτούν βελτιστοποίηση, καθώς μπορούν να χρησιμοποιηθούν δύο λανθάνοντα καρέ-κλειδιά στις χρονικές στιγμές 0 και :

$$\underset{\mathbf{z}^0,\mathbf{z}^T,\mathbf{T}^0,\mathbf{T}^T}{\operatorname{argmin}} \mathcal{E}_{data}^{\mathcal{T}} + \mathcal{E}_{prior}^{0,T}, \tag{18}$$

με τον όρο που αφορά την εκ των προτέρων γνώση για τις πόζες να επιβάλλεται μόνο στα καρέ-κλειδιά και τον όρο των δεδομένων να ορίζεται σε ολόκληρο το παράθυρο $\mathcal{T}$:

$$\mathcal{E}_{data}^{\mathcal{T}} = \sum_{t=0}^{T} \mathcal{E}_{data}^{t}. \tag{19}$$

Για να λύσουμε το πρόβλημα σε ολόκληρο το χρονικό παράθυρο και να προσαρμόσουμε τη λύση μας σε όλους τους διαθέσιμους περιορισμούς δεδομένων, αναδομούμε τα ενδιάμεσα πλαίσια μέσω παρεμβολής:

$$\theta^t = \mathcal{G}(\varsigma^t(\mathbf{z}^0, \mathbf{z}^T)), \begin{bmatrix} \mathbf{R}^t & \mathbf{t}^t \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} \varsigma^t(\mathbf{R}^0, \mathbf{R}^T) & \xi^t(\mathbf{t}^0, \mathbf{t}^T) \\ \mathbf{0} & 1 \end{bmatrix}, \tag{20}$$

με ς^t και ξ^t να είναι αντίστοιχα σφαιρικές και γραμμικές συναρτήσεις παρεμβολής που αντιστοιχούν το t σε ένα κλειστό μοναδιαίο διάστημα ($[0, 1]$) εντός του T για να συνδυάσουν τα σημεία αρχής $t = 0$ και τέλους $t = T$. Ειδικότερα για την πόζα θ, η διαδικασία αυτή βασίζεται σε σημαντικό βαθμό σε ένα καλά εκπαιδευμένο γεννήτικό μοντέλο $\mathcal{G}$ που αποτυπώνει έναν τοπολογικό χώρο $\mathcal{M}$ ο οποίος μπορεί να παρεμβληθεί ομαλά για να ανακατασκευάσει αντίστοιχα ομαλές μεταβάσεις στο χώρο της πόζας.

[0118] Αυτό εισάγει εμμέσως τους περιορισμούς ομαλότητας στον τοπολογικό χώρο αντί ενός ρητού στόχου ομαλότητας, χωρίς να απαιτείται ρύθμιση των αντίστοιχων υπερπαραμέτρων. Η παρούσα τεχνική μπορεί επιπροσθέτως να υλοποιηθεί με έναν κυλυώμενο τρόπο, θεωρώντας το πρώτο καρέ-κλειδί, $t = 0$, ως αρχική συνθήκη και συνεπώς, σταθερή. Με αυτόν τον τρόπο, είναι ταυτόγχρονα αποδοτική και επεκτάσιμη, καθώς δεν απαιτεί επίλυση επί όλων των παραμέτρων σε ένα πλεονάζον χρονικό παράθυρο και, αντίθετα, βελτιστοποιεί μόνο τον λανθάνοντα κώδικα του επόμενου καρέ-κλειδιού $t = T$.

[0119] Κάθε ανακατασκευασμένο διάνυσμα παραμέτρων πόζας θ^t ανακατασκευάζει ένα σύνολο από ορόσημα **604** που χρησιμοποιούνται στον στόχο δεδομένων, ο οποίος μπορεί να είναι ένα σφάλμα L2 ή ο στόχος με επίγνωση του θορύβου της εξίσωσης (15), ή οποιονδήποτε άλλο ανθεκτικό εκτιμητή, χωρίς απώλεια της γενικότητας. Ο στόχος διατυπώνεται ώστε το ανακαταστκευασμένο χρονικό παράθυρο μέσω του κάθε επόμενου λυμένου καρέ-κλειδιού να αντιστοιχεί στα εκτιμώμενα ορόσημα **104** των ενδιάμεσων καρέ **602**, ξεκινώντας από μια σταθερή προηγούμενη εκτίμηση. Χρησιμοποιείται ένα

κατάλληλο γεννητικό μοντέλο **605** που επιλέγεται από μια βάση δεδομένων γεννητικών μοντέλων **110**, η οποία μπορεί να είναι οποιοδήποτε γεννητικό μοντέλο που μπορεί να ανακατασκευάσει μια ακολουθία ποζών που αναπαριστούν ένα τμήμα κίνησης.

[0120] Ένας ειδικευμένος επαγγελματίας καταλαβαίνει ότι οποιοσδήποτε αριθμός καρέ-κλειδιών μπορεί να χρησιμοποιηθεί με διαφορετικές τεχνικές πολυμεταβλητής παρεμβολής, αρκεί να διασχίζουν κατάλληλα τον λανθάνοντα χώρο του εκμαθημένου τοπολογικού χώρου ώστε να παράγουν ρεαλιστικές πόζες. Παρομοίως, ενώ χρησιμοποιήθηκε μια στατική διαδικασία διαχωρισμού των χρονικών παραθύρων, αυτή μπορεί να προσαρμοστεί άμεσα σε μια δυναμική/προσαρμοστική διαδικασία διαχωρισμού χρονικών παραθύρων, με ποικίλα επίπεδα επικάλυψης και αρχικοποίησης των παραμέτρων τους. Ένα ενδεικτικό και καθόλου περιοριστικό παράδειγμα θα ήταν μια επικάλυψη 3 καρέ μεταξύ κάθε τμήματος, με την αρχική κατάσταση των καρέ-κλειδιών του επόμενου τμήματος να αρχικοποιείται, πριν από την επίλυση, χρησιμοποιώντας ένα μοντέλο πρόβλεψης κίνησης που εκτελεί εξωτερική πρόβλεψη χρησιμοποιώντας τη λύση των επικαλυπτόμενων τμημάτων ως είσοδο.

Υπολογιστικό περιβάλλον.

[0121] Το ΣΧΗΜΑ 8 απεικονίζει ένα υποδειγματικό υπολογιστικό περιβάλλον στο οποίο μπορούν να εφαρμοστούν διάφορες υλοποιήσεις της εφεύρεσης και στο οποίο μπορούν να χρησιμοποιηθούν διάφορες υλοποιήσεις της εφεύρεσης. Το περιβάλλον του υπολογιστικού συστήματος είναι μόνο ένα παράδειγμα ενός κατάλληλου υπολογιστικού περιβάλλοντος και δεν έχει σκοπό να υποδηλώσει οποιονδήποτε περιορισμό όσον αφορά το πεδίο χρήσης ή τη λειτουργικότητα. Μπορούν να χρησιμοποιηθούν πολυάριθμα άλλα περιβάλλοντα ή διαμορφώσεις υπολογιστικού συστήματος γενικού ή ειδικού σκοπού. Παραδείγματα γνωστών υπολογιστικών συστημάτων, περιβαλλόντων και/ή διαμορφώσεων που μπορεί να είναι κατάλληλα για χρήση περιλαμβάνουν, μεταξύ άλλων, προσωπικές ηλεκτρονικές συσκευές, όπως έξυπνα τηλέφωνα και έξυπνα ρολόγια, υπολογιστές ταμπλέτας, προσωπικούς υπολογιστές (PC), υπολογιστές διακομιστή,

φορητές συσκευές ή φορητούς υπολογιστές, συστήματα πολλαπλών επεξεργαστών, συστήματα βασισμένα σε μικροεπεξεργαστές, δικτυακούς υπολογιστές, μικροϋπολογιστές, κεντρικούς υπολογιστές, ενσωματωμένα συστήματα, κατανεμημένα υπολογιστικά περιβάλλοντα, όπως το νέφος, που περιλαμβάνουν οποιοδήποτε από τα παραπάνω συστήματα ή συσκευές, και τα παρόμοια.

[0122] Μπορούν να χρησιμοποιηθούν εντολές εκτελέσιμες από υπολογιστή, όπως ενότητες προγραμμάτων που εκτελούνται από υπολογιστή. Γενικά, οι ενότητες προγράμματος περιλαμβάνουν ρουτίνες, προγράμματα, αντικείμενα, στοιχεία, δεδομένα δομές κ.λπ. που εκτελούν συγκεκριμένες εργασίες ή υλοποιούν συγκεκριμένους αφηρημένους τύπους δεδομένων. Μπορούν να χρησιμοποιηθούν περιβάλλοντα υπολογιστών με διασκορπισμένη χρήση, όπου οι εργασίες εκτελούνται από απομακρυσμένες συσκευές επεξεργασίας που συνδέονται μέσω ενός δικτύου επικοινωνιών ή άλλου μέσου μετάδοσης δεδομένων. Σε ένα κατανεμημένο υπολογιστικό περιβάλλον, οι ενότητες προγράμματος και άλλα δεδομένα μπορεί να βρίσκονται τόσο σε τοπικά όσο και σε απομακρυσμένα μέσα αποθήκευσης υπολογιστών, συμπεριλαμβανομένων συσκευών αποθήκευσης μνήμης.

[0123] Αναφορικά με το ΣΧΗΜΑ 7, ένα υποδειγματικό σύστημα για την υλοποίηση πτυχών που περιγράφονται στο παρόν περιλαμβάνει μια υπολογιστική συσκευή, όπως η υπολογιστική συσκευή **1000**. Στην πιο βασική της διαμόρφωση, η υπολογιστική συσκευή **1000** τυπικά περιλαμβάνει τουλάχιστον μία μονάδα επεξεργασίας **1002** και μνήμη **1004**. Ανάλογα με την ακριβή διαμόρφωση και τον τύπο της υπολογιστικής συσκευής, η μνήμη **1004** μπορεί να είναι ασταθής (όπως μνήμη τυχαίας προσπέλασης RAM), σταθερή (όπως μνήμη μόνο για ανάγνωση ROM, μνήμη φλας κλπ.), ή κάποιος συνδυασμός των δύο. Αυτή η πιο βασική διαμόρφωση απεικονίζεται στο ΣΧΗΜΑ 7 με τη διακεκομμένη γραμμή **1006**. Η υπολογιστική συσκευή **1000** μπορεί να έχει πρόσθετα χαρακτηριστικά/λειτουργικότητα. Για παράδειγμα, η υπολογιστική συσκευή **1000** μπορεί να περιλαμβάνει πρόσθετο αποθηκευτικό χώρο (αφαιρούμενο ή/και μη αφαιρούμενο),

συμπεριλαμβανομένων, μεταξύ άλλων, μαγνητικών ή οπτικών δίσκων ή ταινιών. Τέτοια πρόσθετη αποθήκευση απεικονίζεται στο ΣΧΗΜΑ 7 με την αφαιρούμενη αποθήκευση **1008** και τη μη αφαιρούμενη αποθήκευση **1100**. Η υπολογιστική συσκευή **1000**, όπως χρησιμοποιείται εδώ, μπορεί να είναι είτε μια φυσική συσκευή υλικού, είτε μια εικονική συσκευή, είτε ένας συνδυασμός αυτών.

[0124] Η υπολογιστική συσκευή **1000** τυπικά περιλαμβάνει ή είναι εφοδιασμένη με μια ποικιλία μέσων αναγνώσιμων από υπολογιστή. Τα αναγνώσιμα από υπολογιστή μέσα μπορούν να είναι οποιαδήποτε διαθέσιμα μέσα στα οποία μπορεί να έχει πρόσβαση η υπολογιστική συσκευή **1000** και περιλαμβάνουν τόσο ασταθή όσο και σταθερά μέσα, αφαιρούμενα και μη αφαιρούμενα μέσα. Ενδεικτικά και όχι περιοριστικά, τα αναγνώσιμα από υπολογιστή μέσα μπορούν να περιλαμβάνουν μέσα αποθήκευσης υπολογιστή και μέσα επικοινωνίας.

[0125] Τα μέσα αποθήκευσης υπολογιστών περιλαμβάνουν ασταθή και σταθερά, αφαιρούμενα και μη αφαιρούμενα μέσα που υλοποιούνται με οποιαδήποτε μέθοδο ή τεχνολογία για την αποθήκευση πληροφοριών, όπως εντολές αναγνώσιμες από υπολογιστή, δομές δεδομένων, μονάδες προγράμματος ή άλλα δεδομένα. Η μνήμη **1004**, η αφαιρούμενη αποθήκευση **1008** και η μη αφαιρούμενη αποθήκευση **1010** είναι όλα παραδείγματα μέσων αποθήκευσης υπολογιστή. Τα μέσα αποθήκευσης υπολογιστών περιλαμβάνουν, μεταξύ άλλων, τη μνήμη τυχαίας προσπέλασης (RAM), τη μνήμη μόνο για ανάγνωση (ROM), την ηλεκτρικά διαγράψιμη προγραμματιζόμενη μνήμη μόνο για ανάγνωση (EEPROM), τη μνήμη φλας ή άλλη τεχνολογία μνήμης, τη μνήμη μόνο για ανάγνωση compact disc (CD-ROM), τους ψηφιακούς ευέλικτους δίσκους (DVD) ή άλλη οπτική αποθήκευση, τις μαγνητικές κασέτες, τη μαγνητική ταινία, την αποθήκευση μαγνητικών δίσκων ή άλλες συσκευές μαγνητικής αποθήκευσης ή οποιοδήποτε άλλο μέσο που μπορεί να χρησιμοποιηθεί για την αποθήκευση της επιθυμητής πληροφορίας και στο οποίο μπορεί να έχει πρόσβαση η υπολογιστική συσκευή **1000**. Οποιοδήποτε τέτοιο μέσο αποθήκευσης υπολογιστή μπορεί να αποτελεί μέρος της υπολογιστικής

συσκευής **1000**.

[0126] Η υπολογιστική συσκευή **1000** μπορεί επίσης να περιέχει σύνδεση(-εις) επικοινωνίας **1012** που επιτρέπουν στη συσκευή να επικοινωνεί με άλλες συσκευές. Κάθε τέτοια σύνδεση επικοινωνιών **1012** είναι ένα παράδειγμα μέσου επικοινωνίας. Τα μέσα επικοινωνίας ενσωματώνουν συνήθως εντολές αναγνώσιμες από υπολογιστή, δομές δεδομένων, ενότητες προγράμματος ή άλλα δεδομένα σε ένα διαμορφωμένο σήμα δεδομένων, όπως ένα κύμα φορέα ή άλλο μηχανισμό μεταφοράς και περιλαμβάνει οποιοδήποτε μέσο παροχής πληροφοριών. Ο όρος "διαμορφωμένο σήμα δεδομένων" σημαίνει ένα σήμα του οποίου ένα ή περισσότερα από τα χαρακτηριστικά έχουν οριστεί ή μεταβληθεί κατά τρόπο ώστε να κωδικοποιούνται πληροφορίες στο σήμα. Ενδεικτικά και όχι περιοριστικά, τα μέσα επικοινωνίας περιλαμβάνουν ενσύρματα μέσα, όπως ενσύρματο δίκτυο ή απευθείας ενσύρματη σύνδεση, και ασύρματα μέσα, όπως ακουστικά, ραδιοσυχνότητας (RF), υπέρυθρα και άλλα ασύρματα μέσα. Ο όρος μέσα αναγνώσιμα από υπολογιστή, όπως χρησιμοποιείται εδώ, περιλαμβάνει τόσο μέσα αποθήκευσης όσο και μέσα επικοινωνίας. Η υπολογιστική συσκευή **1000** μπορεί επίσης να διαθέτει συσκευή(ες) εισόδου **1014**, όπως πληκτρολόγιο, ποντίκι, στυλό, συσκευή φωνητικής εισόδου, συσκευή εισόδου αφής κ.λπ. Μπορεί επίσης να περιλαμβάνονται συσκευή(ες) εξόδου **1016** όπως οθόνη, ηχεία, εκτυπωτής κ.λπ. Όλες αυτές οι συσκευές είναι γενικά γνωστές και συνεπώς δεν χρειάζεται να συζητηθούν λεπτομερώς στο παρόν, εκτός από τα προβλεπόμενα.

[0127] Ειδικότερα, η υπολογιστική συσκευή **1000** μπορεί να είναι μία από μια πλειάδα υπολογιστικών συσκευών **1000** που συνδέονται μεταξύ τους με ένα δίκτυο **1018**, όπως φαίνεται στο ΣΧΗΜΑ 7. Όπως μπορεί να εκτιμηθεί, το δίκτυο **1018** μπορεί να είναι οποιοδήποτε κατάλληλο δίκτυο- κάθε υπολογιστική συσκευή **1000** μπορεί να συνδέεται σε αυτό μέσω μιας σύνδεσης **1012** με οποιονδήποτε κατάλληλο τρόπο, και κάθε υπολογιστική συσκευή **1000** μπορεί να επικοινωνεί με μία ή περισσότερες από τις άλλες υπολογιστικές συσκευές **1000** στο δίκτυο **1018** με οποιονδήποτε κατάλληλο

τρόπο. Για παράδειγμα, το δίκτυο **1018** μπορεί να είναι ένα ενσύρματο ή ασύρματο δίκτυο εντός ενός οργανισμού ή ενός σπιτιού ή κάτι παρόμοιο, και μπορεί να περιλαμβάνει άμεση ή έμμεση σύνδεση με ένα εξωτερικό δίκτυο, όπως το διαδίκτυο ή κάτι παρόμοιο.

[0128] Θα πρέπει να γίνει κατανοητό ότι οι διάφορες τεχνικές που περιγράφονται στο παρόν μπορούν να εφαρμοστούν μέσω υλικού ή λογισμικού ή, κατά περίπτωση, με συνδυασμό και των δύο. Έτσι, οι μέθοδοι και οι συσκευές του παρόντος αποκαλυπτόμενου αντικειμένου, ή ορισμένες πτυχές ή τμήματά τους, μπορούν να λάβουν τη μορφή κώδικα προγράμματος (δηλαδή οδηγιών) που ενσωματώνεται σε απτά μέσα, όπως μονάδες φλας του καθολικού σειριακού διαύλου (USB), κάρτες μνήμης Secure Digital (SD), CD- ROMs, σκληρούς δίσκους ή οποιοδήποτε άλλο μέσο αποθήκευσης αναγνώσιμο από μηχανή, όπου, όταν ο κώδικας προγράμματος φορτώνεται και εκτελείται από μια μηχανή, όπως ένας υπολογιστής, η μηχανή γίνεται μια συσκευή για την άσκηση του παρόντος αποκαλυπτόμενου αντικειμένου.

[0129] Στην περίπτωση εκτέλεσης κώδικα προγράμματος σε προγραμματιζόμενους υπολογιστές, η υπολογιστική συσκευή περιλαμβάνει γενικά έναν επεξεργαστή, ένα μέσο αποθήκευσης αναγνώσιμο από τον επεξεργαστή (συμπεριλαμβανομένης της ασταθούς και μη ασταθούς μνήμης ή/και των στοιχείων αποθήκευσης), τουλάχιστον μία συσκευή εισόδου και τουλάχιστον μία συσκευή εξόδου. Ένα ή περισσότερα προγράμματα μπορούν να υλοποιούν ή να χρησιμοποιούν τις διαδικασίες που περιγράφονται στην αποκαλυπτόμενη εφεύρεση, π.χ. μέσω της χρήσης μιας διεπαφής προγράμματος εφαρμογής (API), επαναχρησιμοποιήσιμων στοιχείων ελέγχου ή παρόμοιων. Τέτοια προγράμματα μπορούν να υλοποιούνται σε μια υψηλού επιπέδου διαδικαστική ή αντικειμενοστραφή γλώσσα προγραμματισμού για την επικοινωνία με ένα σύστημα υπολογιστή. Ωστόσο, το(α) πρόγραμμα(α) μπορεί να υλοποιηθεί(ουν) σε γλώσσα Assembly ή σε γλώσσα μηχανής, εάν είναι επιθυμητό. Σε κάθε περίπτωση, η γλώσσα προγραμματισμού μπορεί να είναι μεταγλωττισμένη ή διερμηνευμένη και να συνδυάζεται με υλοποιήσεις υλικού. Σε

μια υλοποίηση, το σύστημα μπορεί να αναπτυχθεί χρησιμοποιώντας C/C ++/Python. Αν και οι υποδειγματικές υλοποιήσεις μπορεί να αναφέρονται στη χρησιμοποίηση πτυχών της ποκαλυπτόμενης εφεύρεσης στο πλαίσιο ενός ή περισσότερων αυτόνομων υπολογιστικών συστημάτων, η εφεύρεση δεν περιορίζεται, αλλά μπορεί να υλοποιηθεί σε συνάρτηση με οποιοδήποτε υπολογιστικό περιβάλλον, όπως ένα δίκτυο **1018** ή ένα κατανεμημένο υπολογιστικό περιβάλλον. Ακόμα περισσότερο, πτυχές της αποκαλυπτόμενης εφεύρεσης μπορούν να υλοποιηθούν σε ένα ή περισσότερα τσιπ επεξεργασίας ή συσκευών, και η αποθήκευση μπορεί ομοίως να πραγματοποιηθεί σε ένα πλήθος συσκευών σε ένα δίκτυο **1018**. Τέτοιες συσκευές μπορεί να περιλαμβάνουν για παράδειγμα προσωπικούς υπολογιστές, διακομιστές δικτύου και φορητές συσκευές.

[0130] Ενώ περιγράφονται ορισμένες υλοποιήσεις, αυτές οι υλοποιήσεις παρουσιάζονται μόνο ως παράδειγμα και δεν προορίζονται να περιορίσουν το πεδίο εφαρμογής της εφεύρεσης. Πράγματι, τα νέα συστήματα και οι μέθοδοι που περιγράφονται στο παρόν μπορούν να υλοποιηθούν με ποικίλες άλλες μορφές. Επιπλέον, μπορούν να γίνουν διάφορες παραλείψεις, αντικαταστάσεις και αλλαγές στη μορφή του συστήματος και των μεθόδων που περιγράφονται στο παρόν χωρίς να απομακρυνθούν από το πνεύμα της εφεύρεσης. Το πεδίο εφαρμογής της εφεύρεσης περιλαμβάνει οποιαδήποτε ισοδύναμα αυτής, όπως θα μπορούσε να εκτιμηθεί από κάποιον με συνήθη δεξιότητα στην τέχνη.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Aristidou, Andreas, et al., "Self-similarity analysis for motion capture cleaning." Computer graphics forum. 2018.

Arnab, Anurag, Carl Doersch, and Andrew Zisserman. "Exploiting temporal context for 3D human pose estimation in the wild." IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.

Baran, Ilya, and Jovan Popović. "Automatic rigging and animation of 3d characters." ACM Transactions on graphics. 2007.

Barron, Jonathan T. "A general and adaptive robust loss function." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.

Branco, Paula, et al., "SMOGN: a pre-processing approach for imbalanced regression." First international workshop on learning with imbalanced domains: Theory and applications. PMLR, 2017.

Chatzitofis, Anargyros, et al., "Democap: low-cost marker-based motion capture." International Journal of Computer Vision. 2021.

Chen, Kang, et al., "MoCap-Solver: A neural solver for optical motion capture data." ACM Transactions on Graphics (TOG). 2021.

Cheng, Wei, et al. "Generalizable neural performer: Learning robust radiance fields for human novel view synthesis." arXiv preprint arXiv:2204.11798. 2022.

Fieraru, Mihai, et al., "Aifit: Automatic 3d human-interpretable feedback models for fitness training." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

Ghorbani, Nima, et al., "Soma: Solving optical marker-based mocap automatically." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

Ghorbani, et al., "Auto-labelling of markers in optical motion capture by permutation learning." Advances in Computer Graphics: 36th Computer Graphics International Conference. 2019.

Gong, Yu, et al., "RankSim: Ranking Similarity Regularization for Deep Imbalanced Regression." arXiv preprint arXiv:2205.15236. 2022.

Han, Shangchen, et al. "Online optical marker-based hand tracking with deep labels." ACM Transactions on Graphics (TOG). 2018.

Holland, Paul W., et al. "Robust regression using iteratively reweighted least-squares." Communications in Statistics-theory and Methods. 1977.

Holden, Daniel. "Robust solving of optical motion capture data by denoising." ACM Transactions on Graphics (TOG). 2018.

Huang, Y., Bogo, F., Lassner, C., Kanazawa, A., Gehler, P. V., Romero, J., ... & Black, M. J. "Towards accurate marker-less human shape and pose estimation over time." In international conference on 3D vision. 2017.

Huang, Y., Taheri, O., Black, M. J., & Tzionas, D. "InterCap: Joint Markerless 3D Tracking of Humans and Objects in Interaction." In DAGM German Conference on Pattern Recognition. 2022.

Huang, B., Shu, Y., Zhang, T., & Wang, Y. "Dynamic multi-person mesh recovery from uncalibrated multi-view cameras." In International Conference on 3D Vision. 2021.

Huang, Buzhen, Yuan Shu, Tianshu Zhang, and Yangang Wang. "Dynamic multi-person mesh recovery from uncalibrated multi-view cameras." International Conference on 3D Vision. 2021.

Ingwersen, C. K., Mikkelstrup, C. M., Jensen, J. N., Hannemose, M. R., & Dahl, A. B. "SportsPose-A Dynamic 3D sports pose dataset." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023

Jacobson, Alec, Ilya Baran, Jovan Popovic, and Olga Sorkine. "Bounded biharmonic weights for real-time deformation." ACM Transactions on Graphics. 2011.

Kaufmann, Manuel, et al. "Em-pose: 3d human pose estimation from sparse electromagnetic trackers." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

Kingma, Diederik P., et al. "Auto-encoding variational bayes." arXiv preprint arXiv:1312.6114 (2013). In 2nd International Conference on Learning Representations, (ICLR). 2014.

Loper, Matthew, et al., "MoSh: motion and shape capture from sparse markers." ACM Transactions on Graphics (TOG). 2014.

Loper, Matthew, et al. "SMPL: A skinned multi-person linear model." ACM transactions on graphics (TOG). 2015.

Mahmood, Naureen, et al. "AMASS: Archive of motion capture as surface shapes." Proceedings of the IEEE/CVF international conference on computer vision. 2019.

Menéndez, M. L., et al. "The jensen-shannon divergence." Journal of the Franklin Institute. 1997.

Nibali, Aiden, et al. "Numerical coordinate regression with convolutional neural networks." arXiv preprint arXiv:1801.07372. 2018.

Nibali, Aiden, et al. "3d human pose estimation with 2d marginal heatmaps." IEEE Winter Conference on Applications of Computer Vision (WACV). 2019.

Osman, Ahmed AA, et al., "Star: Sparse trained articulated human body regressor." Proceedings of the 16th European Conference on Computer Vision. 2020.

Osman, Ahmed AA, et al. "SUPR: A Sparse Unified Part-Based Human Representation." Proceedings of the 17th European Conference on Computer Vision. 2022.

Pavlakos, Georgios, et al. "Expressive body capture: 3d hands, face, and body from a single image." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.

Pavllo, Dario, et al. "Real-time neural network prediction for handling two-hands mutual occlusions." Computers & Graphics. 2019.

Rempe, D., Birdal, T., Hertzmann, A., Yang, J., Sridhar, S., & Guibas, L. J. (2021). Humor: 3d human motion model for robust pose estimation. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 11488-11499).

Ren, Jiawei, et al. "Balanced mse for imbalanced visual regression." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

Rezende, Danilo, et al. "Variational inference with normalizing flows." International conference on machine learning, (PMLR). 2015.

Ronneberger, Olaf, et al. "U-net: Convolutional networks for biomedical image segmentation." 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2015.

Rong, Yu, et al., "Chasing the tail in monocular 3d human reconstruction with prototype memory." IEEE Transactions on Image Processing. 2022.

Saini, N., Huang, C. H. P., Black, M. J., & Ahmad, A. "SmartMocap: Joint Estimation of Human and Camera Motion Using Uncalibrated RGB Cameras." IEEE Robotics and Automation Letters. 2023.

Schubert, Tobias, et al. "Automatic initialization for skeleton tracking in optical motion capture." IEEE International Conference on Robotics and Automation (ICRA). 2015.

Silva, Aníbal, et al., "Model Optimization in Imbalanced Regression." Proceedings of the 25th International Conference on Discovery Science. 2022.

Steininger, Michael, et al. "Density-based weighting for imbalanced regression." Machine Learning. 2021.

Sun, Xiao, et al. "Integral human pose regression." Proceedings of the European conference on computer vision (ECCV). 2018.

Torgo, Luís, et al. "Resampling strategies for regression." Expert Systems. 2015.

Torgo, Luis, et al., "Utility-based regression." PKDD. 2007.

Torgo, Luís, et al. "Smote for regression." 16th Portuguese Conference on Artificial Intelligence. 2013.

Varol, Gul, et al. "Learning from synthetic humans." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

Xu, Hongyi, et al. "Ghum & ghuml: Generative 3d human shape and articulated pose models." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

Xue, Han, et al. "OMAD: Object Model with Articulated Deformations for Pose Estimation and Retrieval." Proceedings of the British Machine Vision Conference. 2021.

Yan, Haonan, et al. "Ultrapose: Synthesizing dense pose with 1 billion points by human-body decoupling 3d model." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

Yang, Yuzhe, et al. "Delving into deep imbalanced regression." International Conference on Machine Learning. PMLR, 2021.

Ye, Hang, et al. "Faster VoxelPose: Real-time 3D Human Pose Estimation by Orthographic Projection." 17th European Conference on Computer Vision, (ECCV). 2022.

Ye, V., Pavlakos, G., Malik, J., & Kanazawa, A. Decoupling human and camera motion from videos in the wild. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.

Yu, Tao, et al. "Function4d: Real-time human volumetric capture from very sparse consumer rgbd sensors." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern

Recognition. 2021.

Zanfir, Andrei, Elisabeta Marinoiu, and Cristian Sminchisescu. "Monocular 3d pose and shape estimation of multiple people in natural scenes-the importance of multiple scene constraints." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

Zhang, Feng, et al. "Distribution-aware coordinate representation for human pose estimation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

Zhang, Richard. "Making convolutional networks shift-invariant again." International conference on machine learning (PMLR). 2019.

Zuffi, Silvia, et al. "3D menagerie: Modeling the 3D shape and pose of animals." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2017.

Hang, Jianwei, et al. "Abnormal value processing method for three-dimensional trajectory data and optical motion capture method", WO2020133447A1, WIPO. 2020.

Michael J. Black, Nima Ghorbani, "Method and system for labelling motion-captured points", EP4163882A2, European Patent Office. 2023.

ΑΞΙΩΣΕΙΣ

Αυτό που υποστηρίζεται είναι:

1 Ένα σύστημα (**100**) βασισμένο σε μηχανική μάθηση που υλοποιείται μέσω υπολογιστή και μπορεί να εκτελεί καταγραφή κίνησης χρησιμοποιώντας δείκτες που προσαρτώνται στο σώμα, με το σύστημα να περιλαμβάνει:

(a) Μία εγκατάσταση οπτικών αισθητήρων (**101**) που παρέχει ανεπεξέργαστες μη δομημένες εκτιμήσεις της θέσης των φυσικών δεικτών (**102**) οι οποίοι είναι προσαρτημένοι στο σώμα ανθρώπου, ζώου ή αντικειμένου,

(b) Μία βάση δεδομένων παραμετρικών μοντέλων σώματος (**108**),

(c) Ένα νευρωνικό δίκτυο (**103**), επιλεγμένο από μια βάση δεδομένων νευρωνικών δικτύων (**107**), το οποίο χρησιμοποιεί τις ανεπεξέργαστες μη δομημένες εκτιμήσεις θέσης δεικτών για να εκτιμήσει ένα σύνολο ορόσημων που σχετίζονται με το σώμα (**104**), εκτελώντας ταυτόχρονα αποθορυβοποίηση των μετρήσεων εισόδου, επισήμανση ορόσημων, επίλυση και ολοκλήρωση, όπου η επιλογή σχετίζεται με,

i. το παραμετρικό μοντέλο σώματος που χρησιμοποιήθηκε για την εκπαίδευση του νευρωνικού δικτύου, ώστε να ταιριάζει με το εικονιζόμενο υποκείμενο/αντικείμενο, και

ii. την διάταξη των δεικτών στο σώμα ώστε να ταιριάζουν με την είσοδο του εκπαιδευόμενου νευρωνικού δικτύου,

(d) Ένα υποσύστημα βαθμονόμησης σχήματος (**109**) που χρησιμοποιεί το εκτιμώμενο σύνολο ορόσημων που σχετίζονται με το υποκείμενο/αντικείμενο για να επιλύσει το σχήμα της εξωτερικής

τρισδιάστατης επιφάνειας και του εσωτερικού κινηματικού σκελετού του υποκειμένου/αντικειμένου, χρησιμοποιώντας μια αντίστοιχη αναπαράσταση παραμετρικών συντελεστών blendshape, που επιλέγεται από τη βάση δεδομένων παραμετρικών μοντέλων σώματος,

(e) Ένα υποσύστημα βελτιστοποίησης (**105**) που χρησιμοποιεί το βαθμονομημένο σχήμα και ένα γεννητικό μοντέλο πόζας ή/και κίνησης, που επιλέγεται από μια βάση δεδομένων γεννητικών μοντέλων (**110**), για να βελτιώσει την εκτιμώμενη αναπαράσταση των ορόσημων, προσαρμόζοντας τις βαθμονομημένες θέσεις ορόσημων στις εκτιμώμενες, παράγοντας μια βελτιωμένη αναπαράσταση σχήματος και κίνησης (**106**),

(f) Μια βάση δεδομένων (**111**) που αποθηκεύει την καταγεγραμμένη κίνηση, η οποία περιλαμβάνει τις βελτιστοποιημένες παραμέτρους του παραμετρικού μοντέλου, συμπεριλαμβανομένων των αναπαραστάσεων σχήματος και πόζας, με τις εν λόγω αναπαραστάσεις πόζας να περιλαμβάνουν τις θέσεις των αρθρώσεων ή/και τις περιστροφές, οπότε η καταγεγραμμένη κίνηση μαζί με τις αναπαραστάσεις σχήματος και πόζας μπορούν να ανακατασκευάσουν την 4Δ κίνηση του υποκειμένου/αντικειμένου.

2 Το σύστημα της αξίωσης **1**,όπου το παραμετρικό μοντέλο σώματος περιλαμβάνει ένα σύνολο από blendshapes που έχουν δημιουργηθεί από καλλιτέχνη, ρυθμισμένα χειροκίνητα και βάρη δέρματος που έχουν ζωγραφιστεί από καλλιτέχνη.

3 Το σύστημα της αξίωσης **1**, όπου το παραμετρικό μοντέλο σώματος βασίζεται σε στατιστική μοντελοποίηση, είτε ντετερμινιστική είτε στοχαστική, ενός συνόλου τρισδιάστατων σαρώσεων ανθρώπινου σώματος. Αυτά μπορεί να περιλαμβάνουν πρόσθετους συντελεστές blendshapes και κοινούς παλινδρομικούς παράγοντες.

4 Το σύστημα της αξίωσης **1**, όπου τα εκτιμώμενα από το νευρωνικό δίκτυο ορόσημα περιλαμβάνουν ένα σύνολο δεικτών. Το σύνολο δεικτών μπορεί να περιλαμβάνει τους πραγματικούς δείκτες που έχουν τοποθετηθεί στο υποκείμενο/αντικείμενο, αλλά δεν περιορίζεται σε αυτούς. Πρόσθετοι δείκτες μπορούν να εκτιμηθούν ανάλογα με τη διαμόρφωση της εκπαίδευσης.

5 Το σύστημα των αξιώσεων **1**, **2**, **3**, όπου τα εκτιμώμενα από το νευρωνικό δίκτυο ορόσημα περιλαμβάνουν τις αρθρώσεις του σώματος του ανθρώπου ή του ζώου ή του αντικειμένου. Οι εσωτερικές αρθρώσεις επιλύονται από το μοντέλο, παράγοντας μη παρατηρούμενα ορόσημα.

6 Το σύστημα των αξιώσεων **1**, **2**, **3**, όπου τα εκτιμώμενα από το νευρωνικό δίκτυο ορόσημα περιλαμβάνουν σημεία επιφάνειας που αντιστοιχούν στις θέσεις στις οποίες έχουν τοποθετηθεί οι δείκτες. Αυτές οι εκτιμώμενες επιφανειακές θέσεις δεν είναι μετρήσιμες από το οπτικό σύστημα που παρέχει την είσοδο.

7 Το σύστημα των αξιώσεων **1**, **2**, **3**, όπου τα εκτιμώμενα από το νευρωνικό δίκτυο ορόσημα περιλαμβάνουν δείκτες και αρθρώσεις. Το νευρωνικό δίκτυο επιλύει τόσο τους δείκτες εισόδου όσο και τις θέσεις των αρθρώσεων του ανθρώπινου σώματος ή του σώματος του ζώου ή του αντικειμένου.

8 Το σύστημα των αξιώσεων **1**, **2**, **3**, όπου τα εκτιμώμενα από το νευρωνικό δίκτυο ορόσημα περιλαμβάνουν σημεία επιφάνειας και αρθρώσεις. Το νευρωνικό δίκτυο επιλύει τόσο τις θέσεις επιφάνειας που τοποθετούνται με δείκτες όσο και τις εσωτερικές αρθρώσεις.

9 Το σύστημα της αξίωσης **1**, με οποιαδήποτε μετατροπή των μοντέλων σώματος όπως ορίζονται στις αξιώσεις **2** και **3** και τους τρόπους εκπαίδευσης και εκτίμησης συνδυασμών ορόσημων που ορίζονται στις αξιώσεις **4**, **5**, **6**, **7**, ;ή **8**.

10 Το σύστημα των αξιώσεων **1** έως **9**, όπου το υποσύστημα βαθμονόμησης σχήματος χρησιμοποιείται για την ταυτόχρονη επίλυση της πόζας και του σχήματος χρησιμοποιώντας τα εκτιμώμενα ορόσημα.

11 Το σύστημα των αξιώσεων **1** έως **9**, όπου το υποσύστημα βαθμονόμησης σχήματος χρησιμοποιεί μόνο τις κατά ζεύγη σχέσεις των εκτιμώμενων ορόσημων για την επίλυση μόνο των συντελεστών blendshape.

12 Το σύστημα των αξιώσεων **1** έως **11**, όπου η βελτιστοποίηση (**105**) βελτιστοποιεί την περιοχή αβεβαιότητας των ορόσημων που εκτιμώνται από το νευρωνικό δίκτυο επιπρόσθετα των παραμέτρων του παραμετρικού μοντέλου σώματος.

13 Μία μέθοδος κατά την οποία ένα νευρωνικό δίκτυο (**103**) εκπαιδεύεται για να εκτελεί εξισορροπημένη παλινδρόμηση (**200**) χρησιμοποιώντας μια ανισοβαρή κατανομή δεδομένων εκπαίδευσης, η μέθοδος περιλαμβάνει τα εξής στάδια:

(a) Κατά τη διάρκεια ενός πρώτου σταδίου, εκπαίδευται ένα αυτοκωδικοποιητικό γεννητικό μοντέλο, που περιλαμβάνει έναν κωδικοποιητή και έναν αποκωδικοποιητή, χρησιμοποιώντας τα ανισοβαρή δεδομένα, όπου:

i. Ο κωδικοποιητής χρησιμοποιείται για την εξαγωγή μιας λανθάνουσας κωδικής αναπαράστασης των δειγμάτων.

ii. Ο αποκωδικοποιητής χρησιμοποιείται για να εκτιμήσει την ανακατασκευή της αναπαράστασης των δειγμάτων εισόδου χρησιμοποιώντας την κωδικοποιημένη λανθάνουσα αναπαράσταση.

(b) Κατά τη διάρκεια ενός δεύτερου σταδίου, παράγονται συνθετικά δείγματα με χρήση του αποκωδικοποιητή του αυτοκωδικοποιητικού γεννητικού μοντέλου για τη συμπλήρωση των διαθέσιμων δεδομένων και αποδίδονται

τιμές σημαντικότητας στα δεδομένα εκπαίδευσης με χρήση του εν λόγω αυτοκωδικοποιητικού γεννητικού μοντέλου,

i. Με τον τρόπο αυτό, το δεύτερο στάδιο περιλαμβάνει την εκπαίδευση ενός νευρωνικού δικτύου που εκτελεί παλινδρόμηση, χρησιμοποιώντας δείγματα δεδομένων ή/και τιμές σημαντικότητας,

ii. Με τον τρόπο αυτό, τα δείγματα δεδομένων προέρχονται είτε από τα ανισοβαρή δεδομένα εκπαίδευσης, είτε από τα συνθετικά δείγματα που παράγονται με τη χρήση του αποκωδικοποιητή του αυτοκωδικοποιητικού γεννητικού μοντέλου, είτε από συνδυασμό και των δύο,

iii. Και με την τρόπο αυτό αποδίδεται μια τιμή σημαντικότητας σε κάθε δείγμα δεδομένων με μια συνάρτηση που χρησιμοποιεί τις εισόδους και τις εξόδους του αυτοκωδικοποιητικού γεννητικού μοντέλου.

14 Η μέθοδος της αξίωσης **13**, όπου το αυτοκωδικοποιητικό γεννητικό μοντέλο είναι ένας μη αντιστρέψιμος μεταβλητός αυτοκωδικοποιητής που περιλαμβάνει πλήρως συνδεδεμένα, συνελικτικά ή έναν συνδυασμό πλήρως συνδεδεμένων και συνελικτικών επιπέδων, ή έναν συνδυασμό τέτοιων αντιστρέψιμων μεταβλητών αυτοκωδικοποιητών.

15 Η μέθοδος της αξίωσης **13**, όπου το αυτοκωδικοποιητικό γεννητικό μοντέλο είναι ένα αντιστρέψιμο πιθανολογικό μοντέλο ή μια αλυσίδα αντιστρέψιμων πιθανολογικών μοντέλων που λειτουργούν με διαφορίσιμους και αντιστρέψιμους μετασχηματισμούς, με διαφορίσιμο αντίστροφο.

16 Η μέθοδος σύμφωνα με οποιαδήποτε από τις αξιώσεις **13**, **14**, **15**, όπου ο αποκωδικοποιητής παράγει δείγματα συνθετικών δεδομένων χρησιμοποιώντας ως είσοδο τυχαίους λανθάνοντες κώδικες που προέρχονται από ομοιόμορφη,

κανονική ή οποιαδήποτε άλλη κατανομή.

17 Η μέθοδος σύμφωνα με οποιαδήποτε από τις αξιώσεις **13**, **14**, **15**, όπου ο αποκωδικοποιητής παράγει δείγματα συνθετικών δεδομένων χρησιμοποιώντας ως είσοδο λανθάνοντες κώδικες που προέρχονται από μια ομοιόμορφη, κανονική ή οποιαδήποτε άλλη κατανομή, γύρω από έναν αριθμό προκαθορισμένων σταθερών λανθανόντων κωδίκων που χρησιμοποιούνται ως κέντρα των κατανομών με ποικίλες τυπικές αποκλίσεις.

18 Η μέθοδος σύμφωνα με την αξίωση **17**, όπου ένας αριθμός συνθετικών δειγμάτων δεδομένων χρησιμοποιείται για την ανακατασκευή νέων συνθετικών δειγμάτων δεδομένων με χρήση πολυμεταβλητής παρεμβολής με τυχαίες παραμέτρους παρεμβολής.

19 Η μέθοδος σύμφωνα με την αξίωση **18**, όπου δύο συνθετικά δείγματα δεδομένων χρησιμοποιούνται για την ανακατασκευή ενός νέου συνθετικού δείγματος δεδομένων με χρήση σφαιρικής γραμμικής παρεμβολής με τυχαία παράμετρο παρεμβολής που δειγματοληπτείται από μία ομοιόμορφη κατανομή προκαθορισμένου εύρους με κατώτερο όριο το 0 και ανώτερο όριο το 1.

20 Η μέθοδος σύμφωνα με οποιαδήποτε από τις αξιώσεις **17**, **18**, **19**, όπου οι προκαθορισμένοι σταθεροί λανθάνοντες κώδικες είναι οποιοιδήποτε από τους ακόλουθους:

(a) δημιουργημένοι χειροκίνητα,

(b) επιλέγονται χειροκίνητα από οποιαδήποτε διαθέσιμα δεδομένα, τα οποία επιλέγονται μετά την εξαγωγή μιας στατιστικής αναπαράστασης των δεδομένων με τη χρήση της τιμής σημαντικότητας που αποδίδεται σε κάθε δείγμα δεδομένων, όπου η επιλογή πραγματοποιείται:

i. αυτόματα, ή

ii. χειροκίνητα

(c) ή με οποιονδήποτε συνδυασμό τους.

21 Η μέθοδος σύμφωνα με οποιαδήποτε από τις αξιώσεις **13** έως **20**, όπου η τιμή σημαντικότητας αποδίδεται χρησιμοποιώντας οποιοδήποτε από τα ακόλουθα:

(a) την είσοδο και την ανακατασκευασμένη έξοδο του αυτοκωδικοποιητικού γεννητικού μοντέλου, ή,

(b) την αναπαράσταση λανθάνοντα κώδικα του αυτοκωδικοποιητικού γεννητικού μοντέλου, ή,

(c) οποιουδήποτε συνδυασμού τους.

22 Η μέθοδος σύμφωνα με οποιαδήποτε από τις αξιώσεις **13** έως **21**, όπου η τιμή σημαντικότητας αποδίδεται στα δείγματα πραγματικών δεδομένων από τα ανισοβαρή δεδομένα εκπαίδευσης, αλλά καθορίζεται σε μια καθορισμένη τιμή για τα παραγόμενα δείγματα συνθετικών δεδομένων.

23 Η μέθοδος σύμφωνα με οποιαδήποτε από τις αξιώσεις **13** έως **21**, όπου η τιμή σημαντικότητας για όλα τα δείγματα είναι σταθερή στη μονάδα.

24 Η μέθοδος σύμφωνα με οποιαδήποτε από τις αξιώσεις **13** έως **23**, όπου τα δείγματα συνθετικών δεδομένων δεν χρησιμοποιούνται κατά την εκπαίδευση του μοντέλου.

25 Η μέθοδος σύμφωνα με οποιαδήποτε από τις αξιώσεις **13** έως **24**, όπου η τιμή σημαντικότητας προσαρμόζεται ως συνάρτηση του βήματος εκπαίδευσης.

26 Η μέθοδος σύμφωνα με οποιαδήποτε από τις αξιώσεις **13** έως **25**, όπου η τιμή σημαντικότητας είναι σταθερή στη μονάδα μέχρι ένα προκαθορισμένο βήμα

εκπαίδευσης και μετά από αυτό το βήμα εκπαίδευσης αποδίδεται στα δείγματα δεδομένων.

27 Η μέθοδος σύμφωνα με οποιαδήποτε από τις αξιώσεις **13** έως **26**, όπου τα παραγόμενα δείγματα συνθετικών δεδομένων χρησιμοποιούνται κατά την εκπαίδευση του μοντέλου μετά από ένα προκαθορισμένο βήμα εκπαίδευσης.

28 Ένα σύστημα που λειτουργεί για να εκπαιδεύσει ένα μοντέλο νευρωνικού δικτύου (**103**) ώστε να εκτελεί ισορροπημένη παλινδρόμηση χρησιμοποιώντας δεδομένα εκπαίδευσης, που περιλαμβάνουν δείγματα δεδομένων, τα οποία παρουσιάζουν ανισοβαρή κατανομή. Το σύστημα περιλαμβάνει:

(a) Μέσα, λειτουργικά κατά τη διάρκεια ενός πρώτου σταδίου, τα οποία είναι προσαρμοσμένα να εκπαιδεύουν ένα αυτοκωδικοποιητικό γεννητικό μοντέλο (**207**), το οποίο περιλαμβάνει έναν κωδικοποιητή και έναν αποκωδικοποιητή, χρησιμοποιώντας τα ανισοβαρή δεδομένα, όπου:

i. Ο κωδικοποιητής χρησιμοποιείται για την εξαγωγή ενός λανθάνων κώδικα αναπαράστασης των δειγμάτων.

ii. Ο αποκωδικοποιητής χρησιμοποιείται για να εκτιμήσει την ανακατασκευή (**401**) της αναπαράστασης των δειγμάτων εισόδου χρησιμοποιώντας το λανθάνων κώδικα.

(b) Μέσα που λειτουργούν κατά τη διάρκεια ενός δεύτερου σταδίου, τα οποία είναι προσαρμοσμένα να παράγουν συνθετικά δείγματα (**202**) χρησιμοποιώντας τον αποκωδικοποιητή του αυτοκωδικοποιητικού γεννητικού μοντέλου για να συμπληρώνουν τα δεδομένα εκπαίδευσης, τα εν λόγω μέσα είναι προσαρμοσμένα να αποδίδουν τιμές σημαντικότητας (**208**) στα δεδομένα εκπαίδευσης χρησιμοποιώντας το εν λόγω αυτοκωδικοποιητικό γεννητικό μοντέλο:

i. Όπου, κατά τη διάρκεια του δεύτερου σταδίου, τα εν λόγω μέσα είναι προσαρμοσμένα να εκπαιδεύουν ένα νευρωνικό δίκτυο που εκτελεί παλινδρόμηση, χρησιμοποιώντας δείγματα δεδομένων και/ή τιμές σημαντικότητας,

ii. Όπου, κατά τη διάρκεια του δεύτερου σταδίου, τα εν λόγω μέσα είναι προσαρμοσμένα ώστε να αντλούν τα δείγματα δεδομένων είτε από τα ανισοβαρή δεδομένα εκπαίδευσης είτε από τα συνθετικά δείγματα που παράγοται χρησιμοποιώντας τον αποκωδικοποιητή του αυτοκωδικοποιητικού γεννητικού μοντέλου ή από συνδυασμό και των δύο,

iii. Και όπου, κατά τη διάρκεια του εν λόγω δεύτερου σταδίου, τα εν λόγω μέσα προσαρμόζονται ώστε να αποδίδουν μια τιμή σημαντικότητας σε κάθε δείγμα δεδομένων με μια συνάρτηση που χρησιμοποιεί τις εισόδους και τις εξόδους του αυτοκωδικοποιητικού γεννητικού μοντέλου.

29 Ένα σύστημα σύμφωνα με την αξίωση **28**, όπου το αυτοκωδικοποιητικό γεννητικό μοντέλο είναι ένας μη αντιστρέψιμος μεταβλητός αυτοκωδικοποιητής που περιλαμβάνει πλήρως συνδεδεμένα, συνελικτικά ή συνδυασμό πλήρως συνδεδεμένων και συνελικτικών επιπέδων ή συνδυασμό τέτοιων μη αντιστρέψιμων μεταβλητών αυτοκωδικοποιητών.

30 Ένα σύστημα σύμφωνα με την αξίωση **28**, όπου το αυτοκωδικοποιητικό γεννητικό μοντέλο είναι ένα αντιστρέψιμο πιθανολογικό μοντέλο ή μια αλυσίδα αντιστρέψιμων πιθανολογικών μοντέλων που λειτουργούν με διαφορίσιμους και αντιστρέψιμους μετασχηματισμούς, με διαφορίσιμο αντίστροφο.

31 Ένα σύστημα σύμφωνα με οποιαδήποτε από τις αξιώσεις **28**,**29**, ή **30**, όπου ο

αποκωδικοποιητής περιλαμβάνει μέσα για τη δημιουργία συνθετικών δειγμάτων δεδομένων χρησιμοποιώντας ως είσοδο τυχαίους λανθάνοντες κώδικες που προέρχονται από ομοιόμορφη, κανονική ή οποιαδήποτε άλλη κατανομή.

32 Ένα σύστημα σύμφωνα με οποιαδήποτε από τις αξιώσεις **28**,**29**, ή **30**, όπου ο αποκωδικοποιητής περιλαμβάνει μέσα για τη δημιουργία συνθετικών δειγμάτων δεδομένων χρησιμοποιώντας ως είσοδο λανθάνοντες κώδικες που προέρχονται από μια ομοιόμορφη, κανονική ή οποιαδήποτε άλλη κατανομή γύρω από έναν αριθμό προκαθορισμένων σταθερών λανθανόντων κωδίκων που χρησιμοποιούνται ως κέντρα των κατανομών με ποικίλες κλίμακες.

33 Ένα σύστημα σύμφωνα με την αξίωση **32**, όπου ένας αριθμός συνθετικών δειγμάτων δεδομένων χρησιμοποιείται για την ανακατασκευή νέων συνθετικών δειγμάτων δεδομένων χρησιμοποιώντας πολυμεταβλητή παρεμβολή με τυχαίες παραμέτρους παρεμβολής.

34 Ένα σύστημα σύμφωνα με την αξίωση **33**, όπου δύο δείγματα συνθετικών δεδομένων χρησιμοποιούνται για την ανακατασκευή ενός νέου δείγματος συνθετικών δεδομένων με χρήση σφαιρικής γραμμικής παρεμβολής με τυχαία παράμετρο παρεμβολής που δειγματοληπτείται ομοιόμορφα από ένα προκαθορισμένο εύρος με κατώτερο όριο το 0 και ανώτερο όριο το 1.

35 Ένα σύστημα σύμφωνα με οποιαδήποτε από τις αξιώσεις **32**, **33**, ή **34**, όπου οι προκαθορισμένοι σταθεροί λανθάνοντες κώδικες είναι οποιοιδήποτε από τους εξής:

(a) δημιουργήθηκαν χειροκίνητα,

(b) επιλέγονται χειροκίνητα από όλα τα διαθέσιμα δεδομένα,

(c) επιλέγονται μετά την εξαγωγή μιας στατιστικής αναπαράστασης των

δεδομένων με τη χρήση της τιμής σημαντικότητας που αποδίδεται σε κάθε δείγμα δεδομένων, όπου πραγματοποιείται η επιλογή:

i. αυτόματα, ή

ii. χειροκίνητα

(d) ή οποιοσδήποτε συνδυασμός των α ή β ή γ αυτών.

36 Ένα σύστημα σύμφωνα με οποιαδήποτε από τις αξιώσεις **28** έως **35**, όπου τα μέσα απόδοσης τιμών σημαντικότητας αποδίδουν τιμές χρησιμοποιώντας οποιοδήποτε από τα ακόλουθα:

(a) την είσοδο και την ανακατασκευασμένη έξοδο του αυτοκωδικοποιητικού γεννητικού μοντέλου, ή,

(b) την αναπαράσταση λανθάνουσας κωδικοποίησης του γεννητικού μοντέλου αυτόματης κωδικοποίησης, ή,

(c) οποιονδήποτε συνδυασμό τους.

37 Ένα σύστημα σύμφωνα με οποιαδήποτε από τις αξιώσεις **28** έως **35**, όπου τα μέσα απόδοσης τιμών σημαντικότητας, αποδίδουν τιμές στα δείγματα πραγματικών δεδομένων από τα ανισοβαρή δεδομένα εκπαίδευσης, όπου οι εν λόγω σημαντικές τιμές μένουν σταθερές σε μια καθορισμένη τιμή για τα παραγόμενα δείγματα συνθετικών δεδομένων.

38 Ένα σύστημα σύμφωνα με οποιαδήποτε από τις αξιώσεις **28** έως **35**, όπου η τιμή σπουδαιότητας για όλα τα δείγματα είναι σταθερή στη μονάδα.

39 Ένα σύστημα σύμφωνα με οποιαδήποτε από τις αξιώσεις **28** έως **37**, όπου τα δείγματα συνθετικών δεδομένων δεν χρησιμοποιούνται κατά την εκπαίδευση του μοντέλου.

40 Ένα σύστημα σύμφωνα με οποιαδήποτε από τις αξιώσεις **28** έως **38**, όπου η τιμή σημαντικότητας προσαρμόζεται ως συνάρτηση του βήματος εκπαίδευσης.

41 Ένα σύστημα σύμφωνα με οποιαδήποτε από τις αξιώσεις **28** έως **39**, όπου η τιμή σπουδαιότητας είναι σταθερή στη μονάδα μέχρι ένα προκαθορισμένο βήμα εκπαίδευσης και μετά από αυτό το βήμα εκπαίδευσης αποδίδεται στα δείγματα δεδομένων.

42 Ένα σύστημα σύμφωνα με οποιαδήποτε από τις αξιώσεις **28** έως **40**, όπου τα παραγόμενα δείγματα συνθετικών δεδομένων χρησιμοποιούνται κατά την εκπαίδευση του μοντέλου μετά από ένα προκαθορισμένο βήμα εκπαίδευσης.

43 Σύστημα σύμφωνα με οποιαδήποτε από τις αξιώσεις **28** έως **42**, όπου το εκπαιδευμένο μοντέλο νευρωνικού δικτύου προστίθεται στη βάση δεδομένων νευρωνικών δικτύων οποιουδήποτε από τα συστήματα των αξιώσεων **1** έως **12**.

44 Μέθοδος για την επεξεργασία μιας καταγεγραμμένης κίνησης, χρησιμοποιώντας ένα γεννητικό νευρωνικό δίκτυο που μπορεί να παράγει ένα τμήμα κίνησης χρησιμοποιώντας έναν αριθμό λανθανόντων διανυσμάτων/κωδίκων, όπου η καταγεγραμμένη κίνηση αναπαρίσταται ως διαδοχικά καρέ πόζας, τα εν λόγω καρέ πόζας περιλαμβάνουν θέσεις ορόσημων και/ή περιστροφές/μετασχηματισμούς, η μέθοδος περιλαμβάνει τα εξής στάδια:

(a) Διαχωρισμό της καταγεγραμμένης κίνησης σε κομμάτια χρονικού παραθύρου:

i. κάθε τμήμα αποτελείται από ένα συνεχές σε χρόνο σύνολο καρέ πόζας,

ii. όπου τα κομμάτια μπορούν να επικαλύπτονται,

(b) Για κάθε τμήμα χρονικού παραθύρου, αρχικοποίηση ενός συνόλου λανθανόντων κωδίκων που χρησιμοποιειούνται από το γεννητικό

νευρωνικό δίκτυο για τη δημιουργία ενός τμήματος κίνησης που ταιριάζει στο αντίστοιχο τμήμα χρονικού παραθύρου,

(c) Για κάθε τμήμα χρονικού παραθύρου, βελτιστοποίηση του συνόλου αρχικοποιημένων λανθανόντων κωδίκων ώστε να ταιριάζουν καλύτερα με τα καρέ πόζας εντός κάθε τμήματος χρονικού παραθύρου ελαχιστοποιώντας την απόσταση μεταξύ του τμήματος κίνησης που παράγεται με τη χρήση του γεννητικού νευρωνικού δικτύου και των παρατηρήσεων που περιλαμβάνουν τα καρέ πόζας, τα οποία καρέ πόζας περιλαμβάνονται στο διαχωρισμένο τμήμα χρονικού παραθύρου,

45 Η μέθοδος σύμφωνα με την αξίωση **44**, όπου το γεννητικό νευρωνικό δίκτυο παράγει ένα τμήμα κίνησης χρησιμοποιώντας έναν μόνο λανθάνοντα κώδικα.

46 Η μέθοδος σύμφωνα με την αξίωση **44**, όπου το γεννητικό νευρωνικό δίκτυο είναι ένα αυτοπαλινδρομικό μοντέλο που παράγει ένα τμήμα κίνησης χρησιμοποιώντας έναν αριθμό λανθανόντων κωδίκων.

47 Η μέθοδος σύμφωνα με την αξίωση **44**, όπου το γεννητικό νευρωνικό δίκτυο παράγει ένα τμήμα κίνησης χρησιμοποιώντας έναν αριθμό διαδοχικών λανθανόντων κωδίκων.

48 Η μέθοδος σύμφωνα με την αξίωση **47**, στην οποία επιβάλλονται πρόσθετοι περιορισμοί απόστασης μεταξύ των βελτιστοποιημένων λανθανόντων κωδίκων για να διασφαλιστεί η χρονική διαδοχή τους.

49 Η μέθοδος σύμφωνα με την αξίωση **44**, όπου το γεννητικό νευρωνικό δίκτυο παράγει ένα σύνολο καρέ-κλειδιών πόζας χρησιμοποιώντας έναν αριθμό λανθανόντων κωδίκων και στη συνέχεια ανακατασκευάζει το τμήμα κίνησης χρησιμοποιώντας πολυμεταβλητή παρεμβολή μεταξύ των καρέ πόζας.

50 Η μέθοδος σύμφωνα με την αξίωση **49**, στην οποία χρησιμοποιούνται δύο καρέ-κλειδιά πόζας, το πρώτο είναι το πρώτο καρέ του χρονικού παραθύρου και το δεύτερο είναι το τελευταίο καρέ του χρονικού παραθύρου εντός του τμήματος του χρονικού παραθύρου, και το τμήμα κίνησης ανακατασκευάζεται μέσω σφαιρικής γραμμικής παρεμβολής μεταξύ των δύο παραγόμενων καρέ-κλειδιών.

51 Η μέθοδος σύμφωνα με τις αξιώσεις **44** έως **50**, όπου το γεννητικό νευρωνικό δίκτυο εκπαιδεύτηκε με περιορισμούς στην κατανομή των λανθανόντων κωδίκων και οι εν λόγω περιορισμοί χρησιμοποιούνται επίσης στους βελτιστοποιημένους λανθάνοντες κώδικες κατά τη διαδικασία βελτιστοποίησης.

52 Η μέθοδος σύμφωνα με τις αξιώσεις **44** έως **51**, όπου το γεννητικό νευρωνικό δίκτυο είναι ένα μοντέλο αυτοκωδικοποιητή, που περιλαμβάνει έναν κωδικοποιητή και έναν αποκωδικοποιητή, και του οποίου ο κωδικοποιητής χρησιμοποιείται για την αρχικοποίηση του συνόλου λανθανόντων κωδίκων.

53 Η μέθοδος σύμφωνα με την αξίωση **50**, στην οποία χρησιμοποιείται μια διαδικασία βελτιστοποίησης ολισθαίνοντος παραθύρου που περιλαμβάνει τα ακόλουθα βήματα:

(a) τα κομμάτια του χρονικού παραθύρου χωρίζονται ώστε να επικαλύπτονται από ένα μόνο καρέ, με αποτέλεσμα το τελευταίο του προηγούμενου κομματιού να συμπίπτει με το πρώτο του τρέχοντος κομματιού, κάνοντας ουσιαστικά κάθε ζεύγος διαδοχικών κομματιών να επικαλύπτεται από ένα μόνο καρέ,

(b) το πρώτο καρέ-κλειδί του πρώτου χρονικού παραθύρου αρχικοποιείται και δεν βελτιστοποιείται,

(c) το δεύτερο καρέ-κλειδί του πρώτου τμήματος του χρονικού παραθύρου βελτιστοποιείται έτσι ώστε το τμήμα κίνησης να ταιριάζει με τα καρέ πόζας του τμήματος του χρονικού παραθύρου,

(d) για κάθε τμήμα χρονικού παραθύρου που ακολουθεί το πρώτο, τα ακόλουθα βήματα επαναλαμβάνονται μέχρι να υποβληθεί σε επεξεργασία ολόκληρη η καταγεγραμμένη κίνηση, που περιλαμβάνει όλα τα τμήματα χρονικού παραθύρου:

i. το δεύτερο καρέ-κλειδί του προηγούμενου τμήματος χρονικού παραθύρου χρησιμοποιείται ως το πρώτο καρέ-κλειδί του τρέχοντος τμήματος χρονικού παραθύρου και δεν βελτιστοποιείται,

ii. το δεύτερο καρέ-κλειδί του τρέχοντος τμήματος του χρονικού παραθύρου βελτιστοποιείται έτσι ώστε το τμήμα κίνησης να ταιριάζει με τα καρέ πόζας του τμήματος του χρονικού παραθύρου.

54 Η μέθοδος σύμφωνα με την αξίωση **53**, όπου το γεννητικό νευρωνικό δίκτυο εκπαιδεύτηκε με περιορισμούς κατανομής λανθανόντων κωδίκων και οι εν λόγω περιορισμοί χρησιμοποιούνται επίσης στους βελτιστοποιημένους λανθάνοντες κώδικες κατά τη διαδικασία βελτιστοποίησης.

55 Η μέθοδος σύμφωνα με τις αξιώσεις **53** ή **54**, όπου το γεννητικό νευρωνικό δίκτυο είναι ένα μοντέλο αυτοκωδικοποιητή, που περιλαμβάνει έναν κωδικοποιητή και έναν αποκωδικοποιητή, και του οποίου ο κωδικοποιητής χρησιμοποιείται για την αρχικοποίηση του συνόλου λανθανόντων κωδίκων.

56 Η μέθοδος σύμφωνα με οποιαδήποτε από τις αξιώσεις **44** έως **55**, όπου η διαδικασία βελτιστοποίησης βελτιστοποιεί επιπλέον την περιοχή αβεβαιότητας των εκτιμήσεων των ορόσημων.

57 Ένα σύστημα που λειτουργεί για την επεξεργασία μιας καταγεγραμμένης κίνησης, χρησιμοποιώντας ένα γεννητικό νευρωνικό δίκτυο που λειτουργεί για την παραγωγή ενός τμήματος κίνησης χρησιμοποιώντας έναν αριθμό διανυσμάτων λανθανόντων κωδίκων, όπου η καταγεγραμμένη κίνηση αναπαρίσταται ως διαδοχικά καρέ πόζας, τα οποία περιλαμβάνουν θέσεις ορόσημων και/ή περιστροφών, με το σύστημα να περιλαμβάνει τα ακόλουθα:

(a) Μέσα που λειτουργούν κατά τη διάρκεια ενός πρώτου σταδίου, τα εν λόγω μέσα είναι προσαρμοσμένα να χωρίζουν την καταγεγραμμένη κίνηση σε κομμάτια χρονικού παραθύρου:

i. κάθε τμήμα αποτελείται από ένα συνεχές σε χρόνο σύνολο από καρέ πόζας,

ii. όπου τα κομμάτια μπορούν να επικαλύπτονται,

(b) Μέσα, λειτουργικά για κάθε τμήμα χρονικού παραθύρου, τα οποία είναι προσαρμοσμένα να αρχικοποιούν ένα σύνολο λανθανόντων κωδίκων που χρησιμοποιούνται από το γεννητικό νευρωνικό δίκτυο για τη δημιουργία ενός τμήματος κίνησης που ταιριάζει με το αντίστοιχο τμήμα χρονικού παραθύρου,

(c) Μέσα που λειτουργούν για κάθε τμήμα χρονικού παραθύρου, τα οποία μέσα λειτουργούν για να βελτιστοποιήσουν το σύνολο των αρχικοποιημένων λανθανόντων κωδίκων ώστε να ταιριάζουν με τα καρέ πόζας εντός κάθε τμήματος χρονικού παραθύρου ελαχιστοποιώντας την απόσταση μεταξύ του τμήματος κίνησης που παράγεται με τη χρήση του γεννητικού νευρωνικού δικτύου και των παρατηρήσεων που περιλαμβάνουν τα καρέ πόζας, τα οποία περιλαμβάνουν το διαχωρισμένο τμήμα χρονικού παραθύρου,

58 Ένα σύστημα σύμφωνα με την αξίωση **57**, όπου το γεννητικό νευρωνικό δίκτυο περιλαμβάνει μέσα για τη δημιουργία ενός τμήματος κίνησης χρησιμοποιώντας έναν ενιαίο λανθάνοντα κώδικα.

59 Ένα σύστημα σύμφωνα με την αξίωση **57**, όπου το γεννητικό νευρωνικό δίκτυο είναι ένα αυτοπαλινδρομικό μοντέλο που περιλαμβάνει μέσα για τη δημιουργία ενός τμήματος κίνησης χρησιμοποιώντας έναν αριθμό λανθανόντων κωδίκων.

60 Ένα σύστημα σύμφωνα με την αξίωση **57**, όπου το γεννητικό νευρωνικό δίκτυο περιλαμβάνει μέσα για τη δημιουργία ενός τμήματος κίνησης χρησιμοποιώντας έναν αριθμό διαδοχικών λανθανόντων κωδίκων.

61 Ένα σύστημα σύμφωνα με την αξίωση **60**, όπου επιβάλλονται πρόσθετοι περιορισμοί απόστασης μεταξύ των βελτιστοποιημένων λανθανόντων κωδίκων για να διασφαλιστεί η χρονική διαδοχή τους.

62 Ένα σύστημα σύμφωνα με την αξίωση **57**, όπου το γεννητικό νευρωνικό δίκτυο περιλαμβάνει μέσα για τη δημιουργία ενός συνόλου καρέ-κλειδιών πόζας χρησιμοποιώντας έναν αριθμό λανθανόντων κωδίκων και στη συνέχεια ανακατασκευάζει το τμήμα κίνησης χρησιμοποιώντας πολυμεταβλητή παρεμβολή μεταξύ των καρέ-κλειδιών πόζας.

63 Ένα σύστημα σύμφωνα με την αξίωση **62**, όπου χρησιμοποιούνται δύο καρέ-κλειδιά πόζας, όπου το πρώτο είναι το πρώτο καρέ του χρονικού παραθύρου και το δεύτερο είναι το τελευταίο καρέ του χρονικού παραθύρου εντός του τμήματος του χρονικού παραθύρου, και το τμήμα κίνησης ανακατασκευάζεται μέσω σφαιρικής γραμμικής παρεμβολής μεταξύ των δύο παραγόμενων καρέ-κλειδιών.

64 Ένα σύστημα σύμφωνα με τις αξιώσεις **57** έως **63**, όπου το γεννητικό νευρωνικό δίκτυο εκπαιδεύτηκε με περιορισμούς κατανομής λανθανόντων κωδίκων και οι

εν λόγω περιορισμοί χρησιμοποιούνται επίσης στους βελτιστοποιημένους λανθάνοντες κώδικες κατά τη διαδικασία βελτιστοποίησης.

65 Ένα σύστημα σύμφωνα με τις αξιώσεις **57** έως **64**, όπου το γεννητικό νευρωνικό δίκτυο είναι ένα μοντέλο αυτοκωδικοποιητή, που περιλαμβάνει έναν κωδικοποιητή και έναν αποκωδικοποιητή, και του οποίου ο κωδικοποιητής περιλαμβάνει μέσα για την αρχικοποίηση του συνόλου λανθανόντων κωδίκων.

66 Ένα σύστημα σύμφωνα με τις αξιώσεις **57** έως **65**, όπου χρησιμοποιείται μια διαδικασία βελτιστοποίησης ολισθαίνοντος παραθύρου που περιλαμβάνει τα ακόλουθα:

(a) μέσα που λειτουργούν για τον διαχωρισμό των τμημάτων του χρονικού παραθύρου έτσι ώστε να επικαλύπτονται κατά ένα μόνο καρέ, με αποτέλεσμα το τελευταίο του προηγούμενου τμήματος να συμπίπτει με το πρώτο του τρέχοντος τμήματος, κάνοντας ουσιαστικά κάθε ζεύγος διαδοχικών τμημάτων να επικαλύπτεται κατά ένα μόνο καρέ,

(b) μέσα που λειτουργούν για να αρχικοποιήσουν αλλά όχι να βελτιστοποιήσουν το πρώτο καρέ-κλειδί του πρώτου χρονικού παραθύρου,

(c) μέσα που μπορούν να βελτιστοποιήσουν το δεύτερο καρέ-κλειδί του πρώτου χρονικού παραθύρου έτσι ώστε το τμήμα κίνησης να ταιριάζει με τα καρέ πόζας του χρονικού παραθύρου,

(d) μέσα που λειτουργούν για να επαναλαμβάνουν για κάθε τμήμα χρονικού παραθύρου που ακολουθεί το πρώτο, έως ότου γίνει επεξεργασία ολόκληρης της καταγεγραμμένης κίνησης, που περιλαμβάνει όλα τα τμήματα χρονικού παραθύρου:

i. το δεύτερο καρέ-κλειδί του προηγούμενου τμήματος χρονικού παραθύρου χρησιμοποιείται ως το πρώτο καρέ-κλειδί του τρέχοντος

τμήματος χρονικού παραθύρου και δεν βελτιστοποιείται,

ii. το δεύτερο καρέ-κλειδί του τρέχοντος τμήματος του χρονικού παραθύρου βελτιστοποιείται έτσι ώστε το τμήμα κίνησης να ταιριάζει με τα καρέ πόζας του τμήματος του χρονικού παραθύρου.

67 Το σύστημα σύμφωνα με οποιαδήποτε από τις αξιώσεις **57** έως **65**, όπου η διαδικασία βελτιστοποίησης περιλαμβάνει επιπλέον μέσα βελτιστοποίησης για την περιοχή αβεβαιότητας των εκτιμήσεων των ορόσημων.

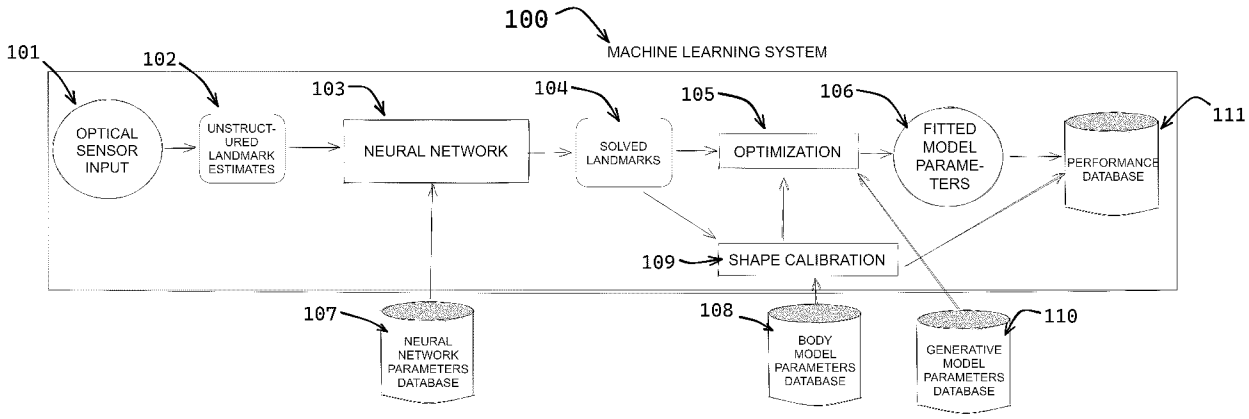

Figure ΣΧΗΜΑ 1

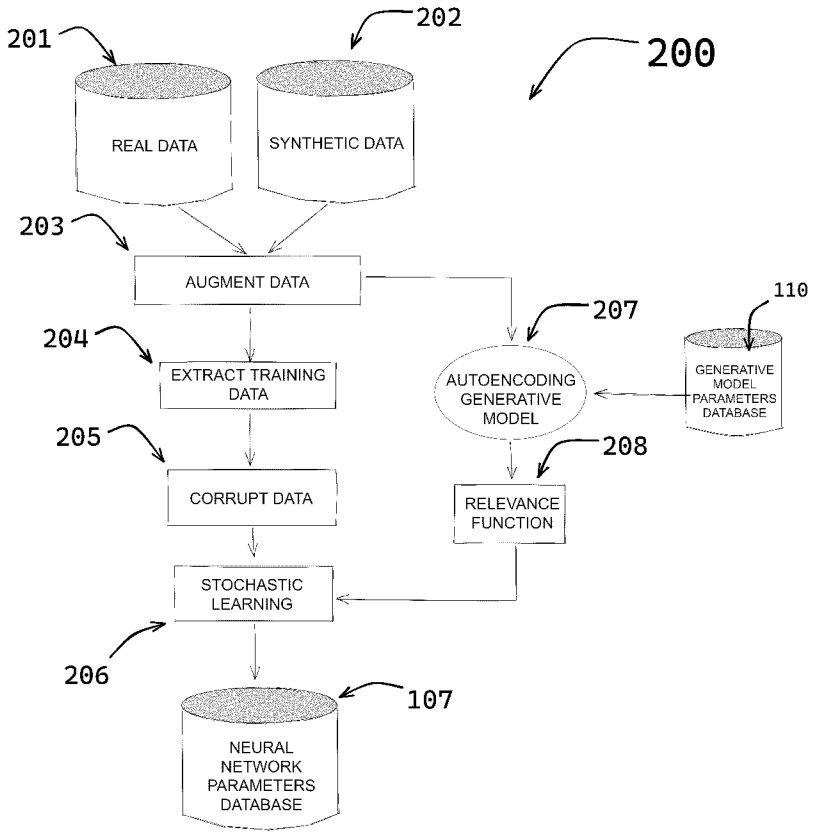

Figure ΣΧΗΜΑ **2**

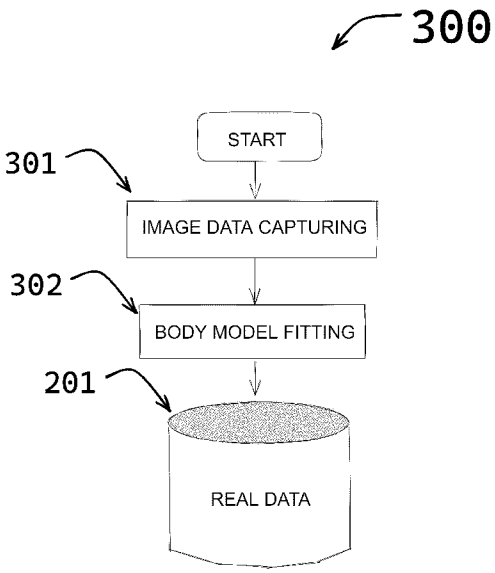

Figure ΣΧΗΜΑ **3**

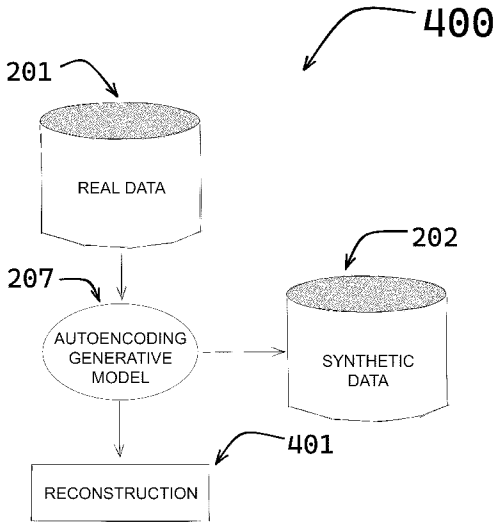

Figure ΣΧΗΜΑ 4

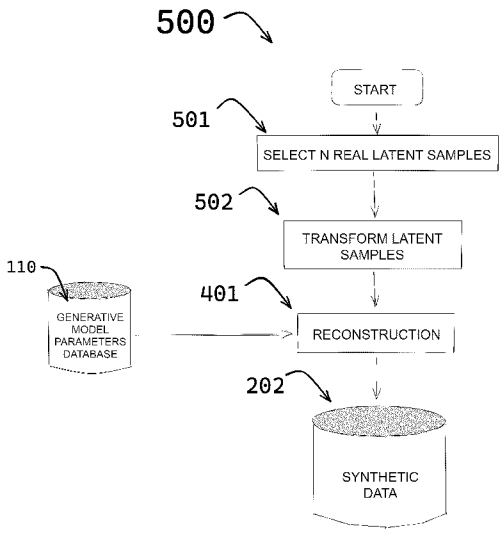

Figure ΣΧΗΜΑ 5

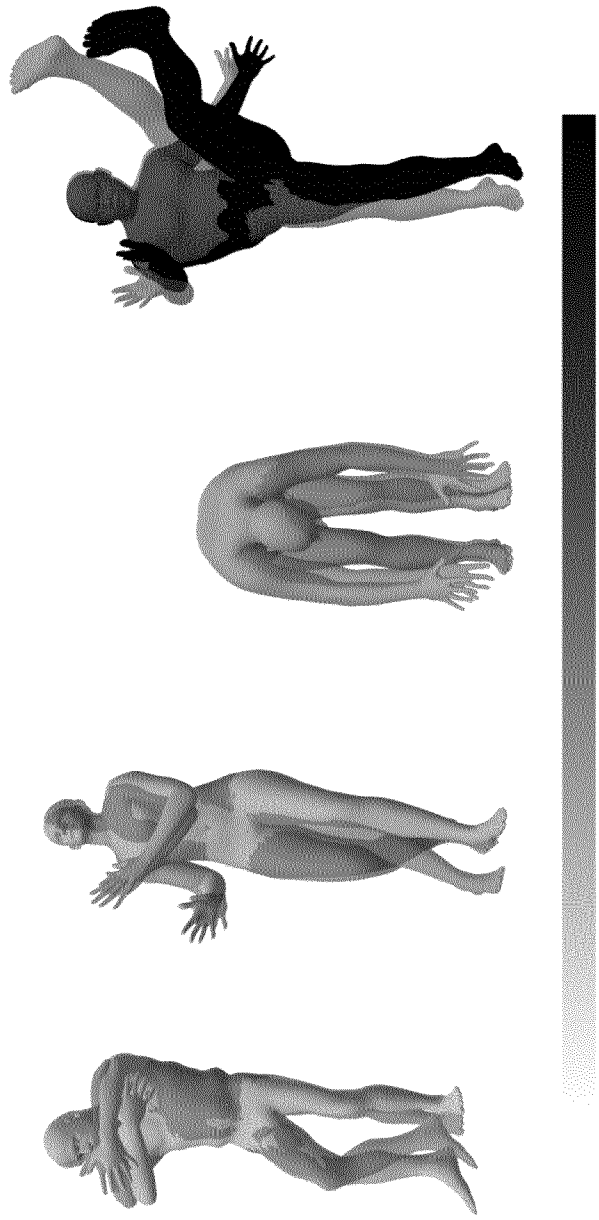

Figure **ΣΧΗΜΑ 6**

6/8

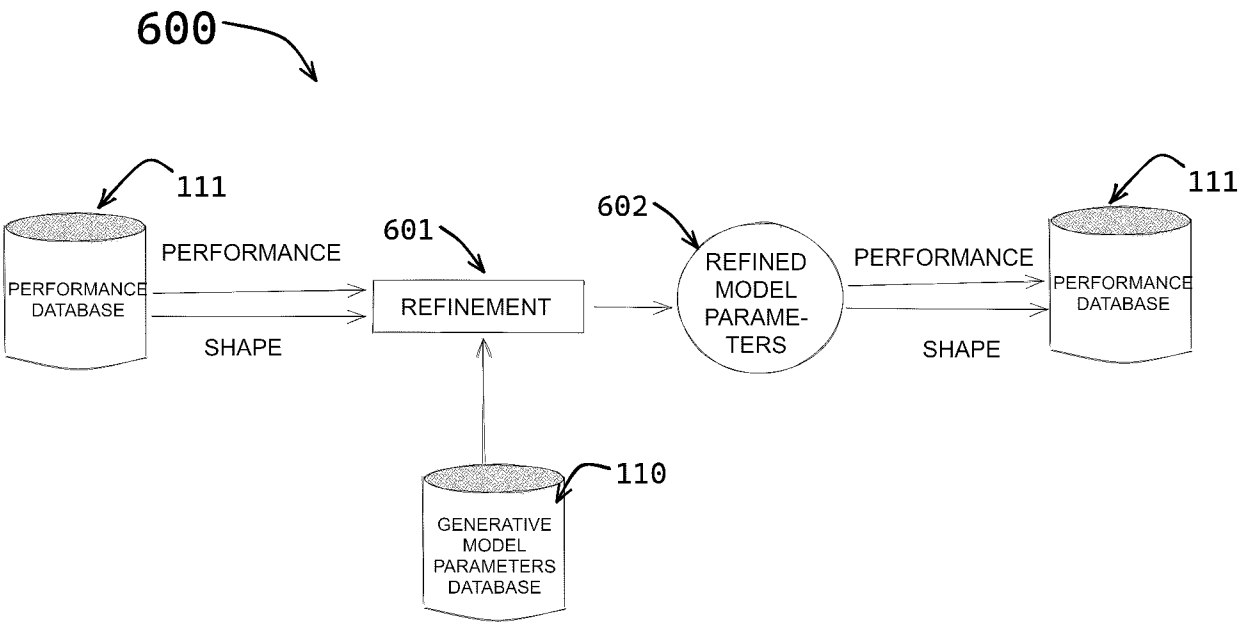

Figure ΣΧΗΜΑ 7

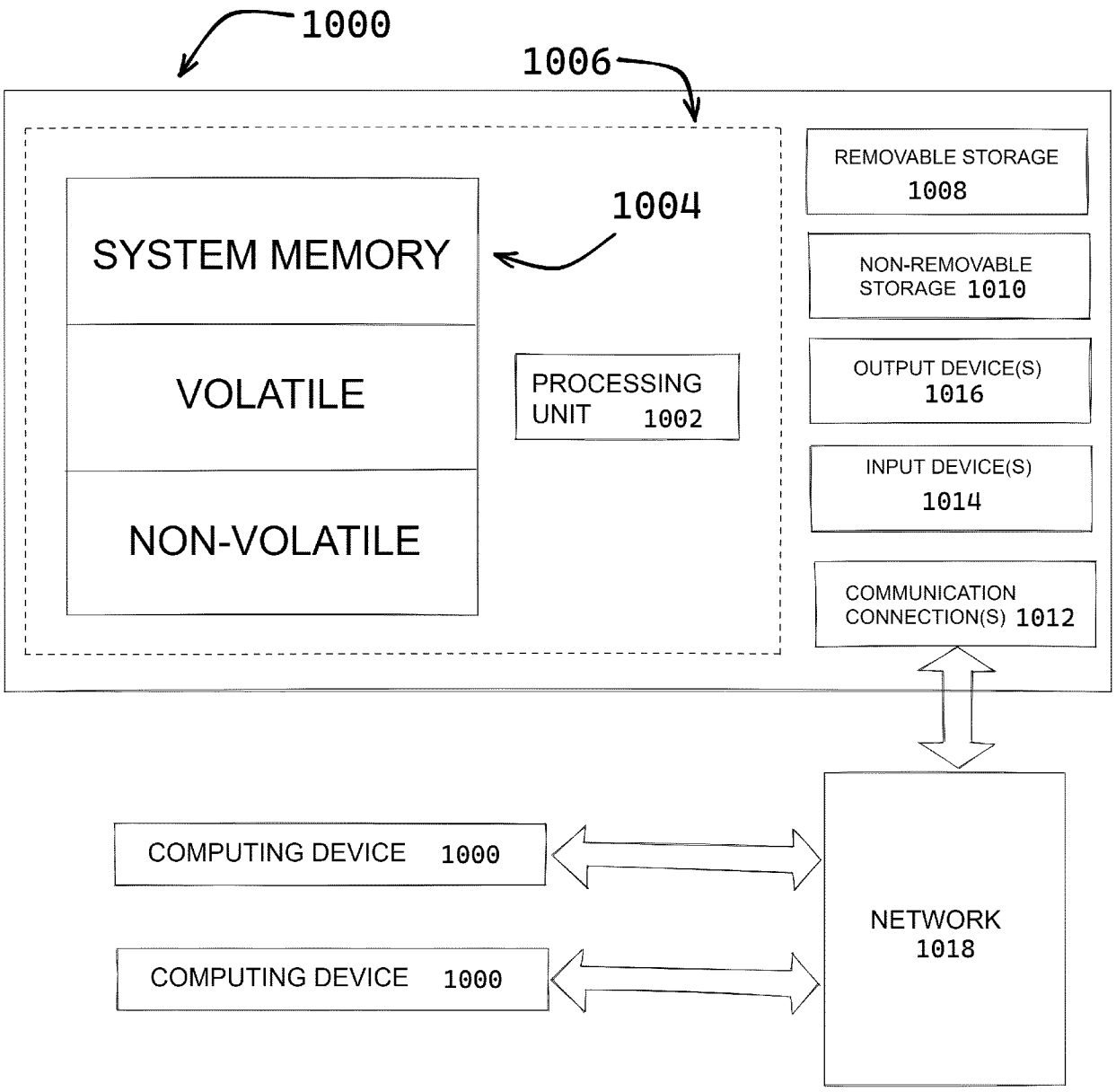

Figure ΣΧΗΜΑ 8