

Dated: December 10, 2025

Respectfully submitted,

By: /s/ Roger Fulghum
Roger Fulghum
Lead Attorney
Texas State Bar No. 00790724
BAKER BOTTS L.L.P.
910 Louisiana St
Houston, TX 77002
Telephone: 713-229-1707
Facsimile: 713-229-2707
Email:
roger.fulghum@bakerbotts.com

Mark Speegle
Texas State Bar No. 24117198
BAKER BOTTS L.L.P.
401 South First Street, Suite 1300
Austin, Texas 78704
Telephone: 512-322-2536
Facsimile: 512-322-3636
Email:
mark.speegle@bakerbotts.com

Helen Trac
California Bar No. 285824
Tesla, Inc.
1501 Page Mill, Building 6
Palo Alto, CA 94304
Telephone: 628-218-1429
Email: htrac@tesla.com

Ashraf Fawzy
DC Bar No. 989132
Telephone: (202) 905-9221
Email: afawzy@tesla.com
Paul Margulies
DC Bar No. 1000297
Telephone: (202) 695-5388
Email: pmargulies@tesla.com
Tesla, Inc.
800 Connecticut Ave. NW
Washington, DC 20006

Counsel for Tesla, Inc.

CERTIFICATE OF SERVICE

I hereby certify that all counsel of record are being served this 10th day of December 2025, by e-mail at the addresses indicated below:

Andrea Leigh Fair
Miller Fair Henry PLLC
1507 Bill Owens Parkway
Longview, TX 75604
andrea@millerfairhenry.com

Charles Austin Ginnings
Taryn Trusty
Timothy E Grochocinski
Nelson Bumgardner Conroy PC
745 McClintock Road, Suite 340
Burr Ridge, IL 60527
austin@nelbum.com
taryn@nelbum.com
tim@nelbum.com

Janson Westmoreland
John Paul Murphy
Nelson Bumgardner Conroy PC
3131 West 7th Street, Suite 300
Fort Worth, TX 76107
janson@nelbum.com
murphy@nelbum.com

Jonathan Hart Rastegar
Patrick Joseph Conroy
Nelson Bumgardner Conroy PC - Dallas
2727 N. Harwood Street, Suite 250
Dallas, TX 75201
JON@nelbum.com
pat@nelbum.com

/s/ Roger Fulghum
Roger Fulghum

UNITED STATES DISTRICT COURT

for the

Eastern District of Texas

PERCEPTIVE AUTOMATA LLC

Plaintiff

v.
TESLA, INC.

Defendant

Civil Action No. 2:25-cv-00742-JRG

SUBPOENA TO TESTIFY AT A DEPOSITION IN A CIVIL ACTION

To: Mercedes-Benz USA, LLC c/o CT Corporation System, 1999 Bryan Street, Suite 900, Dallas, TX 75201
Corporate HQ: One Mercedes-Benz Drive, Sandy Spring, GA 30328

(Name of person to whom this subpoena is directed)

☒ **Testimony:** YOU ARE COMMANDED to appear at the time, date, and place set forth below to testify at a deposition to be taken in this civil action. If you are an organization, you must promptly confer in good faith with the party serving this subpoena about the following matters, or those set forth in an attachment, and you must designate one or more officers, directors, or managing agents, or designate other persons who consent to testify on your behalf about these matters:

Place: Veritex 20 Mansell Court, Suite 300, Roswell, GA 30076	Date and Time: 01/14/2026 9:00 am
--	--------------------------------------

The deposition will be recorded by this method: Stenographic and videotaped

☒ **Production:** You, or your representatives, must also bring with you to the deposition the following documents, electronically stored information, or objects, and must permit inspection, copying, testing, or sampling of the material:

See Attachment A

The following provisions of Fed. R. Civ. P. 45 are attached – Rule 45(c), relating to the place of compliance; Rule 45(d), relating to your protection as a person subject to a subpoena; and Rule 45(e) and (g), relating to your duty to respond to this subpoena and the potential consequences of not doing so.

Date: 12/10/2025

CLERK OF COURT

OR

/s/ Roger Fulghum

Signature of Clerk or Deputy Clerk

Attorney's signature

The name, address, e-mail address, and telephone number of the attorney representing *(name of party)* Defendant, Tesla, Inc., who issues or requests this subpoena, are:

Roger Fulghum, Baker Botts L.L.P., 910 Louisiana, Houston, TX 77002; 713-229-1707; roger.fulghum@bakerbotts.com

Notice to the person who issues or requests this subpoena

If this subpoena commands the production of documents, electronically stored information, or tangible things before trial, a notice and a copy of the subpoena must be served on each party in this case before it is served on the person to whom it is directed. Fed. R. Civ. P. 45(a)(4).

Perceptive Exhibit 2045
IPR2025-01576

Civil Action No. 2:25-cv-00742-JRG

PROOF OF SERVICE

(This section should not be filed with the court unless required by Fed. R. Civ. P. 45.)

I received this subpoena for *(name of individual and title, if any)* _____
on *(date)* _____ .

☐ I served the subpoena by delivering a copy to the named individual as follows: _____

_____ on *(date)* _____ ; or

☐ I returned the subpoena unexecuted because: _____
_____ .

Unless the subpoena was issued on behalf of the United States, or one of its officers or agents, I have also
tendered to the witness the fees for one day's attendance, and the mileage allowed by law, in the amount of
\$ _____ .

My fees are \$ _____ for travel and \$ _____ for services, for a total of \$ 0.00 .

I declare under penalty of perjury that this information is true.

Date: _____

Server's signature

Printed name and title

Server's address

Additional information regarding attempted service, etc.:

Federal Rule of Civil Procedure 45 (c), (d), (e), and (g) (Effective 12/1/13)

(c) Place of Compliance.

(1) For a Trial, Hearing, or Deposition. A subpoena may command a person to attend a trial, hearing, or deposition only as follows:

- (A) within 100 miles of where the person resides, is employed, or regularly transacts business in person; or
- (B) within the state where the person resides, is employed, or regularly transacts business in person, if the person
 - (i) is a party or a party's officer; or
 - (ii) is commanded to attend a trial and would not incur substantial expense.

(2) For Other Discovery. A subpoena may command:

- (A) production of documents, electronically stored information, or tangible things at a place within 100 miles of where the person resides, is employed, or regularly transacts business in person; and
- (B) inspection of premises at the premises to be inspected.

(d) Protecting a Person Subject to a Subpoena; Enforcement.

(1) Avoiding Undue Burden or Expense; Sanctions. A party or attorney responsible for issuing and serving a subpoena must take reasonable steps to avoid imposing undue burden or expense on a person subject to the subpoena. The court for the district where compliance is required must enforce this duty and impose an appropriate sanction—which may include lost earnings and reasonable attorney's fees—on a party or attorney who fails to comply.

(2) Command to Produce Materials or Permit Inspection.

(A) *Appearance Not Required.* A person commanded to produce documents, electronically stored information, or tangible things, or to permit the inspection of premises, need not appear in person at the place of production or inspection unless also commanded to appear for a deposition, hearing, or trial.

(B) *Objections.* A person commanded to produce documents or tangible things or to permit inspection may serve on the party or attorney designated in the subpoena a written objection to inspecting, copying, testing, or sampling any or all of the materials or to inspecting the premises—or to producing electronically stored information in the form or forms requested. The objection must be served before the earlier of the time specified for compliance or 14 days after the subpoena is served. If an objection is made, the following rules apply:

- (i) At any time, on notice to the commanded person, the serving party may move the court for the district where compliance is required for an order compelling production or inspection.
- (ii) These acts may be required only as directed in the order, and the order must protect a person who is neither a party nor a party's officer from significant expense resulting from compliance.

(3) Quashing or Modifying a Subpoena.

(A) *When Required.* On timely motion, the court for the district where compliance is required must quash or modify a subpoena that:

- (i) fails to allow a reasonable time to comply;
 - (ii) requires a person to comply beyond the geographical limits specified in Rule 45(c);
 - (iii) requires disclosure of privileged or other protected matter, if no exception or waiver applies; or
 - (iv) subjects a person to undue burden.
- (B) *When Permitted.* To protect a person subject to or affected by a subpoena, the court for the district where compliance is required may, on motion, quash or modify the subpoena if it requires:

(i) disclosing a trade secret or other confidential research, development, or commercial information; or

(ii) disclosing an unretained expert's opinion or information that does not describe specific occurrences in dispute and results from the expert's study that was not requested by a party.

(C) *Specifying Conditions as an Alternative.* In the circumstances described in Rule 45(d)(3)(B), the court may, instead of quashing or modifying a subpoena, order appearance or production under specified conditions if the serving party:

- (i) shows a substantial need for the testimony or material that cannot be otherwise met without undue hardship; and
- (ii) ensures that the subpoenaed person will be reasonably compensated.

(e) Duties in Responding to a Subpoena.

(1) Producing Documents or Electronically Stored Information. These procedures apply to producing documents or electronically stored information:

(A) *Documents.* A person responding to a subpoena to produce documents must produce them as they are kept in the ordinary course of business or must organize and label them to correspond to the categories in the demand.

(B) *Form for Producing Electronically Stored Information Not Specified.* If a subpoena does not specify a form for producing electronically stored information, the person responding must produce it in a form or forms in which it is ordinarily maintained or in a reasonably usable form or forms.

(C) *Electronically Stored Information Produced in Only One Form.* The person responding need not produce the same electronically stored information in more than one form.

(D) *Inaccessible Electronically Stored Information.* The person responding need not provide discovery of electronically stored information from sources that the person identifies as not reasonably accessible because of undue burden or cost. On motion to compel discovery or for a protective order, the person responding must show that the information is not reasonably accessible because of undue burden or cost. If that showing is made, the court may nonetheless order discovery from such sources if the requesting party shows good cause, considering the limitations of Rule 26(b)(2)(C). The court may specify conditions for the discovery.

(2) Claiming Privilege or Protection.

(A) *Information Withheld.* A person withholding subpoenaed information under a claim that it is privileged or subject to protection as trial-preparation material must:

- (i) expressly make the claim; and
- (ii) describe the nature of the withheld documents, communications, or tangible things in a manner that, without revealing information itself privileged or protected, will enable the parties to assess the claim.

(B) *Information Produced.* If information produced in response to a subpoena is subject to a claim of privilege or of protection as trial-preparation material, the person making the claim may notify any party that received the information of the claim and the basis for it. After being notified, a party must promptly return, sequester, or destroy the specified information and any copies it has; must not use or disclose the information until the claim is resolved; must take reasonable steps to retrieve the information if the party disclosed it before being notified; and may promptly present the information under seal to the court for the district where compliance is required for a determination of the claim. The person who produced the information must preserve the information until the claim is resolved.

(g) Contempt.

The court for the district where compliance is required—and also, after a motion is transferred, the issuing court—may hold in contempt a person who, having been served, fails without adequate excuse to obey the subpoena or an order related to it.

Attachment A

ATTACHMENT A

Pursuant to Rule 45 of the Federal Rules of Civil Procedure, Tesla is seeking the production of the documents requested below for inspection and copying. The due date for production set in this subpoena is January 14, 2026. Tesla is also seeking deposition testimony from Mercedes on the documents and topics set out below. The time and date for the deposition is January 14, 2026 at 9:00 am. Tesla is willing to negotiate with Mercedes an alternative date and time for both the production of documents, and a deposition. Materials produced in this litigation will be governed by the Protective Order entered in this case, dated November 21, 2025. *See* Attachment B.

DEFINITIONS

1. The terms “You”, “Your”, and “Mercedes” mean Mercedes-Benz USA, LLC, and each of that entity’s parents, subsidiaries, divisions, affiliates, predecessors, assigns, successors, and acquired assets or business units, and any of its present or former officers, directors, trustees, employees, agents, representatives, attorneys, patent agents, and all other persons acting on its behalf.

2. The term “Mercedes Prior Art Vehicles” means the any vehicle implementing automatic emergency breaking or other autonomous features (including, but not limited to Mercedes-Benz S-, E-, and C-Class) that was designed, developed, tested, operated, offered for sale, used in commerce, or otherwise used by Mercedes, whether in the United States or another country, models before February 6, 2019. *See* Attachments C, D, E.

INSTRUCTIONS

1. Tesla is seeking documents and information related to prior art for certain patents that were originally issued to Perceptive Automata, Inc. and later assigned to the plaintiff in this lawsuit, Perceptive Automata, LLC. Tesla is willing to meet and confer and answer any questions

concerning the material sought in this subpoena. Please reach out to Roger Fulghum (roger.fulghum@bakerbotts.com or mobile 281-217-5031) or Mark Speegle (mark.speegle@bakerbotts.com or mobile 254-749-9120) with any questions concerning this subpoena.

REQUESTS FOR PRODUCTION

1. Documents sufficient to identify all Mercedes Prior Art Vehicles that were in public use, publicly demonstrated, offered for sale, used in commerce, or otherwise available to the public before each of the following dates: February 6, 2019; January 30, 2019; and July 5, 2017.

2. For each Mercedes Prior Art Vehicle, documents sufficient to establish when such Mercedes Prior Art Vehicle was first in public use, publicly demonstrated, offered for sale, used in commerce, or otherwise available to the public.

3. Documents sufficient to show the operation of the automatic emergency braking system for each Mercedes Prior Art Vehicle, including the use of “stereo-based pedestrian safety system[s].” *See* Attachment C at 1445; *see also* Attachment D at 174; Attachment E at 494.

4. Documents sufficient to show the operation of (including inputs and outputs to) all machine learning models, algorithms, routines, methods, functions, or structures in relation to automatic emergency braking systems in each Mercedes Prior Art Vehicle, including but not limited to pedestrian dynamics models. *See, e.g.*, Attachment C at 1446-47, Attachment D at 174, Attachment E at 494.

5. Documents sufficient to show the hardware implemented in automatic emergency braking systems in each Mercedes Prior Art Vehicle, including but not limited to (i) cameras, lidar, radar, or other visual sensors and (ii) vehicle control systems.

6. Documents sufficient to show how Mercedes collected, stored, labeled, and/or generated training data used to train machine learning models implemented in or executed by automatic emergency braking systems in Mercedes Prior Art Vehicles. *See, e.g.*, Attachment C at 1448-49, Attachment D at 179-80, Attachment E at 497-98.

DEPOSITION TOPICS

1. All documents produced in response to this subpoena, including the authenticity thereof and whether each such documents is a record of an act, event, condition, opinion, or diagnosis that (i) was made at or near the time by — or from information transmitted by — someone with knowledge; (ii) was kept in the course of a regularly conducted activity of a business, organization, occupation, or calling, whether or not for profit; and (iii) was made as a regular practice of that activity.

2. The identity of all Mercedes Prior Art Vehicles that were in public use, publicly demonstrated, offered for sale, used in commerce, or otherwise available to the public before each of the following dates: February 6, 2019; January 30, 2019; and July 5, 2017.

3. The operation of the automatic emergency braking system for each Mercedes Prior Art Vehicles, including that of any components or software thereof, in relation to or within any computers, servers, cloud instances, or the like.

4. The operation of (including inputs and outputs to) all machine learning models, algorithms, routines, methods, functions, or structures in relation to or within each Mercedes Prior Art Vehicle.

5. Attachment C, including the contents and the circumstances around its authorship and publication.

6. Attachment D, including the contents and the circumstances around its authorship and publication.

7. Attachment E, including the contents and the circumstances around its authorship and publication.

8. The operation and implementation of any “stereo-based pedestrian safety system[s]” in relation to emergency braking systems in the Mercedes Prior Art Vehicles. *See* Attachment C at 1445; *see also* Attachment D at 174; Attachment E at 494.

9. The hardware implemented in automatic emergency braking systems in each Mercedes Prior Art Vehicle, including but not limited to (i) cameras, lidar, radar, or other visual sensors and (ii) vehicle control systems.

10. The collection, storage, labeling, and/or generating of training data used to train machine learning models executed by or implemented in automatic emergency braking systems in the Mercedes Prior Art Vehicles. *See, e.g.*, Attachment C at 1448-49, Attachment D at 179-80, Attachment E at 497-98.

Attachment B

**IN THE UNITED STATES DISTRICT COURT
FOR THE EASTERN DISTRICT OF TEXAS
MARSHALL DIVISION**

PERCEPTIVE AUTOMATA LLC,	§	
	§	
<i>Plaintiff,</i>	§	
	§	Case No. 2:25-cv-00742-JRG
v.	§	
	§	JURY TRIAL DEMANDED
TESLA, INC.	§	
	§	
<i>Defendant.</i>	§	

PROTECTIVE ORDER

WHEREAS, Plaintiff Perceptive Automata LLC and Defendant Tesla, Inc., hereafter referred to as “the Parties,” believe that certain information that is or will be encompassed by discovery demands by the Parties involves the production or disclosure of trade secrets, confidential business information, or other proprietary information;

WHEREAS, the Parties seek a protective order limiting disclosure thereof in accordance with Federal Rule of Civil Procedure 26(c):

THEREFORE, it is hereby stipulated among the Parties and ORDERED that:

1. Each Party may designate as confidential for protection under this Order, in whole or in part, any document, information or material that constitutes or includes, in whole or in part, confidential or proprietary information or trade secrets of the Party or a Third Party to whom the Party reasonably believes it owes an obligation of confidentiality with respect to such document, information or material (“Protected Material”). Protected Material shall be designated by the Party producing it by affixing a legend or stamp on such document, information or material as follows: “CONFIDENTIAL.” The word “CONFIDENTIAL” shall be placed clearly on each page of the Protected Material (except deposition and hearing

transcripts) for which such protection is sought. For deposition and hearing transcripts, the word “CONFIDENTIAL” shall be placed on the cover page of the transcript (if not already present on the cover page of the transcript when received from the court reporter) by each attorney receiving a copy of the transcript after that attorney receives notice of the designation of some or all of that transcript as “CONFIDENTIAL.”

2. Any document produced under Patent Rules 2-2, 3-2, and/or 3-4 before issuance of this Order with the designation “Confidential” or “Confidential - Outside Attorneys’ Eyes Only” shall receive the same treatment as if designated “RESTRICTED - ATTORNEYS’ EYES ONLY” under this Order, unless and until such document is redesignated to have a different classification under this Order.
3. With respect to documents, information or material designated “CONFIDENTIAL,” “RESTRICTED - ATTORNEYS’ EYES ONLY,” or “RESTRICTED CONFIDENTIAL - SOURCE CODE” (“DESIGNATED MATERIAL”),¹ subject to the provisions herein and unless otherwise stated, this Order governs, without limitation: (a) all documents, electronically stored information, and/or things as defined by the Federal Rules of Civil Procedure; (b) all pretrial, hearing or deposition testimony, or documents marked as exhibits or for identification in depositions and hearings; (c) pretrial pleadings, exhibits to pleadings and other court filings; (d) affidavits; and (e) stipulations. All copies, reproductions, extracts, digests and complete or partial summaries prepared from any DESIGNATED MATERIALS shall also be considered DESIGNATED MATERIAL and treated as such

¹ The term DESIGNATED MATERIAL is used throughout this Protective Order to refer to the class of materials designated as “CONFIDENTIAL,” “RESTRICTED - ATTORNEYS’ EYES ONLY,” or “RESTRICTED CONFIDENTIAL - SOURCE CODE,” both individually and collectively.

under this Order.

4. A designation of Protected Material (i.e., “CONFIDENTIAL,” “RESTRICTED - ATTORNEYS’ EYES ONLY,” or “RESTRICTED CONFIDENTIAL SOURCE CODE”) may be made at any time. Production of documents, information, or material that has not been designated as DESIGNATED MATERIAL, whether inadvertently or otherwise, shall not be deemed a waiver in whole or in part of a claim for confidential treatment. Any party that inadvertently or otherwise produces Protected Material without designating it as DESIGNATED MATERIAL may request destruction of that Protected Material by notifying the recipient(s), as soon as reasonably possible after the Producing Party becomes aware of the inadvertent or unintentional disclosure, and providing replacement Protected Material that is properly designated. The recipient(s) shall then destroy all copies of the inadvertently or unintentionally produced Protected Materials and any documents, information or material derived from or based thereon.
5. “CONFIDENTIAL” documents, information and material may be disclosed only to the following persons, except upon receipt of the prior written consent of the designating party, upon order of the Court, or as set forth in paragraph 12 herein:
 - (a) Outside Counsel of Record in this Action for the Parties;
 - (b) employees of such counsel assigned to and reasonably necessary to assist such counsel in the litigation of this Action;
 - (c) in-house counsel for the Parties who either have responsibility for making decisions dealing directly with the litigation, or who are assisting outside counsel in the litigation;
 - (d) up to and including three (3) designated representatives of each of the Parties to the extent reasonably necessary for the litigation of this Action, except that either party may in good faith request the other party’s consent to designate one or more additional representatives, the other party shall not unreasonably withhold such consent, and the requesting party may seek leave of Court to designate such

additional representative(s) if the requesting party believes the other party has unreasonably withheld such consent;

- (e) outside consultants or experts (*i.e.*, not existing employees or affiliates of a Party or an affiliate of a Party) retained for the purpose of this litigation, provided that: (1) such consultants or experts are not presently employed by the Parties hereto for purposes other than this Action; (2) before access is given, the consultant or expert has completed the Undertaking attached as Exhibit A hereto and the same is served upon the Producing Party with a current curriculum vitae of the consultant or expert, including a list of other cases from the prior five (5) years in which the individual has provided a report or testified (at trial or at deposition, at least ten (10) days before access to the Protected Material is to be given to that consultant or Undertaking to object to and notify the Receiving Party in writing that it objects to disclosure of Protected Material to the consultant or expert. The Parties agree to promptly confer and use good faith to resolve any such objection. If the Parties are unable to resolve any objection, the objecting Party may file a motion with the Court within fifteen (15) days of the notice, or within such other time as the Parties may agree, seeking a protective order with respect to the proposed disclosure. The objecting Party shall have the burden of proving the need for a protective order. No disclosure shall occur until all such objections are resolved by agreement or Court order;
 - (f) independent litigation support services, including persons working for or as court reporters, graphics or design services, jury or trial consulting services, and photocopy, document imaging, and database services retained by counsel and reasonably necessary to assist counsel with the litigation of this Action; and
 - (g) the Court and its personnel.
6. A Party shall designate documents, information or material as “CONFIDENTIAL” only upon a good faith belief that the documents, information or material contains confidential or proprietary information or trade secrets of the Party or a Third Party to whom the Party reasonably believes it owes an obligation of confidentiality with respect to such documents, information or material.
7. Documents, information or material produced pursuant to any discovery request in this Action, including but not limited to Protected Material designated as DESIGNATED MATERIAL, shall be used by the Parties only in the litigation of this Action and shall not be used for any other purpose. Any person or entity who obtains access to DESIGNATED

MATERIAL or the contents thereof pursuant to this Order shall not make any copies, duplicates, extracts, summaries or descriptions of such DESIGNATED MATERIAL or any portion thereof except as may be reasonably necessary in the litigation of this Action. Any such copies, duplicates, extracts, summaries or descriptions shall be classified DESIGNATED MATERIALS and subject to all of the terms and conditions of this Order.

8. To the extent a Producing Party believes that certain Protected Material qualifying to be designated CONFIDENTIAL is so sensitive that its dissemination deserves even further limitation, the Producing Party may designate such Protected Material “RESTRICTED - ATTORNEYS’ EYES ONLY,” or to the extent such Protected Material includes extremely sensitive DESIGNATED MATERIAL representing computer source code and/or live data (that is, data as it exists residing in a database or databases) as well as associated “Readme” files that are stored in the same directory as the source code in the ordinary course of business (collectively, “Source Code Material”), the Producing Party may designate such Protected Material as “RESTRICTED CONFIDENTIAL - SOURCE CODE.”
9. For Protected Material designated RESTRICTED - ATTORNEYS’ EYES ONLY, access to, and disclosure of, such Protected Material shall be limited to individuals listed in paragraphs 5(a-c) and (e-g); provided, however, that access by in-house counsel pursuant to paragraph 5(c) be limited to in-house counsel who exercise no competitive decision-making authority on behalf of the client.
10. For Protected Material designated RESTRICTED CONFIDENTIAL - SOURCE CODE, the following additional restrictions apply:
 - (a) Access to a Party’s Source Code Material shall be provided only on “stand-alone” computer(s) (that is, the computer may not be linked to any network, including a local area network (“LAN”), an intranet or the Internet) and on which all access ports have been disabled, as necessary and appropriate to prevent and protect

against any unauthorized copying, transmission, removal, or other transfer of any Source Code outside or away from the computer on which the Source Code is provided for inspection. Additionally, except as provided in paragraph 10(u) below, the stand-alone computer(s) may only be located at or near the offices of the Producing Party's Outside Counsel of Record or In-House Counsel of Record if the Party is not represented by outside counsel. The Producing Party may visually monitor the activities of the Receiving Party's representatives during any source code review, but only to ensure that there is no unauthorized recording, copying, or transmission of the source code;

- (b) No recordable media or recordable devices, including without limitation sound recorders, computers, cellular telephones, peripheral equipment, cameras, CDs, DVDs, or drives of any kind shall be permitted into the source code review room.
- (c) Parties shall ensure that appropriate software sufficient for viewing, and searching the Source Code Material is installed on the Secure Source Code Computers. The Receiving Party's counsel and/or experts may request that commercially available software for viewing, and searching source code be installed on the Secure Source Code Computer. The Receiving Party must provide the Producing Party with electronic media containing such software, or links to a public location from which such software may be downloaded, and any necessary licenses at least four (4) days in advance of the date upon which the Receiving Party wishes to have the software tools available for use on the Secure Source Code Computers. Examples of software may include SciTools Understand, Eclipse, Android Studio, Microsoft Visual Studio, Notepad++, UltraEdit, SlickEdit, Beyond Compare, CodeSuite from Software Analysis, PowerGrep, Mozilla Firefox, Adobe Acrobat, or other similar software that provides for viewing, and searching source code;
- (d) The Receiving Party shall make reasonable efforts to restrict its requests for such access to the stand-alone computer(s) to normal business hours, which for purposes of this paragraph shall be 9:00 a.m. through 6:00 p.m. The Parties agree to cooperate in good faith such that maintaining the Producing Party's Source Code Material at or near the offices of its Outside Counsel of Record, or In-House Counsel of Record if the Party is not represented by outside counsel, shall not unreasonably hinder the Receiving Party's ability to efficiently and effectively conduct the prosecution or defense of this Action;
- (e) The Producing Party shall provide the Receiving Party with information explaining how to start, log on to, and operate the stand-alone computer(s) in order to access the produced Source Code Material on the stand-alone computer(s);
- (f) The Producing Party will produce Source Code Material in computer searchable format on the stand-alone computer(s) as described above;
- (g) Access to Protected Material designated RESTRICTED CONFIDENTIAL -

SOURCE CODE shall be limited to Outside Counsel of Record and up to three (3) outside consultants or experts² (i.e., not existing employees or affiliates of a Party or an affiliate of a Party) retained for the purpose of this litigation and approved to access such Protected Materials pursuant to paragraph 5(e) above. A receiving Party may include excerpts of Source Code Material in a pleading, exhibit, expert report, discovery document, deposition transcript, other Court document, provided that the Source Code Documents are appropriately marked under this Order, restricted to those who are entitled to have access to them as specified herein, and, if filed with the Court, filed under seal in accordance with the Court's rules, procedures and orders;

- (h) To the extent portions of Source Code Material are quoted in a Source Code Exhibit, either (1) the entire Source Code Exhibit will be stamped and treated as RESTRICTED CONFIDENTIAL - SOURCE CODE or (2) those pages containing quoted Source Code Material will be separately stamped and treated as RESTRICTED CONFIDENTIAL - SOURCE CODE and, if filed with the Court, filed under seal in accordance with the Court's rules, procedures, and orders;
- (i) Except as set forth in paragraph 10(l) below, no electronic copies of Source Code Material shall be made without prior written consent of the producing Party, except as necessary to create documents which, pursuant to the Court's rules, procedures and order, must be filed or served electronically.
- (j) The Receiving Party shall not be permitted to make its own printouts and photocopies of Source Code Material. At the Receiving Party's request, the Producing Party shall provide up to three (3) copies of all printouts of Source Code Material, all of which shall be designated and clearly labeled "RESTRICTED CONFIDENTIAL SOURCE CODE";
- (k) If the Receiving Party's Outside Counsel of Record, consultants, or experts obtain printouts or photocopies of Source Code Material, the Receiving Party shall ensure that such outside counsel, consultants, or experts keep the printouts or photocopies in a secured locked area in the offices of such outside counsel, consultants, or expert. The Receiving Party may also temporarily keep the printouts or photocopies at: (i) the Court for any proceedings(s) relating to the Source Code Material, for the dates associated with the proceeding(s); (ii) the sites where any deposition(s) relating to the Source Code Material are taken, for the dates associated with the deposition(s); and (iii) any intermediate location reasonably necessary to transport the printouts or photocopies (e.g., a hotel prior to a Court proceeding or deposition); and
- (l) A Producing Party's Source Code Material may only be transported by the

² For the purposes of this paragraph, an outside consultant or expert is defined to include the outside consultant's or expert's direct reports and other support personnel, such that the disclosure to a consultant or expert who employs others within his or her firm to help in his or her analysis shall count as a disclosure to a single consultant or expert.

Receiving Party at the direction of a person authorized under paragraph 10(e) above to another person authorized under paragraph 10(e) above, on paper or removable encrypted electronic media (*e.g.*, a DVD, CD-ROM, or flash memory “stick”) via hand carry, Federal Express or other similarly reliable courier. Source Code Material may not be transported or transmitted electronically over a network of any kind, including a LAN, an intranet, or the Internet. Source Code Material may only be transported electronically for the purpose of Court proceeding(s) as set forth in paragraph 10(k) above and is at all times subject to the transport restrictions set forth herein. But, for those purposes only, the Source Code Materials may be loaded onto a stand-alone computer.

Notes

- (m) The Receiving Party’s outside counsel of record and/or authorized consultants and experts pursuant to paragraph 5(g) above shall be entitled to take handwritten notes relating to the Source Code Material including identifying the file name and line number of the code, but may not copy entire lines of code. Such notes shall be labeled “RESTRICTED CONFIDENTIAL - SOURCE CODE.” The Receiving Party shall be under no obligation to produce handwritten notes. No copies of all or any portion of the Source Code Material may leave the source code review room. The Receiving Party may take handwritten notes. All notes are protected by attorney client privilege. No other written or electronic record of the Source Code Material is permitted except as otherwise provided in this Protective Order.
- (n) Unless otherwise agreed in advance by the parties in writing, the Receiving Party’s authorized counsel and/or authorized consultants and experts pursuant to paragraph 5(g) above shall remove all notes, documents, and all other materials from the source code review room that may contain work product and/or attorney-client privileged information at the end of each day. Materials inadvertently left in the source code review room do not operate as a waiver of the attorney work product doctrine or any other applicable privilege and shall be returned to the owner promptly. The Producing Party shall not be responsible for any items left in the source code review room.

Printouts and Copies

- (o) The Receiving Party may request a printout of reasonable portions of the Source Code Material. The Receiving Party shall not request the printouts of Source Code Material in order to review blocks of Source Code Material in the first instance (*i.e.*, as an alternative to reviewing Source Code Material electronically on the source code computers) as the Parties acknowledge and agree that the purpose of the protections herein would be frustrated by printing portions of Source Code Material for review and analysis elsewhere.
- (p) In the event that the Producing Party objects that identified files are not reasonably necessary, the Producing Party shall make such objection known to the Receiving

Party within three (3) business days of the identification of any files by the Receiving Party. The Producing Party will make itself available within three (3) business days of making the objection for a meet and confer. If, after meeting and conferring, the Producing Party and the Receiving Party cannot resolve the objection, the Producing Party shall seek a judicial resolution of whether or not the identified Source Code Material in question is reasonably necessary to any case preparation activity. The parties recognize that the volume and format of source code at issue in the case is unknown. Nevertheless, the Receiving Party will make a good faith effort to not exceed the agreed limit of fifteen (15) consecutive pages and an aggregate of more than six-hundred fifty (650) pages of printouts of Source Code Material. Requests to print additional pages beyond these limits will not be unreasonably withheld by the Producing Party. In the event Producing Party refuses a request to print additional pages, the Receiving Party will bear the burden of demonstrating the additional printouts are relevant to the needs of the case. At the request of the Receiving Party, in the absence of any objection the Producing Party shall provide within five (5) days of such request, a printout of the Source Code Material requested by the Receiving Party. The Producing Party shall print and label such Source Code Material with the complete path name, if applicable, and append production numbers and the designation "RESTRICTED CONFIDENTIAL - SOURCE CODE."

- (q) Unless agreed by the parties, images or copies of Source Code Material shall not be included in correspondence between the Parties (references to production numbers shall be used instead). The parties will make a good faith effort to omit or minimize the use of images or copies of Source Code Material in pleadings, exhibits, and other papers.
- (r) The Receiving Party shall maintain a record of any individual who has inspected any portion of the Source Code Material in electronic or paper form. The log shall include the number of paper copies, names of the recipients of paper copies, and locations where the paper copies are stored. Upon three (3) business day's advance notice to the Receiving Party by the Producing Party, the Receiving Party shall provide a copy of this log to the Producing Party. The Parties agree the purpose of the log is to evidence compliance with this Order and that the log may not be used for examination of a witness. All copies of Source Code Material shall be printed with an "anti-copying" watermark that reveals whether such printouts have been photocopied.
- (s) Consistent with and subject to paragraph 10(k) above, Source Code Material must be kept at all times in a secured, locked and hardened container (e.g., a safe or lockbox) that prevents duplication of or unauthorized access to the Source Code Material.

Depositions

- (t) For all depositions, Receiving Party may request that the witness have access to a

Secure Source Code Computer during the deposition three (3) days prior to the deposition. Upon such request, Producing Party will ensure a Secure Source Code Computer is available with all code during examination with reasonable search tools so that the witness can be questioned about the code. If a Source Code Material document has been marked up or altered in any way by the deponent, the Producing Party shall copy the exhibit in its marked form and provide one copy to outside counsel for the Receiving Party. Copies of Source Code Material that are marked as deposition exhibits shall not be provided to the Court Reporter or attached to deposition transcripts; rather, the deposition record will identify the exhibit by its production numbers.

11. Any attorney representing a Party, whether in-house or outside counsel, and any person associated with a Party that receives the other Party's Protected Material that is designated RESTRICTED - ATTORNEYS' EYES ONLY and/or RESTRICTED CONFIDENTIAL - SOURCE CODE (collectively "HIGHLY SENSITIVE MATERIAL"), who obtains, receives, has access to, or otherwise learns, in whole or in part, HIGHLY SENSITIVE MATERIAL under this Order shall not prepare, prosecute, supervise, or assist in the preparation or prosecution of any patent application pertaining to the field of the invention of the patents-in-suit on behalf of the Receiving Party or its acquirer, successor, predecessor, or other affiliate during the pendency of this Action and for one year after its conclusion, including any appeals. This provision does not prohibit a Party's counsel of record or experts in this litigation from participating in or representing it in reexamination proceedings, Post-Grant Review proceedings, Inter Partes Review proceedings, or Covered Business Method Review proceedings involving any patent, including the patents-in-suit, provided they do not advise on, consult on, prepare, draft, edit, or otherwise provide input into any claims, or amendment to specifications or claims in those proceedings. To ensure compliance with the purpose of this provision, each Party shall create an "Ethical Wall" between those persons with access to HIGHLY SENSITIVE MATERIAL and any individuals who, on behalf of the Party or its acquirer, successor,

predecessor, or other affiliate, prepare, prosecute, supervise or assist in the preparation or prosecution of any patent application pertaining to the field of invention of the patent-in-suit.

12. Unless otherwise provided by this Order or provided by the Court, confidentiality obligations imposed by this Order shall remain in effect following the conclusion of this lawsuit.
13. Nothing in this Order shall require production of documents, information or other material that a Party contends is protected from disclosure by the attorney-client privilege, the work product doctrine, or other privilege, doctrine, or immunity. If documents, information or other material subject to a claim of attorney-client privilege, work product doctrine, or other privilege, doctrine, or immunity is inadvertently or unintentionally produced, such production shall in no way prejudice or otherwise constitute a waiver of, or estoppel as to, any such privilege, doctrine, or immunity. Any Party that inadvertently or unintentionally produces documents, information or other material it reasonably believes are protected under the attorney-client privilege, work product doctrine, or other privilege, doctrine, or immunity may obtain the return of such documents, information or other material by promptly notifying the recipient(s) and providing a privilege log for the inadvertently or unintentionally produced documents, information or other material. The recipient(s) shall gather and return all copies of such documents, information or other material to the Producing Party, except for any pages containing privileged or otherwise protected markings by the recipient(s), which pages shall instead be destroyed and certified as such to the Producing Party. If a Receiving Party believes that a produced document or other information may reasonably be subject to a claim of privilege, then the Receiving Party

shall promptly inform the Producing Party of the beginning BATES number of the document or, if no BATES number is available, shall otherwise inform the Producing Party of the information.

14. There shall be no disclosure of any DESIGNATED MATERIAL by any person authorized to have access thereto to any person who is not authorized for such access under this Order. The Parties are hereby ORDERED to safeguard all such documents, information and material to protect against disclosure to any unauthorized persons or entities.
15. Nothing contained herein shall be construed to prejudice any Party's right to use any DESIGNATED MATERIAL in taking testimony at any deposition or hearing provided that the DESIGNATED MATERIAL is only disclosed to a person(s) who is: (i) eligible to have access to the DESIGNATED MATERIAL by virtue of his or her employment with the designating party, (ii) identified in the DESIGNATED MATERIAL as an author, addressee, or copy recipient of such information, (iii) although not identified as an author, addressee, or copy recipient of such DESIGNATED MATERIAL, has, in the ordinary course of business, seen such DESIGNATED MATERIAL, (iv) a current or former officer, director or employee of the Producing Party or a current or former officer, director or employee of a company affiliated with the Producing Party; (v) counsel for a Party, including outside counsel and in-house counsel (subject to paragraph 9 of this Order); (vi) an independent contractor, consultant, and/or expert retained for the purpose of this litigation; (vii) court reporters and videographers; (viii) the Court; or (ix) other persons entitled hereunder to access to DESIGNATED MATERIAL. DESIGNATED MATERIAL shall not be disclosed to any other persons unless prior authorization is obtained from counsel representing the Producing Party or from the Court.

16. Parties may, at the deposition or hearing or within thirty (30) days after receipt of a deposition or hearing transcript, designate the deposition or hearing transcript or any portion thereof as “CONFIDENTIAL,” “RESTRICTED - ATTORNEY’ EYES ONLY,” or “RESTRICTED CONFIDENTIAL - SOURCE CODE” pursuant to this Order. Access to the deposition or hearing transcript so designated shall be limited in accordance with the terms of this Order. Until expiration of the 30-day period, the entire deposition or hearing transcript shall be treated as confidential.
17. Any DESIGNATED MATERIAL that is filed with the Court shall be filed under seal and shall remain under seal until further order of the Court. The filing party shall be responsible for informing the Clerk of the Court that the filing should be sealed and for placing the legend “FILED UNDER SEAL PURSUANT TO PROTECTIVE ORDER” above the caption and conspicuously on each page of the filing. Exhibits to a filing shall conform to the labeling requirements set forth in this Order. If a pretrial pleading filed with the Court, or an exhibit thereto, discloses or relies on confidential documents, information or material, such confidential portions shall be redacted to the extent necessary and the pleading or exhibit filed publicly with the Court.
18. The Order applies to pretrial discovery. Nothing in this Order shall be deemed to prevent the Parties from introducing any DESIGNATED MATERIAL into evidence at the trial of this Action, or from using any information contained in DESIGNATED MATERIAL at the trial of this Action, subject to any pretrial order issued by this Court.
19. A Party may request in writing to the other Party that the designation given to any DESIGNATED MATERIAL be modified or withdrawn. If the designating Party does not agree to redesignation within ten (10) days of receipt of the written request, the requesting

Party may apply to the Court for relief. Upon any such application to the Court, the burden shall be on the designating Party to show why its classification is proper. Such application shall be treated procedurally as a motion to compel pursuant to Federal Rules of Civil Procedure 37, subject to the Rule's provisions relating to sanctions. In making such application, the requirements of the Federal Rules of Civil Procedure and the Local Rules of the Court shall be met. Pending the Court's determination of the application, the designation of the designating Party shall be maintained.

20. Each outside consultant or expert to whom DESIGNATED MATERIAL is disclosed in accordance with the terms of this Order shall be advised by counsel of the terms of this Order, shall be informed that he or she is subject to the terms and conditions of this Order, and shall sign an acknowledgment that he or she has received a copy of, has read, and has agreed to be bound by this Order. A copy of the acknowledgment form is attached as Appendix A.
21. To the extent that any discovery is taken of persons who are not Parties to this Action ("Third Parties") and in the event that such Third Parties contended the discovery sought involves trade secrets, confidential business information, or other proprietary information, then such Third Parties may agree to be bound by this Order.
22. To the extent that discovery or testimony is taken of Third Parties, the Third Parties may designate as "CONFIDENTIAL" or "RESTRICTED - ATTORNEYS' EYES ONLY" any documents, information or other material, in whole or in part, given by such Third Parties. The Third Parties shall have ten (10) days after production of such documents, information or other materials to make such a designation. Until that time period lapses or until such a designation has been made, whichever occurs sooner, all documents, information or other

material so produced or given shall be treated as “CONFIDENTIAL” in accordance with this Order.

23. Within sixty (60) days of final termination of this Action, including any appeals, all DESIGNATED MATERIAL, including all copies, duplicates, abstracts, indexes, summaries, descriptions, and excerpts or extracts thereof (excluding excerpts or extracts incorporated into any privileged memoranda of the Parties and materials which have been admitted into evidence in this Action), shall at the Producing Party’s election either be returned to the Producing Party or be destroyed. No copies shall be retained by a Receiving Party or other individuals retained to assist them in this litigation, except counsel may retain copies of pleadings in their files and the terms of this Protective Order shall continue to apply to such copies. The language contained in this paragraph shall not preclude counsel from retaining documents constituting attorney work product as defined by applicable law, or from retaining one copy of all such work product documents as part of a permanent litigation file that is otherwise subject to the confidentiality restrictions set forth herein. The Receiving Party shall verify the return or destruction by affidavit furnished to the Producing Party, upon the Producing Party’s request. Nothing in this order requires a Party to destroy any information it is required by law to retain. The Parties shall not be required to delete information that may reside on electronic disaster recovery systems that are overwritten in the normal course of business, or information that may reside in electronic files which are not reasonably accessible. However, parties and their Counsel shall not retrieve, access, nor use any Protected Material from said electronic disaster recovery systems or other not-reasonably accessible data sources after the conclusion of the matter. The parties shall make a good faith effort to locate and destroy any case materials

designated RESTRICTED CONFIDENTIAL - SOURCE CODE.

24. The failure to designate documents, information or material in accordance with this Order and the failure to object to a designation at a given time shall not preclude the filing of a motion at a later date seeking to impose such designation or challenging the propriety thereof. The entry of this Order and/or the production of documents, information and material hereunder shall in no way constitute a waiver of any objection to the furnishing thereof, all such objections being hereby preserved.
25. Any Party knowing or believing that any other party is in violation of or intends to violate this Order and has raised the question of violation or potential violation with the opposing party and has been unable to resolve the matter by agreement may move the Court for such relief as may be appropriate in the circumstances. Pending disposition of the motion by the Court, the Party alleged to be in violation of or intending to violate this Order shall discontinue the performance of and/or shall not undertake the further performance of any action alleged to constitute a violation of this Order.
26. Production of DESIGNATED MATERIAL by each of the Parties shall not be deemed a publication of the documents, information and material (or the contents thereof) produced so as to void or make voidable whatever claim the Parties may have as to the proprietary and confidential nature of the documents, information or other material or its contents.
27. The Receiving Party may not upload to, share with, or otherwise use DESIGNATED MATERIAL of the Producing Party on any generative artificial intelligence (GenAI) or large language model (LLM) tools or platforms including any third-party, public, or internal tools unless the Producing Party approves the use of such tool or the Court issues an Order approving the use of such tool upon request by the Receiving Party.

28. Nothing in this Order shall be construed to effect an abrogation, waiver or limitation of any kind on the rights of each of the Parties to assert any applicable discovery or trial privilege.
29. Each of the Parties shall also retain the right to file a motion with the Court (a) to modify this Order to allow disclosure of DESIGNATED MATERIAL to additional persons or entities if reasonably necessary to prepare and present this Action and (b) to apply for additional protection of DESIGNATED MATERIAL.

So ORDERED and SIGNED this 21st day of November, 2025.



RODNEY GILSTRAP
UNITED STATES DISTRICT JUDGE

**IN THE UNITED STATES DISTRICT COURT
FOR THE EASTERN DISTRICT OF TEXAS
MARSHALL DIVISION**

PERCEPTIVE AUTOMATA LLC,
Plaintiff,

v.

TESLA, INC.,
Defendant.

§
§
§
§
§
§
§
§

Case No. 2:25-cv-00742-JRG

JURY TRIAL DEMANDED

**APPENDIX A
UNDERTAKING OF EXPERTS OR CONSULTANTS REGARDING
PROTECTIVE ORDER**

I, _____, declare that:

1. My address is _____.
My current employer is _____.
My current occupation is _____.
2. I have received a copy of the Protective Order in this action. I have carefully read and understand the provisions of the Protective Order.
3. I will comply with all of the provisions of the Protective Order. I will hold in confidence, will not disclose to anyone not qualified under the Protective Order, and will use only for purposes of this action any information designated as “CONFIDENTIAL,” “RESTRICTED - ATTORNEYS’ EYES ONLY,” or “RESTRICTED CONFIDENTIAL - SOURCE CODE” that is disclosed to me.
4. Promptly upon termination of these actions, I will return all documents and things designated as “CONFIDENTIAL,” “RESTRICTED - ATTORNEYS’ EYES ONLY,” or “RESTRICTED CONFIDENTIAL - SOURCE CODE” that came into my possession,

and all documents and things that I have prepared relating thereto, to the outside counsel for the party by whom I am employed.

5. I hereby submit to the jurisdiction of this Court for the purpose of enforcement of the Protective Order in this action.

I declare under penalty of perjury that the foregoing is true and correct.

Signature _____

Date _____

Attachment C

Analysis of pedestrian dynamics from a vehicle perspective

Julian F. P. Kooij^{†,1,2}, Nicolas Schneider^{‡,1,2} and Darius M. Gavrila^{†,1,2}

Abstract—Accurate motion models are key to many tasks in the intelligent vehicle domain, but simple Linear Dynamics (e.g. Kalman filtering) do not exploit the spatio-temporal context of motion. We present a method to learn Switching Linear Dynamics of object tracks observed from within a driving vehicle. Each switching state captures object dynamics as a mean motion with variance, but also has an additional spatial distribution on where the dynamic is seen relative to the vehicle. Thus, both an object's previous movements and current location will make certain dynamics more probable for subsequent time steps. To train the model, we use Bayesian inference to sample parameters from the posterior, and jointly learn the required number of dynamics. Unlike Maximum Likelihood learning, inference is robust against overfitting and poor initialization.

We demonstrate our approach on an ego-motion compensated track dataset of pedestrians, and illustrate how the switching dynamics can make more accurate path predictions than a mixture of linear dynamics for crossing pedestrians.

I. INTRODUCTION

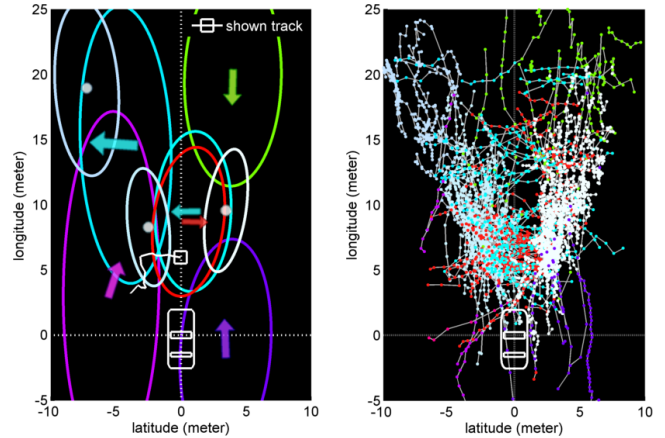
Research in the intelligent vehicles domain has over the years resulted in improved driver assistance systems by successfully applying pattern recognition (e.g. computer vision) and machine learning techniques on complex real-world data. Our company Daimler, for example, has introduced an innovative stereo-based pedestrian safety system in its 2013-2014 Mercedes-Benz S-, E-, and C-Class models, which incorporates fully automatic emergency braking.

Accurate object motion models are key for sophisticated driver assistance systems, that perform tracking of multiple objects and path prediction for situation analysis. Motion dynamics are commonly modeled using a Linear Dynamical System (LDS) with Gaussian noise, for which the well known Kalman Filter can be used to maintain a optimal probabilistic estimate of the true motion state. The underlying assumption that all target motions can be described by the same dynamics can be too simplistic though. One can improve this model by assuming that targets exhibit one out of a set of possible dynamics, i.e. a Mixture of LDS [1]. However, sometimes targets exhibit various dynamics even within the same track. Especially pedestrians can quickly change from, for example, standing to walking when crossing a road [2]. Their behaviors also exhibit spatial structure, e.g. one typically walks parallel to the road on the sidewalk and does not stand still in front of a vehicle.

In this paper, we model object dynamics with a Switching LDS, extended with spatial distributions on where different types of motion occur near the vehicle. We train the model on



(a) Pedestrian dynamics (mean directions and locations) learned from track data, such as the one shown here of a crossing person, as seen from a vehicle.



(b) Top-down view of the situation (vehicle at origin). (c) All pedestrian tracks; positions are color-coded with used dynamics.

Fig. 1. (a), (b) Illustration of learned pedestrian dynamics. For each dynamic the average motion direction is shown as an arrow (or as a spot where the average direction is near zero), as well as the spatial region on the ground plane. The process noise of the dynamics (motion uncertainty) is not shown here. (c) Pedestrian tracks from which dynamics were learned.

real-world pedestrian data recorded from a moving vehicle. Figure 1 illustrates our model and used track data¹.

Unsupervised learning raises questions on determining the number of dynamics needed, and on avoiding overfitting in dynamics with few samples. EM-like schemes also rely on good initial estimates (e.g. [1] required an external clustering method) and Maximum Likelihood learning is prone to overfitting. We therefore use a Bayesian inference scheme to tackle these issues in a more principled way, and learn the number of dynamics from the data without overfitting.

[†] J.F.P.Kooij@uva.nl, [‡] Nicolas.Schneider@daimler.com,

[†] D.M.Gavrila@daimler.com

¹ Environment Perception, Daimler R&D, Ulm, Germany

² Intelligent Systems Laboratory, Univ. of Amsterdam, The Netherlands

¹A video animation is available by following the links from <http://www.gavrila.net>

II. RELATED WORK

Motion models are essential to tasks as tracking [3], anomaly detection [4], [5] and path prediction [2], [6], found in a variety of application domains, from intelligent vehicles [1], [2], [6], [7], [8], [9] to fixed-viewpoint surveillance [4], [5], [10], [11]. In intelligent vehicles for instance, trackers commonly build tracks from detections (e.g. for pedestrians [7], [12], [13]), and use motion models to reduce false positives and resolve data association [7], [9].

When both the motion dynamics and observations can be expressed as a linear transformation of a latent state with added Gaussian noise, then the resulting model is a Linear Dynamical System (LDS). The Kalman Filter (KF) [3] can then be used for efficient forward filtering, maintaining uncertainty over the state by a single normal distribution. For batch processing, the Kalman Smoother extends the KF by a backward pass to incorporate information from future time steps in state estimation. The Extended KF (EKF) [6] or Unscented KF (UKF) [9], [14] are used in certain cases where the observations are not linear transformations of the underlying state space. But the LDS is not an optimal model for track data that contains various dynamics. A Mixture of LDS is an extension where a track uses one out of multiple possible dynamics [1], see Fig. 2(a). But in many cases a target can change dynamics within a single track [5], [15], such as when a person starts walking after standing still.

The Switching Linear Dynamical System (SLDS) [16], [17] allows for different dynamics at each time step, conditioning the linear dynamics on an additional hidden Markov chain of switching states (see Fig. 2(b)) which have their own prior and transition probability. Unfortunately, exact inference in a SLDS is intractable since the latent state depends on the full history of switching states [16]. Expert knowledge can be exploited to some extent to define human motion classes [11], [15], but not all model parameters are known or optimal for a target application. Therefore, we must resort to approximate inference for both training and filtering. For on-line filtering, the Interacting Multiple Models (IMM) algorithm [3], [2], [6] filters a track simultaneously with the various dynamics, and mixes their predictions. Alternatively, [16] presents a collapsed Gibbs sampler [18] for the switching states by integrating over the latent positions.

Gibbs sampling can also be used for learning, though other learning schemes using Viterbi or Variational inference are possible too [19]. Fox et al. [17] presented a Gibbs sampler for a SLDS in a Dirichlet Process [20], estimating both the parameters and number of dynamics. In [5] we took a similar approach to learn person dynamics in a surveillance scenario. Additionally, we learned spatial distributions of where different states occur, while hierarchically clustering tracks that exhibit similar state transitions too.

Our contributions in this paper are two-fold. First, instead of one or multiple Kalman filters to model object dynamics, we use a SLDS and extend it with additional spatial distributions to capture where specific dynamics occur. This does not exclude the possibility of the same dynamic (e.g.

moving left) at spatially distant regions, since this can be represented by separate switching states with similar process noise but different spatial distributions. We describe how to use Bayesian inference to learn both the number of switching dynamics and their parameters. By sampling from the posterior, this approach is more robust to random initialization and overfitting than EM-based learning. Second, our approach is a novel way in the intelligent vehicle domain to analyze real-world track data captured from a vehicle. We show that the method discovers meaningful and spatially localized dynamics for pedestrians. As an example, we show how this structure could yield more accurate path prediction than a baseline using a Mixture of LDS.

III. APPROACH

To train a particular model $p(X|\Theta)$, one needs to search the parameter space Θ and evaluate how probable a particular solution is for the (training) data X . A common approach to learning is to find the Maximum Likelihood (ML) solution,

$$\theta_{ML}^* = \operatorname{argmax}_{\theta} p(X|\Theta = \theta). \quad (1)$$

The predictive density for some new data point x' is then simply $p(x'|\theta_{ML}^*)$. However, this solution is prone to overfitting, especially when there are many parameters, or the training data is sparse. A more principled way to counter these effects is to use an appropriate prior distribution $p(\theta)$ on the parameters we wish to learn, and use Bayes' rule to find the posterior for a Maximum A-Posteriori (MAP) solution,

$$\theta_{MAP}^* = \operatorname{argmax}_{\theta} p(\Theta = \theta|X) = \operatorname{argmax}_{\theta} p(X|\theta)p(\theta), \quad (2)$$

Even better, the posterior could be used for full Bayesian inference by integrating over all possible parameters and account for uncertainty, $p(x'|X) = \int p(x'|\theta)p(\theta|X)d\theta$.

Generally, this integral cannot be solved analytically and exact Bayesian inference is unfortunately intractable. Still, there are various methods for approximation available [18]. Markov Chain Monte Carlo (MCMC) techniques do this by sequentially generating samples $\theta^{(i)}$ of the posterior, using some stochastic process f such that $\theta^{(i)} \sim f(\theta^{(i-1)})$.

Gibbs sampling is a form of MCMC. It updates one or a few hidden variables at a time, sampling them from their marginal distribution while keeping all other hidden variables fixed to their current value. We thus reduce inference in a complex model to iteratively sampling the hidden variables from a few simpler distributions. In comparison to EM [1], MCMC is less sensitive to the initial state since it can move away from local maxima. On sparse data, samples reflect the prior instead of overfitting. As more data is obtained, sampling might become computationally more expensive, but the posterior also contains less uncertainty. Since there is less variance in the samples, fewer iterations can be used instead.

In this paper, we use Gibbs sampling to explore the parameter space of our model. The sampler could be used as an any-time algorithm, by running it in the background and keeping track of the best MAP parameters found at any time, or to approximate full Bayesian inference.

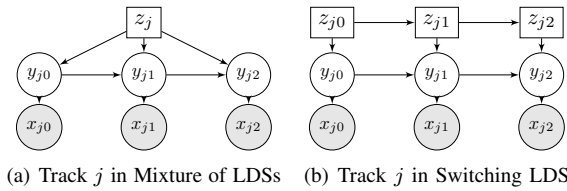


Fig. 2. Graphical model of a Mixture of Linear Dynamical Systems, and a Switching Linear Dynamical System (SLDS). Here discrete nodes are rectangular, continuous ones are round, and observed nodes are shaded. Each model is depicted for three time slices. When modeling a set of tracks as a Mixture of LDSs, then each track has a single dynamic or class label which indicates which of the LDS parameters it uses. The SLDS on the other hand has a dynamic state not per track but per time instance.

A. Model

Let the training data consist of a set of J tracks, each being a sequence observed positions of the target. We will index tracks by suffix j , and let $X_j = \{x_{j1}, \dots, x_{jT_j}\}$ be the T_j observations. Observed positions are typically noisy measurements of latent true positions, which we represent as the corresponding sequence $Y_j = \{y_{j1}, \dots, y_{jT_j}\}$. The SLDS additionally has a latent Markov chain of switching states for each time step, $Z_j = \{z_{j1}, \dots, z_{jT_j}\}$ that determine the used dynamics, see Figure 2(b). So if there are K dynamics, then it holds that $1 \leq z_{jt} \leq K$.

The process noise $\mathcal{N}(m_k, Q_k)$ is non-zero mean and conditioned on the switching state, resulting in the dynamics

$$y_{jt} = Ay_{jt-1} + q_{jt} \quad q_{jt} \sim \mathcal{N}(m_{z_{jt}}, Q_{z_{jt}}) \quad (3)$$

$$x_{jt} = Cy_{jt} + r_{jt} \quad r_{jt} \sim \mathcal{N}(0, R) \quad (4)$$

where q_{jt} is the non-zero mean process noise at time t , and r_{jt} the zero mean observation noise. Matrices A and C are fixed and are determined by the type of states, observations and kinematics used. The temporal transitions in the chain of switching states are parameterized by a $K \times K$ transition matrix $\Pi = [\pi_1; \dots; \pi_K]$, where each row π_k is a parameter vector of K elements for the Categorical distribution over next states given current state k ,

$$p(z_{jt+1} = k' | z_{jt} = k) = \text{Cat}(k' | \pi_k) = \pi_{k(k')}. \quad (5)$$

i.e. the k' -th element of the vector π_k .

We add an additional Normal distribution over the observations to express where a certain motion dynamic is likely to occur. The spatial distribution of switching state k , is parameterized by (μ_k^S, Σ_k^S) ,

$$x_{jt} \sim \mathcal{N}(\mu_{z_{jt}}^S, \Sigma_{z_{jt}}^S). \quad (6)$$

Now there are two predictive density terms for the observations, i.e. the LDS term (4), and the spatial term (6). We use then the product of both as the final predictive density.

In conclusion, the complete set of model parameters is $\Theta = (R, \forall k \{m_k, Q_k, \pi_k, \mu_k^S, \Sigma_k^S\})$.

B. Conjugate Priors

The priors of the model are conjugate (i.e. which have the convenient property that the posterior distributions have the

same parametric form [18]) and shared by all K states,

$$R \sim \mathcal{W}^{-1}(\xi_0^R) \quad (7)$$

$$m_k, Q_k \sim \mathcal{NW}^{-1}(\xi_0^Q) \quad (8)$$

$$\mu_k^S, \Sigma_k^S \sim \mathcal{NW}^{-1}(\xi_0^S) \quad (9)$$

$$\pi_k \sim \text{Stick}(\alpha) \quad (10)$$

with $\xi_0^R = \{\nu_0^R, \Psi_0^R\}$ the parameters of an Inverse-Wishart (IW) distribution, a distribution over covariance matrices, and $\xi_0^Q = \{\mu_0^Q, \kappa_0^Q, \nu_0^Q, \Psi_0^Q\}$ and ξ_0^S the parameters for a Normal-Inverse-Wishart (NIW), a joint distribution over mean and covariance. Intuitively, IW parameters (ν, Ψ) represent ν ‘virtual’ data points with sample covariance matrix Ψ/ν . For low ν , IW spreads its density over many covariance matrices (i.e. we are uncertain of the true covariance that generated data with that sample covariance) and the distribution peaks on the sample covariance as ν increases (i.e. low uncertainty on the covariance). The NIW extends IW with a Normal over a mean value, see [18] for details.

Finally, α is the concentration parameter of the Stick-breaking distribution. This distribution, from which the multinomial parameters π_k are sampled, is a distribution over vectors of possibly *infinite* elements that sum to one. If the vector of component weights in a mixture model is distributed along a Stick-breaking distribution, it is said to be an infinite or Dirichlet Process mixture model [20], [17]. Note that an infinite mixture model can represent finite mixture models of any size, as these are just cases where all except a few of the infinite weights are zero. Thus instead of fixing the cluster count up front, we can learn how many non-zero weights are needed for the training data. Parameter α expresses a prior preference for a few large or many smaller clusters. The advantage of having a concentration parameter is that the actual number of clusters in the data is typically unknown a-priori, and varies from dataset to dataset.

Instead of implementing Stick-breaking explicitly, which requires additional math and more memory management when clusters are added and removed, it has been shown [20] that we can instead use a Dirichlet distribution with a parameter *vector* α_0 of fixed and finite length K ,

$$\pi_k \sim \text{Dir}(\alpha_0), \quad \text{where } \alpha_0 = [\alpha/K, \dots, \alpha/K]. \quad (11)$$

This yields the same distribution as the Stick-breaking prior of Eq. (10) in the infinite limit of $K \rightarrow \infty$. In practice, we regard K as an upper-limit on the actual number of clusters, which is discovered from the data. As sampling converges, certain elements in the vectors π_k will go to zero, and the actual number of clusters is found by counting the number clusters with non-zero probability.

C. Gibbs sampling procedure

Like in [5], [17], we use Gibbs sampling to search the parameter space by iteratively executing the following steps.

a) *Sampling dynamics* $R, m_k, Q_k, \mu_k^S, \Sigma_k^S$: At this stage, we assume that the switching states Z of all tracks are fixed. Due to conjugacy, the posterior distributions of

the model parameters have the same form as their priors. Hence, at the i -th Gibbs iteration they are sampled as,

$$R^{(i)} \sim \mathcal{W}^{-1}(\xi_+^R) \quad (12)$$

$$m_k^{(i)}, Q_k^{(i)} \sim \mathcal{N}\mathcal{W}^{-1}(\xi_{+k}^Q) \quad (13)$$

$$\mu_k^{S(i)}, \Sigma_k^{S(i)} \sim \mathcal{N}\mathcal{W}^{-1}(\xi_{+k}^S) \quad (14)$$

with the posterior parameters $\xi_+^R, \xi_{+k}^Q, \xi_{+k}^S$ computed from the prior parameters and sufficient statistics of the data. Note that all dynamics share the same priors ξ_0^Q, ξ_0^S , but have different posteriors ξ_{+k}^Q, ξ_{+k}^S after considering the data.

To illustrate, if we would know the true values of all Y_j , then we would also know the true motion vectors $q_{jt} = y_{jt} - Ay_{jt-1}$ (Eq. (3)), and observation noise vectors $r_{jt} = x_{jt} - Cy_{jt}$ (Eq. (4)). The motion vectors q_{jt} with $z_{jt} = k$ are then considered samples from the k -th process noise distribution $\mathcal{N}(m_k, Q_k)$. Accordingly, the posterior $\xi_{+k}^Q = (\mu_{+k}^Q, \kappa_{+k}^Q, \nu_{+k}^Q, \Psi_{+k}^Q)$ would be computed from the sample mean, $\hat{q}_k$, and scatter matrix (c.f. [18]) $\hat{S}_k = \sum_{jt|z_{jt}=k} ((q_{jt} - \hat{q}_k)(q_{jt} - \hat{q}_k)^\top)$ of those N_k vectors, as

$$\mu_{+k}^Q = \frac{\kappa_0^Q \mu_0^Q + N_k \hat{q}_k}{\kappa_0^Q + N_k} \quad (15)$$

$$\kappa_{+k}^Q = \kappa_0^Q + N_k \quad (16)$$

$$\nu_{+k}^Q = \nu_0^Q + N_k \quad (17)$$

$$\Psi_{+k}^Q = \Psi_0^Q + \hat{S}_k + \frac{\kappa_0^Q N_k}{\kappa_0^Q + N_k} (\hat{q}_k - \mu_0^Q)(\hat{q}_k - \mu_0^Q)^\top. \quad (18)$$

However, we do not know the true values of the Y_j . But a forward and backward Kalman filter can provide posterior distributions over the latent positions, i.e. the output of a Kalman smoother, which are all normal. One could then sample *pseudo*-true positions Y_t from those posteriors [17], but this is computationally demanding and may lead to slower convergence. Instead, we use the smoothed estimates of the joint distributions $\mathcal{N}(y_{jt-1}, y_{jt}|X_j, Z_j)$ to compute the distributions over the motion vectors $p(q_{jt}|X_j, Z_j)$, which are normals too and parameterized as $\mathcal{N}(q_{jt}|\mu_{jt}, \Sigma_{jt})$. In Eq. (15)-(18) we then use the expected values of $\hat{q}_k$ and $\hat{S}_k$,

$$\mathbb{E}[\hat{q}_k] = \frac{1}{N_k} \sum_{jt|z_{jt}=k} (\mathbb{E}[q_{jt}]) = \frac{1}{N_k} \sum_{jt|z_{jt}=k} (\mu_{jt}) \quad (19)$$

$$\mathbb{E}[\hat{S}_k] = \sum_{jt|z_{jt}=k} ((\mu_{jt} - \mathbb{E}[\hat{q}_k])(\mu_{jt} - \mathbb{E}[\hat{q}_k])^\top + \Sigma_{jt}). \quad (20)$$

The parameters ξ_+^R and ξ_+^S of the posterior over the observation noise and spatial distribution are computed similarly (c.f. [5]) from the Kalman smoother output for Y_t .

b) Integrating out transitions π_k . The posterior distribution over π_k , given prior α_0 , and the evidence consisting of the states z_{jt} following a $z_{jt-1} = k$, is again a Dirichlet distribution, hence the Dirichlet is a conjugate prior [18]. The posterior parameter vector of this distribution are

$$\alpha_{+k} = [\alpha_{0(0)} + N_{0|k}, \dots, \alpha_{0(K)} + N_{K|k}] \quad (21)$$

and $N_{k'|k}$ are the number of $z_{jt} = k'$ that follow a $z_{jt-1} = k$. Optionally, as in [17], an extra prior weight τ can be

added to self-transitions $N_{k|k}$. This enforces self-transitions even more, and lets sampling converge faster. Finally, it is even possible to integrate all π_k out analytically to make the sampler even more efficient (c.f. [20]). Using (5) and (21),

$$p(z_{jt} = k' | z_{jt-1} = k) = \int \text{Cat}(k' | \pi_k) \text{Dir}(\pi_k | \alpha_+) d\pi_k \\ \propto \alpha_{0(k')} + N_{k'|k} \quad (22)$$

So common transitions become more probable. This self-reinforcing property is what drives the clustering process.

c) Sampling switching states Z_j . Now we resample the latent switching states of the tracks, using the updated model parameters. Again, we will not sample values for the hidden positions Y_j , but instead directly sample from the distribution $p(Z_j | X_j, \Theta)$ for all tracks [5], [16], [17]. In [16] it is shown that, for fixed model parameters Θ , a Kalman Information filter can be used for a SLDS to efficiently compute

$$p(z_{jt}|X_j, Z_{j,-t}, \Theta) \\ \propto p(z_{jt}|z_{jt-1})p(z_{jt+1}|z_{jt})p(x_{jt}|z_{jt}, x_{j1:t-1}) \\ \times \int p(x_{jt+1:T}|y_{jt}, z_{jt+1:T})p(y_{jt}|x_{j1:t}, z_{j1:t})dy_{jt} \quad (23)$$

where the first two terms are the transition probability and likelihood, for which we use the result in Equation (22). The third term is the predictive distribution of the observation, found by the forward Kalman filter, and the last integral can be analytically solved and expressed in term of the forward and backward filtering statistics (c.f. [16]).

IV. EXPERIMENTS

We have used our motion model to learn movement dynamics of pedestrians from the perspective of a vehicle.

A. Dataset

The dataset consists of pedestrian trajectories extracted from images recorded in and around various European cities, with an embedded stereo vision system in a car. Since our aim here is to learn dynamics, and not to solve detections (correct vs. false) and track assignment, unoccluded pedestrians have been manually annotated in the video frames with bounding boxes. Then, a pedestrian's distance in each frame was measured in a depth image created from stereo vision, so each bounding box results in a 2D top-down measurement (latitude and longitude) relative to the camera coordinate system. However, tracks built from these measurements contain the combined motion of the moving vehicle and pedestrian. Each track is therefore ego-motion compensated, and placed in a canonical ground plane coordinate system such that they are spatially aligned in a meaningful way. The coordinate system we used was the camera coordinate system in the last frame where the pedestrian is visible, such that paths of pedestrians crossing, or walking along, the road are best aligned. This process is illustrated in Figure 3.

The camera framerate is not constant over all recordings, thus tracks were subsampled to 0.4 seconds per frame (we have also experimented with other framerates, but we did not observe substantial differences in the learned dynamics).

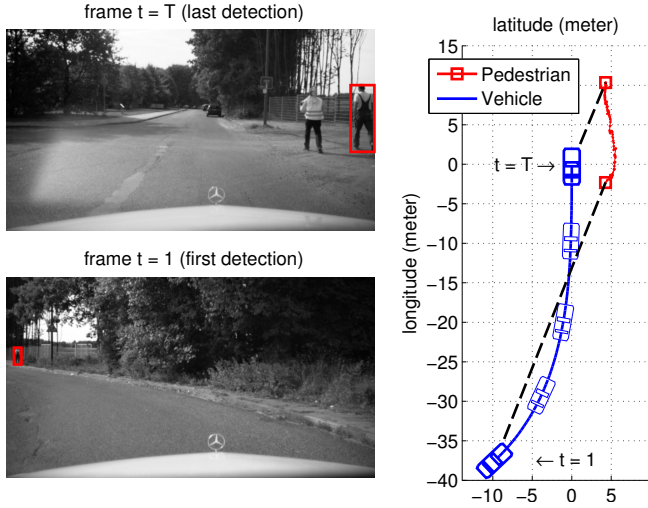


Fig. 3. Constructing a pedestrian track (in red) from the perspective of a driving vehicle (in blue). The pedestrian is seen first at the vehicle's left side and later at the right, due to the vehicle ego-motion. By projecting the ego-motion compensated track in the vehicle's final coordinate system, the track reflects a person walking along the road (i.e. along the longitude axis).

Finally, tracks that are too short or contain gaps due to occlusion are removed. The resulting data consists of 423 tracks and in a total 4103 observations, see Figure 1(c). Most tracks have 6 to 30 observations with a median of 8, though two very long tracks have 55 and 65 time steps. Due to inaccuracies in annotation, depth estimation and ego-motion compensation, the tracks are not noise free, which motivates our use of linear dynamics with observation noise.

B. Learning pedestrian dynamics

The velocity range at which pedestrians move is limited, and a filter should not adapt to high velocities that are anomalous. In our experiments we therefore use as state only a 2D latent position, and set A and C (Eq. (3) and (4)) to identity. This results in a fixed-position model with added velocity determined by the non-zero mean process noise of the switching state, similar to [5]. We applied the Gibbs sampler on the pedestrian track dataset with the upper limit $K = 20$, and set the prior $\alpha_0 = [1/K, \dots, 1/K]$ (i.e. $\alpha = 1$), $\tau = 100$ for self-transitions, the observation noise prior $\nu_0^R = 10$ and $\Psi_0^R = \begin{bmatrix} .3^2 & 0 \\ 0 & .3^2 \end{bmatrix} \times \nu_0^R$ (i.e. prior has 10 'virtual' noise vectors with standard deviation of 0.3 in both directions, see discussion in Section III-B), the process noise prior $\mu_0^Q = [0, 0]$, $\kappa_0^Q = 1$, $\nu_0^Q = 50$ and $\Psi_0^Q = \begin{bmatrix} .3^2 & 0 \\ 0 & .3^2 \end{bmatrix} \times \nu_0^Q$, and set for the spatial prior μ_0^S to the average observation, $\kappa_0^S = 1$, $\nu_0^S = 50$, $\Psi_0^S = \begin{bmatrix} 5^2 & 0 \\ 0 & 15^2 \end{bmatrix} \times \nu_0^S$. Of course, depending on the application we could use other parameters. For example, we could express preference for more dynamics but with smaller spatial regions.

To initialize the sampler, all Z_j are assigned randomly and initial model parameters are taken from the resulting posterior distributions. Sampling converges after ~ 50 iterations, the so-called *burn-in* period [18]. In our hybrid implementa-

tion in Matlab and C++ (for sampling the switching states), one complete sampling run with $K = 20$ on the entire dataset takes ~ 4 seconds on a single 2.6 Ghz core. This could be further improved by porting all code to C++. Note that 50 iterations (≈ 200 seconds) takes less time than recording the actual dataset did ($\approx 4103 \times 0.4 = 1640$ seconds).

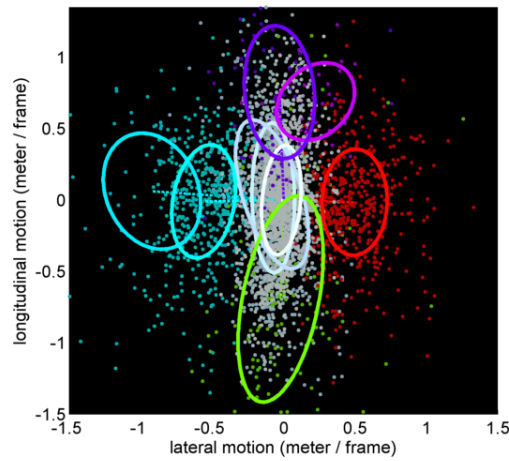
Figure 4(a) shows the sampled process noise after 100 Gibbs iterations, plotted on top of the vectors $\mathbb{E}[q_{jt}]$, the expected true velocities. The hue of each cluster is determined by the velocity direction, while velocities with higher magnitude are more saturated. Due to the Dirichlet prior, we find that only 9 of the 20 possible dynamics are used. In Figure 4(b) the selected values for the Z_j are depicted in the ground plane, after dividing tracks into four sets based on their start and end position for clarity of representation. We see that tracks that have clear lateral motion in front of the car use specialized motion dynamics (in red and blue), and can move from and to states with low lateral motion (colored white). The spatial distributions in Fig. 1(b) show that these low motion areas are located left and right of the car, where longitudinal motion is also typically observed. Indeed, the models captures that longitudinal motion occurs on the sidewalks, and crossing people move fast laterally.

C. Comparison Switching LDS vs. Mixture of LDS

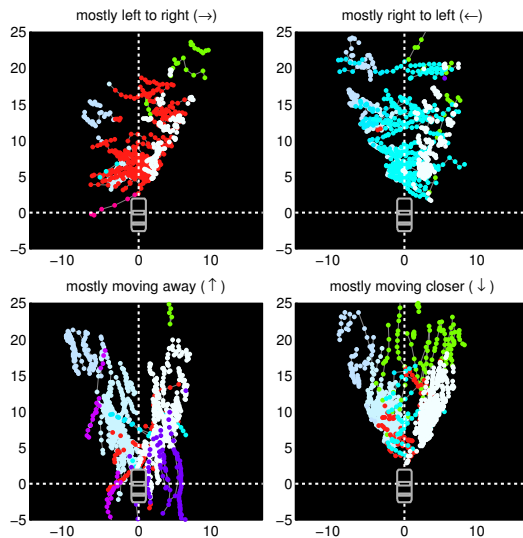
The Gibbs sampler for the SLDS can easily be adapted to a Mixture of LDS, sampling a track's switching state only once and use its dynamics for all time steps. We trained both models on the track data, with the same priors and both with spatial distributions. In figure 5(a) the data log-likelihood $p(X|Y, Z)$ of both models during Gibbs sampling is plotted. It is clear that after the burn-in period the SLDS fits the data better. This difference is illustrated for one track in figure 5(b), where the pedestrian first stands still (observed positions lie close together) and then crosses, and is even more apparent in Figure 5(c) where only the predictive lateral distribution is shown over time. The SLDS selects different states for the standing and walking phase, each with an appropriate mean velocity and low variance. The Mixture of LDS however is forced to use a single dynamic with high variance for such cases that is neither specific for walking nor for standing, and makes therefore less accurate predictions.

V. CONCLUSION

We have presented a SLDS with additional spatial distributions, and demonstrated that it can learn meaningful dynamics that are specialized for different spatial regions around the car. For crossing pedestrians, the capability of the model to switch dynamics enables specialized states for low motion and high motion with a particular direction. This results in dynamics with lower variance (process noise) and thus more accurate predictions. The Gibbs sampler we described is capable of discovering the number of dynamics, starting from random initialization. In this work tracks need to be ego-motion compensated with respect to the last frame. In future work we will look at different ways to incorporate vehicle ego-motion into the prediction.



(a) Process noise of switching states

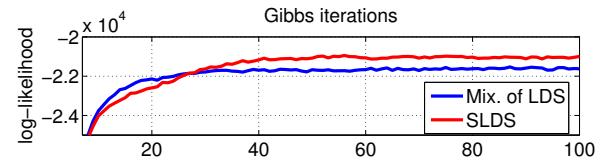


(b) Assigned dynamics (sorted by global motion direction)

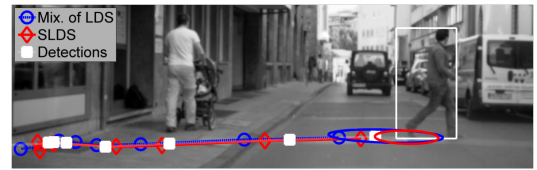
Fig. 4. (a) Ellipses show the process noise $\mathcal{N}(m_k, Q_k)$ of each dynamic. The axes in this plot represent motion, so the origin corresponds to no motion. The dots are the expected true motion vectors $\mathbb{E}[q_{jt}]$. (b) Tracks with assigned switching states of each observation shown in color of (a). Similar to figure 1(c), but splitted in four plots to reduce visual overlap.

REFERENCES

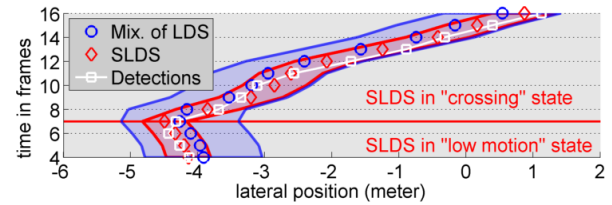
- [1] V. Romero-Cano, J. I. Nieto, and G. Agamennoni, "Unsupervised motion learning from a moving platform," in *Proc. of the IEEE Intelligent Vehicles Symposium*, 2013, pp. 104–108.
- [2] C. G. Keller, C. Hermes, and D. M. Gavrila, "Will the pedestrian cross? Probabilistic path prediction based on learned motion features," *Proc. of the DAGM Symposium on Pattern Recognition*, pp. 386–395, 2011.
- [3] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House Norwood, MA, 1999.
- [4] X. Wang, K. T. Ma, G. W. Ng, and W. E. Grimson, "Trajectory analysis and semantic region modeling using a nonparametric Bayesian model," in *Proc. of the IEEE CVPR*, 2008, pp. 1–8.
- [5] J. F. P. Kooij, G. Englebiene, and D. M. Gavrila, "A non-parametric hierarchical model to discover behavior dynamics from tracks," in *Proc. of the ECCV*. Springer Berlin Heidelberg, 2012, pp. 270–283.
- [6] N. Schneider and D. M. Gavrila, "Pedestrian path prediction with recursive Bayesian filters: A comparative study," in *Proc. of the GCPR*. Springer Berlin Heidelberg, 2013, pp. 174–183.
- [7] D. Geronimo, A. M. Lopez, A. D. Sappa, and T. Graf, "Survey of pedestrian detection for advanced driver assistance systems," *IEEE Trans. on PAMI*, vol. 32, no. 7, pp. 1239–1258, 2010.



(a) The SLDS obtains higher data log-likelihood than Mix. of LDS



(b) Vehicle view of predicted path of pedestrian crossing the road



(c) Lateral distribution of prediction over multiple frames

Fig. 5. (a) Data log-likelihood $p(X|Y, Z)$ on all training data during Gibbs sampling for the Mixture of LDS (in blue) and SLDS (in red). (b) Predictive distributions of both models for a pedestrian that first stands still on the side walk, and then crosses. Markers show mean predictions of previous frames, ellipses show predictive density for current frame. (c) For the same track, the predictive density of just the lateral position over time. The bands show the spatial range of the distribution at two std. dev., markers are on the mean. The red line indicates the moment that the SLDS switches dynamics.

- [8] Z. Chen, D. Ngai, and N. Yung, "Pedestrian behavior prediction based on motion patterns for vehicle-to-pedestrian collision avoidance," in *Proc. of the IEEE ITSC*. IEEE, 2008, pp. 316–321.
- [9] J. Tao and R. Klette, "Tracking of 2d or 3d irregular movement by a family of Unscented Kalman filters," *J. Inf. Commun. Conver. Eng.*, vol. 10, no. 3, pp. 307–314, 2012.
- [10] P. Scovanner and M. F. Tappen, "Learning pedestrian dynamics from the real world," in *Proc. of the IEEE ICCV*, 2009, pp. 381–388.
- [11] G. Antonini, S. V. Martinez, M. Bierlaire, and J. P. Thiran, "Behavioral priors for detection and tracking of pedestrians in video sequences," *IJCV*, vol. 69, no. 2, pp. 159–180, aug 2006.
- [12] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Trans. on PAMI*, vol. 34, no. 4, pp. 743–761, 2012.
- [13] M. Enzweiler and D. M. Gavrila, "Monocular pedestrian detection: Survey and experiments," *IEEE Trans. on PAMI*, vol. 31, no. 12, pp. 2179–2195, 2009.
- [14] M. Meuter, U. Iurgel, S.-B. Park, and A. Kummert, "Unscented Kalman filter for pedestrian tracking from a moving host," in *Proc. of the IEEE Intelligent Vehicles Symposium*, 2008, pp. 37–42.
- [15] C. F. Wakim, S. Capperon, and J. Oksman, "A Markovian model of pedestrian behavior," in *Proc. of the IEEE Int. Conf. on Systems, Man and Cybernetics*, vol. 4. IEEE, 2004, pp. 4028–4033.
- [16] A. V. I. Rosti and M. J. F. Gales, "Rao-Blackwellised Gibbs sampling for switching linear dynamical systems," in *Proc. of the ICASSP*, vol. 1, 2004, pp. 1–809.
- [17] E. Fox, E. Sudderth, M. Jordan, and A. Willsky, "Bayesian nonparametric inference of switching dynamic linear models," *IEEE Trans. on Signal Processing*, vol. 59, no. 4, pp. 1569–1585, 2011.
- [18] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer New York, 2006, vol. 1.
- [19] V. Pavlovic, J. M. Rehg, and J. MacCormick, "Learning switching linear models of human motion," in *NIPS*, 2000, pp. 981–987.
- [20] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical Dirichlet processes," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1566–1581, 2006.

Attachment D

Pedestrian Path Prediction with Recursive Bayesian Filters: A Comparative Study

Nicolas Schneider^{1,2} and Darius M. Gavrilă^{1,2}

¹ Environment Perception, Daimler R&D, Ulm, Germany

² Intelligent Systems Laboratory, Univ. of Amsterdam, The Netherlands

Abstract. In the context of intelligent vehicles, we perform a comparative study on recursive Bayesian filters for pedestrian path prediction at short time horizons ($< 2s$). We consider Extended Kalman Filters (EKF) based on single dynamical models and Interacting Multiple Models (IMM) combining several such basic models (constant velocity/acceleration/turn). These are applied to four typical pedestrian motion types (crossing, stopping, bending in, starting). Position measurements are provided by an external state-of-the-art stereo vision-based pedestrian detector. We investigate the accuracy of position estimation and path prediction, and the benefit of the IMMs vs. the simpler single dynamical models. Special care is given to the proper sensor modeling and parameter optimization. The dataset and evaluation framework are made public to facilitate benchmarking.

1 Introduction

Pedestrian path prediction is an important problem in several application contexts, such as architecture, social robotics and intelligent vehicles. Here we consider the intelligent vehicle context, in view of driver assistance and active pedestrian safety. Strong gains have been made over the years in improving computer vision-based pedestrian recognition performance. This has culminated in first active pedestrian safety systems reaching the market. For example, Mercedes-Benz introduces in its 2013 E- and S-Class models a novel stereo-vision based pedestrian system, which incorporates automatic full emergency braking.

A sophisticated situation assessment requires a precise estimation of the current and future position of the pedestrian with respect to the moving vehicle. A deviation of, say, 30 cm in the estimated lateral position of the pedestrian can make all the difference, as this might place the pedestrian just inside or outside the driving corridor. Current active pedestrian systems are typically designed conservatively in their warning and control strategy, emphasizing the current state rather than prediction, in order to avoid false system activations. Indeed, pedestrian path prediction is a challenging problem, due to the highly dynamic behavior of pedestrians. They can change their walking direction in an instance, or start/stop walking abruptly. As a consequence, sensible prediction horizons are typical short (we consider $< 2s$ in this paper).



Fig. 1. Four typical pedestrian motion types: bending in (top left), stopping (top right), crossing (bottom left) and starting (bottom right) with detection bounding boxes

There has been surprisingly little analysis in previous work of the accuracy of pedestrian state estimation, let alone, that of prediction, in vehicle context. This paper addresses this by providing a quantitative comparative study of recursive Bayesian filters: we consider Extended Kalman Filters (EKF) based on single dynamical models and Interacting Multiple Models (IMM) combining several such basic models (constant velocity/acceleration/turn). These are applied to four typical pedestrian motion types (crossing, stopping, bending in, starting), see Fig. 1. Position measurements are provided by an external state-of-the-art stereo vision-based pedestrian detector. The rationale for focusing on recursive Bayesian filters in connection with modeling pedestrians as point targets is their relatively good performance and low computational cost (especially important in a vehicle context). We investigate the accuracy of position estimation and path prediction, and the benefit of the IMMs vs. the simpler single dynamical models. Special care is given to the proper sensor modeling and parameter optimization.

2 Previous Work

In this section, we focus on pedestrian state estimation based on parametric, recursive Bayesian filters. For an overview of vision-based pedestrian detection and tracking in more general context, see recent surveys (e.g. [7,8]).

A popular choice for target state estimation is the Kalman Filter (KF). Its applicability in real-time systems has been proven over many years for different sensors and application domains [1,3,4,18,21]. State parameters (e.g. position, velocity, acceleration) of the tracked target can be estimated with appropriate dynamical and measurement models. The KF can further be used for prediction by propagating the current state with the dynamical model without the inclusion of new measurements. Work by [3] on FIR-based pedestrian tracking uses a constant acceleration (CA) model in image space. Working in image space, however, makes it difficult to incorporate prior knowledge on the dynamics of pedestrian motion. Therefore, [2] track pedestrians on the ground plane using a KF in an indoor, static stereo camera setup. The use of a linear KF in the

context of video-based pedestrian tracking in the world implies the use of 3D pseudo-measurements (i.e. back projection of 2D measurements); this does not account for the dependency of the longitudinal component of the noise on depth.

More accurate measurement models for the perspective projection of video sensors can be incorporated by means of non-linear Extended (EKF) or Unscented (UKF) Kalman filters. [15] use a UKF in a mono camera setup to track pedestrians on the ground plane (CV model). [19] apply the UKF to measurements from a stereo camera system comparing three different dynamical models (two CV and one constant position (CP) model) where two models have a state space in world coordinates and one in image coordinates.

KF-based approaches have also been used for pedestrian state estimation outside the video-only domain. [9] apply a CV model in a multi-sensor setup with an IR camera and laser scanner. In a previous paper [18], they used two different motion models (CA and CTRV), mentioning advantages of the latter model at near-zero pedestrian speeds. Work by [21] considers a setting where pedestrians wear electronic tags. It uses a KF with a turn motion model including orientation and velocity in polar coordinates (CTRV).

Maneuvering targets can be elegantly accounted for mathematically by means of the Interacting Multiple Model (IMM) framework [4,13]. [11] use an IMM (CP/CV) for analyzing walking vs. stopping pedestrian motion types from a stereo vision sensor on-board a vehicle. [5] use an IMM combining eight CV models with fixed velocities in eight directions. It further contains an online adaption algorithm for the IMM transition probability matrix.

Within the class of non-parametric methods for pedestrian path prediction and action classification, [11] proposes a probabilistic trajectory matching method to estimate whether a pedestrian walking towards the curbside intends to cross or not, when viewed from a stereo vision system on-board a vehicle. [12] considers the complementary case of whether a pedestrian standing will start to walk using a SVM-based classification approach, albeit from a static monocular camera.

Quantitative evaluations of pedestrian state (position) estimation have been few and limited. [3,5,9,15,18,21] do not include any such evaluation. [2] provides accuracy figures only related to its KF approach in indoor setting. [19] uses simulated data to compare CV and CP KFs. Our paper contribution is a broad quantitative study on pedestrian position estimation and path prediction using parametric Bayesian recursive filters in vehicle context. Compared to [11], we consider a wider range of pedestrian motion types. Whereas the IMM used by [11] uses 3D pseudo measurements and KFs, we use a more accurate stereo sensor modeling by EKFs.

3 Recursive Bayesian Filtering

3.1 Kalman Filter

The discrete-time KF estimates a state $\mathbf{x}(t)$ at time step t from measurement $\mathbf{z}(t)$ and previous state $\mathbf{x}(t-1)$ with the dynamical model

$$\mathbf{x}(t) = A\mathbf{x}(t-1) + Bu(t-1) + \boldsymbol{\omega}(t-1) \quad (1)$$

where the relation between measurement and state is given by

$$\mathbf{z}(t) = H\mathbf{x}(t) + \boldsymbol{\nu}(t). \quad (2)$$

A and B are transition matrices for the state $\mathbf{x}$ and the control input u , respectively, $\boldsymbol{\omega}(t-1)$ and $\boldsymbol{\nu}(t)$ are white, zero-mean, uncorrelated noise processes with covariances $\boldsymbol{\omega}(t) \sim (0, Q(t))$ and $\boldsymbol{\nu}(t) \sim (0, R(t))$. The filter process can be described as cycle of the two steps prediction (predicting the state $\mathbf{x}(t-1)$ to the next time step) and correction (updating the predicted state $\hat{\mathbf{x}}(t)$ with the current measurement) [20].

3.2 Interacting Multiple Model Kalman Filter

There are several KF extensions available to cover different motion types and maneuvers (see [13] for an overview), the most common is the Interacting Multiple Model KF (IMM). The IMM models that there is a probability of p_{ij} that the tracking target makes a transition from one type of motion (i) to another (j); these values are captured by the transition probability matrix (TPM). Each iteration of the IMM consists of the three steps: interaction, filtering and model probability update [4]. In the interaction step, the mixing probability $\boldsymbol{\mu}_{ij}(t-1)$ (cond. probability that the target changed its type of motion) is calculated based on model probabilities and the TPM to produce mixed state estimates $\hat{\mathbf{x}}_j^0(t-1)$ and covariances $\hat{P}_j^0(t-1)$ for all models j . The mixed states are used as input in the filtering step where each model is predicted and updated with the standard KF equations. In the last step, the model probabilities are updated based on the measurement likelihood.

3.3 Measurement Model

Measurements come from a pedestrian detector applied on sequences recorded with a stereo camera system. A measurement vector (dropping time index t in the following) $\mathbf{z} = (u, d)$ is derived from the footpoint $p_f = (u, v)$ and the median disparity d of a pedestrian bounding box. The relation of a point in the image $p_i = (u, v)$ and its disparity d to a point $p_c = (x_c, y_c, z_c)$ in the camera coordinate system is given by the perspective camera model [1]:

$$\begin{pmatrix} u \\ v \\ d \end{pmatrix} = \begin{pmatrix} h_1(p_c) \\ h_2(p_c) \\ h_3(p_c) \end{pmatrix} = \begin{pmatrix} u_0 + \frac{f_u x_c}{z_c} \\ v_0 + \frac{f_v y_c}{z_c} \\ \frac{f_u b}{z_c} \end{pmatrix} \quad (3)$$

where $f_u = \frac{f}{s_u}$ and $f_v = \frac{f}{s_v}$ with focal length f , baseline b , horizontal and vertical pixel width s_u and s_v , respectively. Eq. (3) leads to the nonlinear measurement function h . For a position $p_c^g = (x_c, z_c)$ on the groundplane h_2 can be ignored. To predict a measurement at time step t , the predicted state vector $\hat{\mathbf{x}}$ (camera coordinates) has to be projected into the measurement (image) space with $\hat{\mathbf{z}} = h(\hat{\mathbf{x}})$. For the EKF we further need to calculate the Jacobian $H = \frac{\partial h}{\partial \mathbf{x}}$.

Table 1. Mean sojourn times of different target dynamics in the training set (diagonals $P_{i,i}$ of the TPM based on a cycle time of $T \approx 60\text{ ms}$). “Straight Walking” consists of the straight walking segments of starting, stopping and bending in sequences as well as complete crossing sequences. “Maneuver” relates to all other segments. “Turning”, a subset of “Maneuver”, relates to the turning segments within the bending in sequences.

Motion type	mean sojourn time τ_i (s)	$P_{ii}(T) = 1 - T/\tau_i$
Straight Walking	6.66	0.99
Maneuver	1.67	0.96
Turning	2.50	0.98

3.4 Dynamical Models

Several discretized continuous-time dynamical models are considered in this study: the popular constant velocity (white noise acceleration) model (CV), the constant acceleration (Wiener process acceleration) model (CA) and the constant turn model (CT) with Cartesian state vector. These characterized by their state vectors $\mathbf{x}$, transition matrices A and process noise matrices Q . The CV model state vector holds position and velocity ($\mathbf{x} = [x, z, v_x, v_z]$), the CA model further has acceleration ($\mathbf{x} = [x, z, v_x, v_z, a_x, a_z]$) and the CT model turn rate ($\mathbf{x} = [x, z, v_x, v_z, \omega]$) variables. For details, such as transition and process noise matrices, see [4,14].

Several approaches can be taken to specify the TPM. There is the ad-hoc approach to fill the diagonals with values close to one. [4,13] discuss the use of the mean sojourn time (the mean time a target stays in a motion type) for the TPM. Lastly, one could perform parameter optimization of the entries of the TPM directly. In preliminary experiments, we obtained similar best performance with the second and third approaches, thus we selected the sojourn time approach to specify the TPM, derived from a training set, see Section 4 and Table 1.

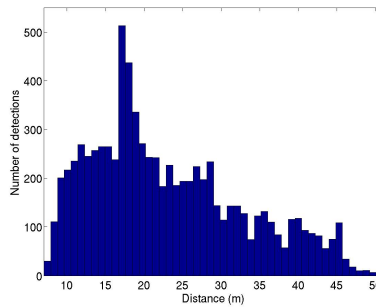
3.5 Ego Motion Compensation

At each time step, the filter state is projected from the previous camera coordinate system to the current one using the inertial motion matrix M_v (vehicle coordinates) based on velocity and yaw rate measured by on-board sensors. The inverse ego motion homography matrix is given by $M_c = D^{-1}M_vD$ (where D defines the relation between camera and vehicle coordinate system). Translational ego compensation is done using t_{M_c} as control vector u (Eq. (1) with $B = I_{2 \times 2}$), the ego rotation is integrated into the transition matrix A_e (exemplary for the CV model) [16]:

$$A_e = \begin{bmatrix} R_{M_c} & 0_{2 \times 2} \\ 0_{2 \times 2} & R_{M_c} \end{bmatrix} A \quad (4)$$

Table 2. Sequences in our dataset recorded with standing and moving vehicle

Sequences	veh. stand.	veh. mov.	total
Bending in	5	18	23
Stopping	5	13	18
Starting	0	9	9
Crossing	3	15	18

**Fig. 2.** Pedestrian distance distribution over entire dataset

4 Experiments¹

Dataset. Image sequences were recorded with a stereo camera system (baseline 22 cm, 16 fps, 1176×640 pixels) mounted behind the windshield of a vehicle. They contain four typical pedestrian motion types: (walking laterally towards the street and) crossing, (walking laterally towards the street and) stopping, (standing at the curbside and) starting (to walk laterally) and (walking alongside the street,) bending in (and crossing). See Fig. 1. The dataset consists of 68 sequences of which 12485 images contain (single) pedestrians. 55 sequences were recorded at vehicle speeds of 20–30 km/h, the others involved a standing vehicle. See Table 2 and Figs. 2 and 6 (bottom left) for further data statistics. The dataset is splitted evenly in a training and test set. The latter was splitted 5-folds for evaluation in Fig. 6 (bottom right); parameter optimization was quite time consuming, so we preferred not vary the training set as part of cross-validation.

Ground Truth (GT) is obtained by manual labeling of pedestrian bounding boxes and computing the median disparity over the rough upper pedestrian body area. The position of the pedestrian in the vehicle coordinate system is calculated with Eq. (3) and the camera-to-vehicle homography matrix. The transformed positions are fitted with a curvilinear model. The GT locations are obtained by longitudinal projections on the fitted curve. Sequences are further labeled with event tags and time-to-event (TTE in frames) values. For stopping

¹ The dataset and evaluation framework are made public for non-commercial research purposes. Follow the links from <http://isla.science.uva.nl/> or contact the 2nd author.

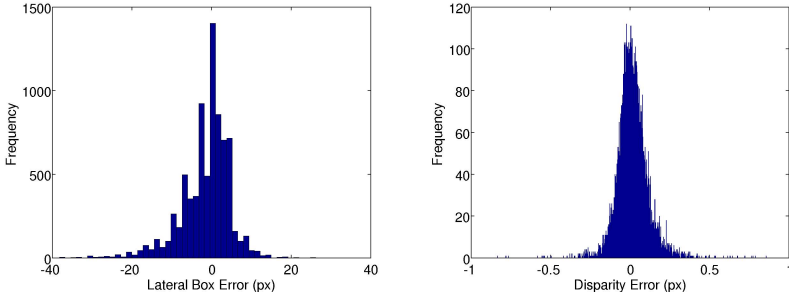


Fig. 3. Measurement error distribution: (left) lateral, u (right) longitudinal, d

Table 3. Optimized process noise parameters (σ_v , σ_a , $\{\sigma_v, \sigma_\omega\}$) for the different filters

Filter	EKF CV	EKF CA	EKF CT	IMM(CV,CA)	IMM(CV,CT*)
Process noise	0.77	0.44	{0.95, 0.90}	(0.70, 0.80)	(0.75, {0.40, 0.90})

pedestrians the last placement of the foot on the ground at the curbside is labeled as $TTE = 0$. For crossing pedestrians, the closest point to the curbside (before entering the roadway), for pedestrians bending in and starting to walk the first moment of visually recognizable body turning or leg movements are labeled with $TTE = 0$. All frames previous to an event have TTE values > 0 , therefore all frames following the event have TTE values < 0 .

A state-of-the-art HOG/linSVM pedestrian detector [6] provides measurements, given region-of-interests supplied by an obstacle detection component using dense stereo data [10]. The resulting bounding boxes are used to calculate a median disparity over the upper pedestrian body area based on the disparity maps. The measurement vector $\mathbf{z} = (u, d)$ is derived using the central lateral position of the bounding box and this median disparity value.

Evaluation Setup. Sequences have an average of 121 measurements (min. 39, max. 274), they start with a minimum of three consecutive measurements, and contain no missing detections longer than five consecutive frames. Evaluation is done with respect to the lateral localization error in TTE interval $[10, -50]$, corresponding to an evaluation interval from 0.60 s before to 3.0 s after the event. At each time step, predictions of up to 32 frames (1.9 s) are made ahead. We use the predict and update functions of the EKF/UKF MATLAB toolbox [17].

Parameter Setting. The measurement noise ν (see Eq. (2)) has been derived statistically on the training set, in terms of lateral bounding box error σ_u and disparity error σ_d of the pedestrian detections w.r.t. the GT. See Fig. 3. The measurement noise matrix $R = \text{diag}(\sigma_u^2, \sigma_d^2)$ is thus set to $\sigma_u = 6.15$ and $\sigma_d = 0.32$.

Process noise ω is determined by $Q(t) = Q^0(t)q$, where $q \in \{\sigma_v^2, \sigma_a^2\}$ and for the CT model, $Q_{5,5}(t) = \frac{\sigma_\omega^2 Q_{5,5}^0(t)}{q}$ [4]. It has been optimized for each filter

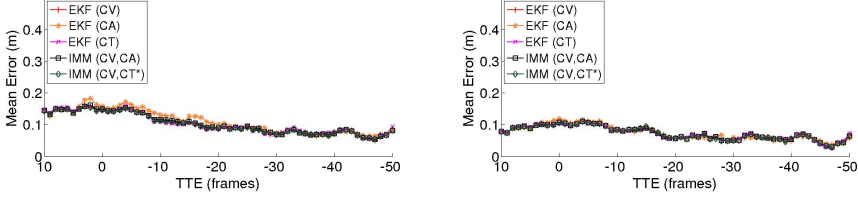


Fig. 4. Position error at current time ($t = 0$) averaged over all sequences: lateral and longitudinal combined (left) and only lateral (right)

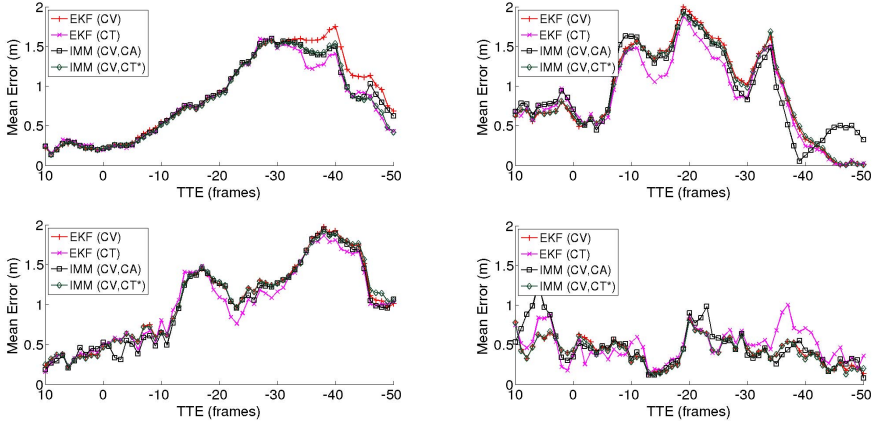


Fig. 5. Mean lateral position error when predicting 32 frames ($t = 1.9$ s) ahead

in terms of positions mean-squared-error (MSE) including all N position state estimates $\mathbf{x}_i(t = 0)$ ($i = 1, \dots, N$) and the corresponding P predictions $\mathbf{x}_i(t = 1, \dots, P)$ ($P = 32$) with the objective function:

$$\arg \min_{\sqrt{q}} \frac{1}{N} \sum_{i=1}^N \left(\sum_{t=0}^P \frac{MSE(\mathbf{x}_i(t))}{P+1} \right) \quad (5)$$

using a discrete parameter search on the training set. The outcome is shown in Table 3. Search space has been discretized using 60 steps for single models with one noise parameter (CV, CA) and 18 steps each for models with two noise parameters (CT, 324 parameter combinations). In a coarse-to-fine fashion discretization for the IMM(CV,CA) could be reduced based on single model results to 9 steps (81 parameter combinations). Including the TPM with a discretization of 12 values per diagonal in the optimization results in 11664 parameter combinations. Furthermore, the CT model has been optimized using only segments of bending in sequences around the labeled turn event ($TTE \in [10 - 50]$). The resulting “turn expert” filter will be termed CT* in the remainder.

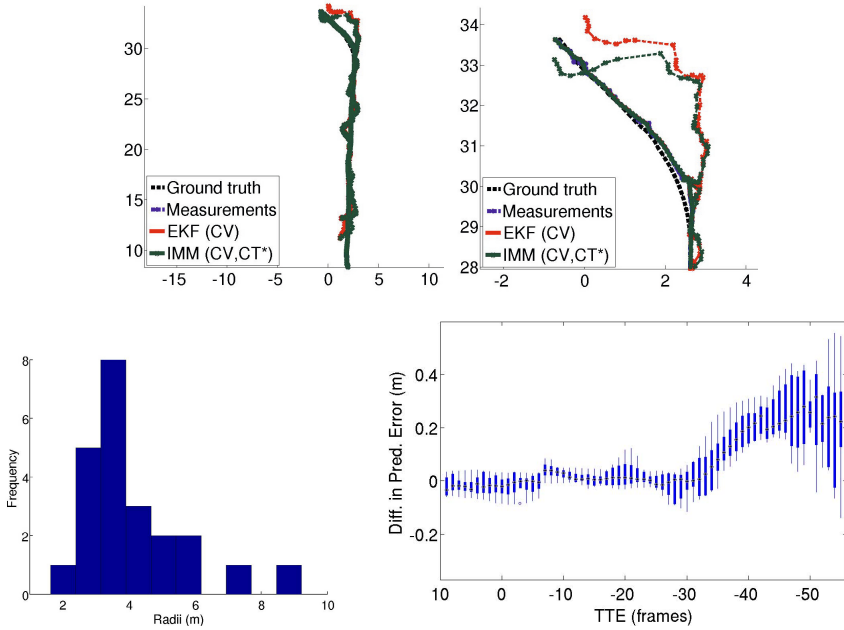


Fig. 6. Bending in: Bird's eye view of an example trajectory showing predictions (32 steps ahead) for various filters (top row, far view and close up, left and right). Turn radius distribution over the bending in sequences (bottom left). Distribution of the prediction error improvement by IMM(CV,CT*) vs. EKF(CV), based on cross-validation.

Results. Fig. 4 shows the position error at current time ($t = 0$), averaged over all sequences. One observes a similar performance for all filters. Performance differences become more evident when predicting 32 frames (1.9 s) ahead, see Fig. 5 (the CA model was removed from the plots since its predictions were far off, i.e. velocities accumulate over the prediction horizon to implausible values). In terms of the single motion models (CV vs. CT), one observes benefits for the CV on the crossing sequences, and benefits for the CT on the others. For example, CT predictions during the turning of the bending in sequence are more accurate by up to 36 cm, compared to CV. The IMM(CV,CT*) combines the best of both worlds, it shows an improvement of up to 30 cm vs. CV. One further observes that IMM(CV,CA) does not outperform CV and lags IMM(CV,CT*), overall. A more detailed analysis of the bending in case is given in Fig. 6.

5 Conclusions

In this paper, we studied several single dynamical models (CV, CA, CT) and IMMs combining such basic models for pedestrian position estimation and path prediction, in vehicle context. Results show no significant performance gain of the more sophisticated IMMs considered vs. the simpler CV, for current position estimation. We attribute this to the high sampling rate and the low measurement error for this application. For path prediction (1.9 s ahead), an IMM(CV,CT*)

involving a constant velocity and a “turn expert” model, is shown to provide an improvement in the lateral position estimation of up to 30 *cm* during maneuvers. Future work involves extending the database, both in terms of motion types considered and in terms of their sample count.

References

1. Barth, A., Franke, U.: Estimating the driving state of oncoming vehicles from a moving platform using stereo vision. *IEEE Trans. ITS* 10(4), 560–571 (2009)
2. Bertozzi, M., et al.: Pedestrian localization and tracking system with Kalman filtering. In: *IEEE Intell. Veh.*, pp. 584–589 (2004)
3. Binelli, E., et al.: A modular tracking system for far infrared pedestrian recognition. In: *IEEE Intell. Veh.*, pp. 759–764 (2005)
4. Blackman, S., Popoli, R.: *Design and Analysis of Modern Tracking Systems*. Artech House Norwood, MA (1999)
5. Burlet, J., et al.: Pedestrian tracking in car parks: An adaptive Interacting Multiple Models based filtering method. In: *Proc. of the IEEE ITSC*, pp. 462–467 (2006)
6. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: *Proc. CVPR*, vol. 1, pp. 886–893. IEEE (2005)
7. Dollár, P., et al.: Pedestrian detection: An evaluation of the state of the art. *IEEE PAMI* 34(4), 743–761 (2012)
8. Enzweiler, M., Gavrila, D.: Monocular pedestrian detection: Survey and experiments. *IEEE PAMI* 31(12), 2179–2195 (2009)
9. Fardi, B., Scheunert, U., Wanielik, G.: Shape and motion-based pedestrian detection in IR images: a multisensor approach. In: *IEEE Intell. Veh.*, pp. 18–23 (2005)
10. Hirschmüller, H.: Stereo processing by semiglobal matching and mutual information. *IEEE PAMI* 30(2), 328–341 (2008)
11. Keller, C.G., Hermes, C., Gavrila, D.M.: Will the pedestrian cross? Probabilistic path prediction based on learned motion features. In: Mester, R., Felsberg, M. (eds.) *DAGM 2011. LNCS*, vol. 6835, pp. 386–395. Springer, Heidelberg (2011)
12. Köhler, S., et al.: Early detection of the pedestrians intention to cross the street. In: *Proc. of the IEEE ITSC*, pp. 1759–1764 (2012)
13. Li, X.R., Jilkov, V.P.: Survey of maneuvering target tracking. Part V. Multiple-model methods. *IEEE Trans. Aerosp. Electron. Syst.* 41(4), 1255–1321 (2005)
14. Li, X.R., Jilkov, V.: Survey of maneuvering target tracking. Part I. Dynamic models. *IEEE Trans. Aerosp. Electron. Syst.* 39(4), 1333–1364 (2003)
15. Meuter, M., et al.: Unscented Kalman filter for pedestrian tracking from a moving host. In: *IEEE Intell. Veh.*, pp. 37–42 (2008)
16. Rabe, C.: *Detection of Moving Objects by Spatio-Temporal Motion Analysis*. Ph.D. thesis, University of Kiel, Kiel, Germany (2011)
17. Särkkä, S., Hartikainen, J., Solin, A.: EKF/UKF toolbox for Matlab v1.3 (2011), <http://becs.aalto.fi/en/research/bayes/ekfukf/>
18. Scheunert, U., et al.: Multi sensor based tracking of pedestrians: a survey of suitable movement models. In: *IEEE Intell. Veh.*, pp. 774–778 (2004)
19. Tao, J., Klette, R.: Tracking of 2d or 3d irregular movement by a family of Unscented Kalman filters. *JICCE* 10(3), 307–314 (2012)
20. Welch, G., Bishop, G.: An introduction to the Kalman filter. In: *Proc. of the ACM SIGGRAPH*. ACM Press, Addison-Wesley, Los Angeles, CA (2001)
21. Westhofen, D., et al.: Transponder- and camera-based advanced driver assistance system. In: *IEEE Intell. Veh.*, pp. 293–298 (2012)

Attachment E

Will the Pedestrian Cross?

A Study on Pedestrian Path Prediction

Christoph G. Keller and Dariu M. Gavrila

Abstract—Future vehicle systems for active pedestrian safety will not only require a high recognition performance but also an accurate analysis of the developing traffic situation. In this paper, we present a study on pedestrian path prediction and action classification at short subsecond time intervals. We consider four representative approaches: two novel approaches (based on Gaussian process dynamical models and probabilistic hierarchical trajectory matching) that use augmented features derived from dense optical flow and two approaches as baseline that use positional information only (a Kalman filter and its extension to interacting multiple models). In experiments using stereo vision data obtained from a vehicle, we investigate the accuracy of path prediction and action classification at various time horizons, the effect of various errors (image localization, vehicle egomotion estimation), and the benefit of the proposed approaches. The scenario of interest is that of a crossing pedestrian, who might stop or continue walking at the road curbside. Results indicate similar performance of the four approaches on walking motion, with near-linear dynamics. During stopping, however, the two newly proposed approaches, with nonlinear and/or higher order models and augmented motion features, achieve a more accurate position prediction of 10–50 cm at a time horizon of 0–0.77 s around the stopping event.

Index Terms—Computer vision, pedestrian safety, prediction methods.

I. INTRODUCTION

PREDICTING the path of a pedestrian is important in several application contexts, such as robot control in human-inhabited environments and driver assistance systems for improved traffic safety. In this paper, we consider the intelligent vehicles context, in which strong gains have been made over the years in improving computer-vision-based pedestrian recognition performance. This has culminated in the first active pedestrian safety systems reaching the market. For example, our company Daimler has recently introduced an innovative stereo-vision-based pedestrian system in its 2013 Mercedes-

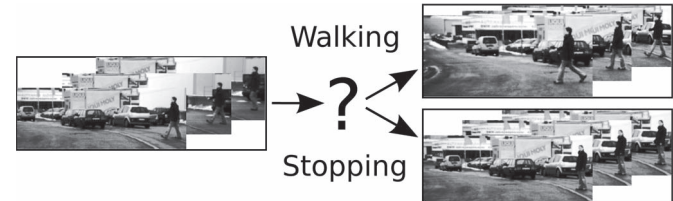


Fig. 1. Pedestrian path prediction and action classification. Where exactly will the pedestrian be in the immediate future? Will the pedestrian cross?

Benz E- and S-Class models, which incorporates automatic full emergency braking.

A sophisticated situation assessment requires a precise estimation of the current and future positions of the pedestrian with respect to the moving vehicle. A deviation of, for example, 30 cm in the estimated lateral position of the pedestrian can make all the difference between a “correct” and an “incorrect” maneuver initiation. One major challenge is the highly dynamic behavior of pedestrians, which can change their walking direction in an instance, or start/stop walking abruptly. As a consequence, prediction horizons for active pedestrian systems are typically short; even so, small performance improvements can produce tangible benefits. For example, accident analysis [1] shows that being able to initiate emergency braking 0.16 s (four frames at 25 Hz) earlier, at a time to collision of 0.66 s, reduces the chance of incurring injury requiring a hospital stay from 50% to 35%, given an initial vehicle speed of 50 km/h.

This paper focuses on the task of predicting the position of pedestrians walking toward the road curbside, when viewed from an approaching vehicle. A secondary question is whether the pedestrian will cross or stop. The setting in Fig. 1 is inspired by an earlier human factors study by Schmidt and Färber [2], which had several test participants watch videos of pedestrians walking toward the curbside and decide whether the pedestrians would stop or cross, at various time instants. The study varied the amount of visual information provided to the test participants and examined its effect on their classification performance. In the baseline case, the pedestrian was fully visible, whereas in other cases, parts of the pedestrian’s body were masked out. Masking the complete pedestrian, and leaving only positional information (bounding box), decreased human accuracy markedly, showing the importance of augmented visual features for this prediction task.

We address the following questions in this paper.

- At the short prediction horizons typical of the traffic safety context, can nonlinear models outperform linear models,

Manuscript received September 25, 2012; revised June 4, 2013, August 6, 2013, and August 23, 2013; accepted September 2, 2013. Date of publication October 17, 2013; date of current version March 28, 2014.

C. G. Keller was with the Image and Pattern Analysis Group, Department of Mathematics and Computer Science, University of Heidelberg, 69117 Heidelberg, Germany. He is now with the Environment Perception Department, Daimler Research and Development, 71034 Böblingen-Hulb, Germany.

D. M. Gavrila is with the Environment Perception Department, Daimler Research and Development, 89081 Ulm, Germany and also with the Intelligent Autonomous Systems Group, University of Amsterdam, 1098 Amsterdam, The Netherlands.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TITS.2013.2280766

or alternatively, can higher order Markov models outperform their first-order counterparts?

- Do augmented visual features (optical flow) improve path prediction and action classification over the use of positional information only?
- How does measurement error (e.g., pedestrian localization error and vehicle egomotion estimation error) affect the results? Can the more complex models still maintain an edge over the simpler ones?

In order to provide answers for the preceding questions, we consider four approaches that differ in their modeling of dynamics and in their use of augmented visual features; together, they cover a broad spectrum of possible approaches. In the category of nonlinear first-order models with augmented visual features, we propose a novel pedestrian path prediction approach, based on Gaussian process dynamical models (GPDMs) [3] and dense optical flow features (see Section III-A). An appealing aspect of this approach is that a low-dimensional latent representation is learned from the data, which takes into account the process dynamics. In the category of nonlinear higher order models with augmented visual features, we propose a novel probabilistic hierarchical trajectory matching (PHTM) approach, based on a low-dimensional motion representation (see Section III-B). Finally, in the category of first-order Markov models using positional information only, and mostly as a baseline, we consider the popular Kalman filter (KF, linear model) and its extension interacting multiple model KF (IMM KF, mixture of linear models) [4] (see Section III-C).

Experimental results on real traffic data are given in Section IV, with pedestrian image location obtained either from ground truth (GT) (optionally corrupted with noise) or obtained by a state-of-the-art pedestrian detection system. Several experimental cases are distinguished (pedestrian stopping versus walking, egovehicle standing versus moving). A discussion of the results, in terms of prediction performance and computational cost, is given in Section V. This paper concludes in Section VI.

II. PREVIOUS WORK

Here, we discuss pedestrian motion model and path prediction techniques. For an overview of vision-based pedestrian detection, we refer to the surveys of Dollar *et al.* [5] and Enzweiler and Gavrila [6].

One way to perform path prediction relies on closed-form solutions for Bayesian filtering; in the KF [4], the current state of a dynamic system can be propagated to the future by means of the underlying linear dynamical model, without the incorporation of new measurements. The same idea can be applied to KF extensions to either multiple linear dynamical models, e.g., the IMM KF [4], or to nonlinear models, e.g., the extended KF or the unscented KF (see [7] and [8] for applications to pedestrian tracking).

An alternative approach to path prediction involves nonparametric stochastic models. Possible trajectories are generated by Monte Carlo simulations, taking into account the respective dynamical models. For example, Keller *et al.* [9] described an integrated vehicle safety system that combines sensing, situation

analysis, decision making, and vehicle control, to automatically brake or evade for pedestrians. Collision detection assumes constant motion from the last estimates of pedestrian position and motion. Abramson and Steux [10] combined a constant motion model with particle filtering. De Nicolao *et al.* [11] distinguished lateral and longitudinal pedestrian velocities and model these independently by a random walk. Wakim *et al.* [12] modeled pedestrian motion by means of four states of a Markov chain, corresponding to standing still, walking, jogging, and running. Each state is associated with probability distributions of magnitude and direction of pedestrian velocity; the state changes are controlled by various transition probabilities. Recently, more complex pedestrian motion models have also accounted for group behavior and spatial layout, e.g., entry/exit points (see Antonini *et al.* [13] for a discussion). These latter approaches, although interesting, are less relevant to the traffic safety domain considered in this paper.

The limited amount of available training data precludes the use of modeling approaches that compute joint probability distributions over time intervals explicitly. Indeed, most pedestrian motion models consist of states that correspond to single time steps and are first-order Markovian. This potentially limits their expressiveness and precision. In contrast, Black and Jepson [14] described an extension of particle filtering to incrementally match trajectory models to input data. It is used for motion classification of 2-D gestures and expression. Sidenbladh *et al.* [15] added an efficient tree search in the context of articulated 3-D human pose recovery. Käfer *et al.* [16] applied this technique to vehicle motion prediction, utilizing the quaternion-based rotationally invariant longest common subsequence (QRLCS) metric for trajectory matching. Keller *et al.* [17] combined positional and optical flow features in the QRLCS matching to perform pedestrian path prediction and action classification (continue walking versus stopping at the curbside) from a vehicle. One of their findings is that humans are still better at this action classification task than the systems considered. Following up on the analysis of pedestrian intention at the curbside, Köhler *et al.* [18] addressed the continue-standing versus starting-to-walk classification task, from a stationary monocular camera. They combined a motion contour image based histogram of oriented gradient (HOG)-like descriptor with a linear support vector machine (SVM). Chen *et al.* [19] proposed a multilevel prediction model, in which the higher levels are long-term predictions based on trajectory clustering matching, whereas the low level uses an autoregressive model to predict the next time step.

A common assumption when dealing with human motion is that measurements in a high-dimensional space can be represented in a low-dimensional nonlinear manifold. Nonlinear dimensionality reduction methods allow learning the internal model of the data (see van der Maaten *et al.* [20] for an overview of techniques). It often depends on the data and the task at hand (e.g., visualization and classification) which of the techniques is best suited. Because measurements from human motions are time dependent, it is desirable to consider the dependence of the data over time. The Gaussian process latent variable model [21], which is a generalization of the probabilistic principal component analysis (PCA) [22], can be extended to model the

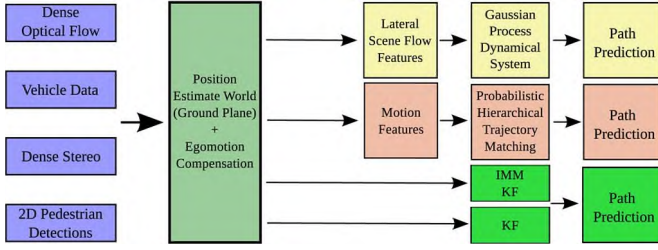


Fig. 2. Overview of considered approaches for pedestrian path prediction.

dynamics of the data. This GPDM [3] allows for nonlinear mapping from the latent space to the observation space, as well as a smooth prediction of latent points. In particular, the mapping (or prediction) of data points on the latent space makes this technique interesting for tracking application. Urtasun *et al.* [23] used a GPDM to track a small number of 3-D body points that have been derived using an image-based tracker. The system is trained using one gait cycle from six subjects and is able to handle several frames of occlusions. Andriluka *et al.* [24] used a dynamic part based limb detector in combination with a GPDM to allow robust detection and tracking in complex scenes with many persons and long-term occlusions. Raskin *et al.* [25] used a GPDM with an articulated model of the human body in combination with an annealed particle filter for tracking and action classification. Action classification is realized by comparing observed sequences with template sequences in latent space.

The contributions of this paper are as follows. We propose two novel approaches that use augmented features derived from dense optical flow for pedestrian path prediction, one GPDM based (see Section III-A) and the other PHTM based (see Section III-B). Furthermore, we present an experimental study on pedestrian path prediction (including baseline KF approaches), based on real video data from a vehicle and actual sensor processing, as opposed to simulated data. Given the documented benefit of dense stereo for pedestrian sensing [26], we use it as input to our approaches. Section III-B is based on our earlier work [17].

III. GENERAL FRAMEWORK

We compare four different approaches for pedestrian path prediction, involving GPDMs, PHTM, KFs, and IMM KFs. For an overview, see Fig. 2.

To allow meaningful comparisons among the systems, several preprocessing components are set equal. Bounding boxes containing pedestrians are supplied from the same detector module. Dense disparity is computed using the semiglobal matching stereo algorithm [27]. Pedestrian positions on the ground plane are obtained by considering the midpoint of the bounding box and the disparity computed over the part of the bounding box that corresponds to the upper body (assuming typical human proportions). The latter involves clustering disparity values using mean shift [28] and selecting the cluster with the largest weight; the median of the corresponding disparity values provides the desired pedestrian distance.

Vehicle egomotion is compensated by rotation and translation of pedestrian positions to a global reference point using a

single-track vehicle model [29] and velocity and yaw rate measurements from onboard sensor data. The two approaches that use augmented visual features (GPDM and PHTM) compute dense optical flow [30] over the bounding boxes provided by the pedestrian detector; this flow is, subsequently, egomotion compensated.

A. GPDM System

The first approach uses scene flow features describing the lateral movement of the pedestrian derived from the dense optical flow field and measured pedestrian distance in the world. Feature dimensionality is reduced by means of a GPDM [3] with a dynamic model in the latent space. To overcome the absence of direct mapping from feature space to latent space, the dynamic model is combined with a particle filter. GPDMs that capture the walking and stopping movements of a pedestrian are separately trained. The learned dynamical models provide optical flow fields at future time instants; future lateral positions can be derived by integration. Longitudinal position is independently estimated by means of a separate KF for each action class (walking versus stopping). Weighting lateral and longitudinal predictions using the probability of each action model results in future pedestrian positions.

1) *Feature Extraction:* Given the lateral component from dense optical flow and a pedestrian distance derived from dense stereo, the lateral velocity of a pedestrian in the world is computed.

With the pedestrian distance (as disparity $disp$), the horizontal component of the optical flow field V_u , the camera base width b , and the camera cycle time Δt , the lateral speed $v_X(m/s)$ of each pixel is computed using

$$v_X = \frac{V_u \cdot b}{disp \cdot \Delta t}. \quad (1)$$

To obtain only flow values located on the pedestrian body, a mask image is generated from the thresholded disparity image, and velocity values corresponding to the background are set to zero. Applying this distance mask also adds rough pedestrian contour information to the feature. Fig. 3 describes the feature extraction steps.

For further use as a feature, this scene flow image is rescaled to 32×16 pixel and concatenated to a feature vector $\mathbf{y}_t \in \mathbb{R}^D$, with $p = 512$. From the scene flow image ($SFlowX$), the lateral velocity of the pedestrian can be directly extracted using the median of velocity values located in the area of the pedestrian upper body (see red box in Fig. 3).

2) *Dynamical Model:* We are interested in a low-dimensional representation $\mathbf{x}_t \in \mathbb{R}^d$ of features $\mathbf{y}_t \in \mathbb{R}^D$ from a pedestrian sequence with $d < D$. This dimensionality reduction is realized using the GPDM [3], [23], [31], which allows modeling the dynamics of the features over time t in the low-dimensional space. For data in latent space $\mathbf{x}_t$, the relation to the input $\mathbf{y}_t$ can be described using

$$\mathbf{y}_t = g(\mathbf{x}_t; \mathbf{B}) + \mathbf{n}_{y,t} \quad (2)$$

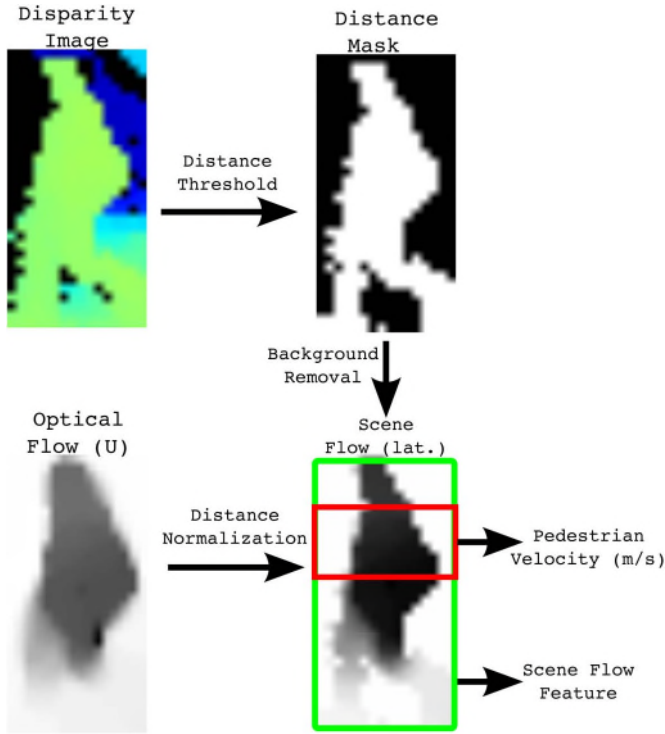


Fig. 3. Feature extraction using dense optical flow and roughly estimated pedestrian contour from dense stereo.

with zero-mean Gaussian noise $\mathbf{n}_{y,t}$ and mapping function g with parameters $\mathbf{B} = [b_1, b_2, \dots]$. Assuming a first-order Markov model, the dynamics of the data in the latent space $\mathbf{x}_1, \dots, \mathbf{x}_t, \dots, \mathbf{x}_N$ can be described using

$$\mathbf{x}_t = f(\mathbf{x}_{t-1}; \mathbf{A}) + \mathbf{n}_{x,t} \quad (3)$$

with zero-mean Gaussian noise $\mathbf{n}_{x,t}$ and mapping function f with parameters $\mathbf{A} = [a_1, a_2, \dots]$.

In a Gaussian process framework, the parameters and basis functions of f and g are marginalized out, and the positions of the latent coordinates are optimized.

a) Latent mapping: In the GPDM framework, the conditional density for the data $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N]^T$ given latent positions $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]^T$ is described using

$$p(\mathbf{Y}|\mathbf{X}, \bar{\beta}, \mathbf{W}) = \frac{|\mathbf{W}|^N}{\sqrt{(2\pi)^{ND} |\mathbf{K}_Y|^D}} \exp \left(-\frac{1}{2} \text{tr} (\mathbf{K}_Y^{-1} \mathbf{Y} \mathbf{W}^2 \mathbf{Y}^T) \right) \quad (4)$$

with kernel matrix $\mathbf{K}_Y$ and kernel hyperparameters $\bar{\beta} = \{\beta_1, \beta_2, \dots\}$ and $\mathbf{W}$. To equally weight all the feature dimensions, the scale parameter is set to $\mathbf{W} = \mathbf{I}$ and is omitted in the following equations. Entries in the kernel matrix are defined using a kernel function $(\mathbf{K}_Y)_{i,j} = k_Y(\mathbf{x}_i, \mathbf{x}_j)$. For our data, we use a radial basis function (RBF) kernel with an additional noise term, i.e.,

$$k_Y(\mathbf{x}_i, \mathbf{x}_j) = \beta_1 \exp \left(-\frac{\beta_2}{2} \|\mathbf{x}_i - \mathbf{x}_j\|^2 \right) + \frac{\delta_{\mathbf{x}_i, \mathbf{x}_j}}{\beta_3}. \quad (5)$$

b) Dynamic mapping: The dynamics of the time series data is incorporated using

$$p(\mathbf{X}|\bar{\alpha}) = \frac{p(\mathbf{x}_1)}{\sqrt{(2\pi)^{(N-1)d} |\mathbf{K}_X|^d}} \times \exp \left(-\frac{1}{2} \text{tr} (\mathbf{K}_X^{-1} \mathbf{X}_{2:N} \mathbf{X}_{2:N}^T) \right) \quad (6)$$

with $\mathbf{X}_{2:N} = [\mathbf{x}_2, \dots, \mathbf{x}_N]^T$, the kernel matrix $\mathbf{K}_X$ constructed from $\mathbf{X}_{1:N-1} = [\mathbf{x}_1, \dots, \mathbf{x}_{N-1}]^T$ with dimensionality $(N-1) \times (N-1)$ and entries $(\mathbf{K}_X)_{i,j} = k_X(\mathbf{x}_i, \mathbf{x}_j)$. A combination of an RBF and a linear kernel with an additional noise is used for the dynamics

$$k_X(\mathbf{x}_i, \mathbf{x}_j) = \alpha_1 \exp \left(-\frac{\alpha_2}{2} \|\mathbf{x}_i - \mathbf{x}_j\|^2 \right) + \alpha_3 \mathbf{x}_i^T \mathbf{x}_j + \alpha_4^{-1} \delta_{\mathbf{x}_i, \mathbf{x}_j} \quad (7)$$

with kernel hyperparameters $\bar{\alpha} = \{\alpha_1, \alpha_2, \dots\}$.

c) Learning the GPDMs: Combining the latent and dynamics mapping defines the model

$$p(\mathbf{X}, \mathbf{Y}, \bar{\alpha}, \bar{\beta}) = p(\mathbf{Y}|\mathbf{X}, \bar{\beta}) p(\mathbf{X}|\bar{\alpha}) p(\bar{\alpha}) p(\bar{\beta}). \quad (8)$$

Learning a GPDM requires finding the latent positions $\mathbf{X}$ and kernel hyperparameters $\mathcal{H} = \{\bar{\alpha}, \bar{\beta}\}$ with respect to the features $\mathbf{Y}$ by minimizing the negative log posterior $-\ln p(\mathbf{X}, \mathcal{H}|\mathbf{Y})$. Minimization can be done using a scaled conjugated gradient (SCG) method [31]. This requires the inverse of the kernel matrix with a complexity of $O(N^3)$ in each optimization iteration. We select $d = 3$ as the latent space dimensionality.

It is difficult to learn a generic GPDM that captures large variations in the data and different motions. Combining trajectory data in which the pedestrian is walking and data in which the pedestrian is stopping results in degenerated models. Urtasun *et al.* [32] introduced additional constraints to prevent the degeneration of models. Selecting the correct constraints for a model that captures the walking and stopping motions of a pedestrian for the used features is difficult, particularly with noisy data. Additionally, the complexity when training the model is increased. To avoid these problems, we train two separate models. The first model is trained using trajectory data segments in which pedestrians are walking. Stopping situations are selected to train the second model. Because the beginning of a stopping action is difficult to define, data from 20 frames (0.91 s) before the stopping of the pedestrians is used. Examples of the two models are plotted in Fig. 4.

d) Mean prediction: With the learned dynamic model, a point $\mathbf{x}$ in the latent space is predicted, and the most likely successor is derived using

$$\mu_{\mathbf{X}}(\mathbf{x}) = \mathbf{X}_{2:N}^T \mathbf{K}_X^{-1} k_{\mathbf{X}}(\mathbf{x}) \quad (9)$$

with the vector $k_{\mathbf{X}}(\mathbf{x})$ containing at the i th entry the results of $k_X(\mathbf{x}, \mathbf{x}_i)$ using training sample $\mathbf{x}_i$.

Fig. 4 illustrates this mean prediction of a point for several frames on the low-dimensional space.

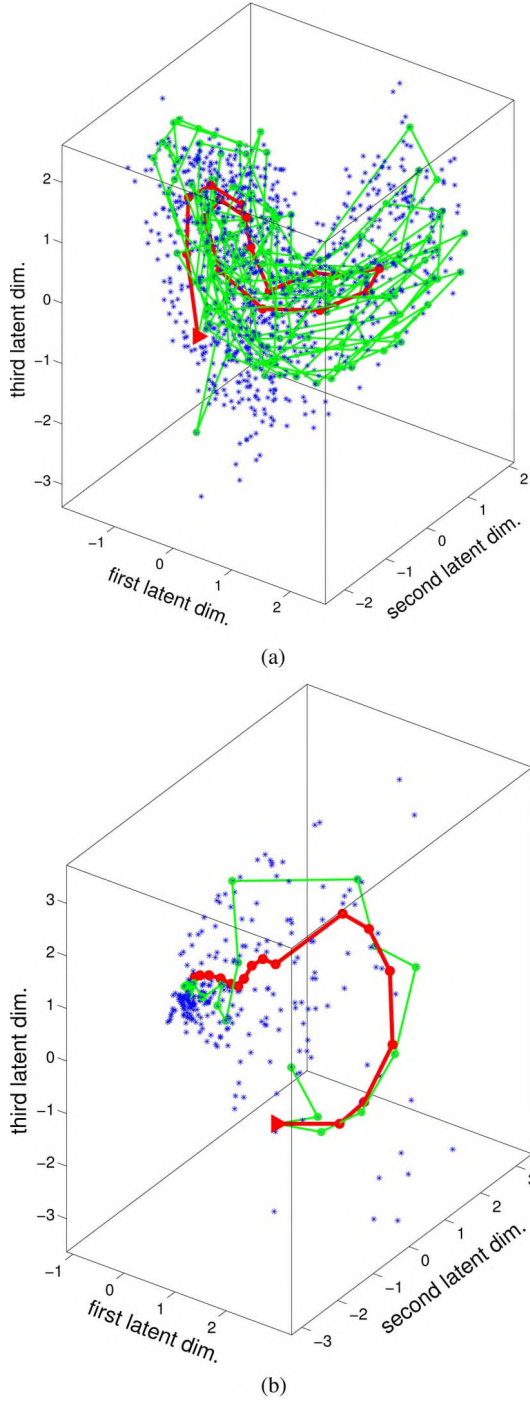


Fig. 4. Traversal of a (•) training trajectory through the (*) learned latent space and (o) mean predictions of a (►) point for 17 frames (0.77 s). Figures depict (a) the walking case and (b) the stopping case. All available training samples are shown.

e) *Latent reconstruction*: A point $\mathbf{x}$ in the latent space is reconstructed using

$$\mu_{\mathbf{Y}}(\mathbf{x}) = \mathbf{Y}^T \mathbf{K}_{\mathbf{Y}}^{-1} k_{\mathbf{Y}}(\mathbf{x}). \quad (10)$$

An example of the reconstructed scene flow feature is shown in Fig. 5.

3) *Multiple Model Particle Filter*: The state of a pedestrian at time t is described using $\phi_t = [\mathbf{x}_t, \mathcal{X}_t]$, where $\mathbf{x}_t \in \mathcal{R}^d$ is a

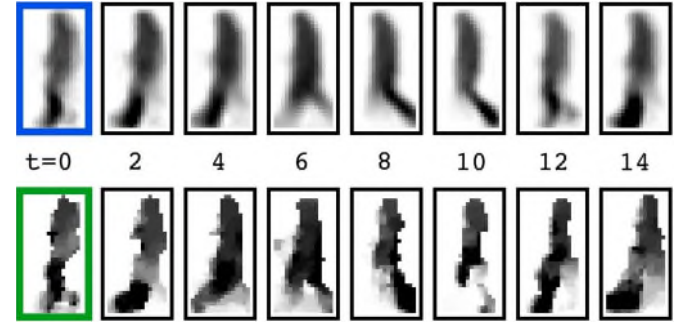


Fig. 5. (Top row) Reconstructed optical flow based on current state ($t = 0$) and state predictions ($t = 2, \dots, 14$) in low-dimensional latent space. (Bottom row) Optical flow that is (will be) actually measured at the corresponding time steps.

point in the low dimensional space and $\mathcal{X}_t$ the lateral pedestrian position in the world. Given an observed motion feature $\mathbf{y}_t$ and observed lateral position $\mathcal{Y}_t$, the probability of a pedestrian state ϕ_t is computed by

$$p(\phi_t | \mathbf{y}_t, \mathcal{Y}_t) = \eta p(\mathbf{y}_t, \mathcal{Y}_t | \phi_t) \int p(\phi_t | \phi_{t-1}) p(\phi_{t-1} | \mathbf{y}_{t-1}, \mathcal{Y}_{t-1}) d\phi_{t-1} \quad (11)$$

with normalization constant η . The probability $p(\phi_t | \phi_{t-1})$ of observing a future state is computed from the GPDM latent space mean prediction.

This distribution is represented by a set of particles $\{\phi_t^{(s)} : s \in \{1, \dots, S\}\}$ with corresponding weight $w_t^{(s)}$ that is propagated using a particle filter. Particles are predicted using the learned GPDM model with the predicted state $\hat{\phi}_t^{(s)} = [\hat{\mathbf{x}}_t^{(s)}, \hat{\mathcal{X}}_t^{(s)}]$ with $\hat{\mathbf{x}}_t^{(s)} = \mu_{\mathbf{X}}(\mathbf{x}_{t-1}^{(s)}) + n_x$, and $\hat{\mathcal{X}}_t^{(s)} = \mathcal{X}_{t-1}^{(s)} + s_{\mathcal{X}}(\hat{\mathbf{y}}_t^{(s)}, \Delta t)$ which computes the traveled distance from the mean velocity derived from the reconstructed scene flow image $\hat{\mathbf{y}}_t^{(s)} = \mu_{\mathbf{Y}}(\hat{\mathbf{x}}_t^{(s)})$ and camera cycle time Δt . For each particle, the noise term n_x is randomly sampled from $\mathcal{N}(\mathbf{0}, I_d \times \sigma_{n_x}^2)$, with an experimentally derived $\sigma_{n_x} = 0.1$.

Scene flow feature similarity is computed using the Euclidean distance

$$d_f(\mathbf{y}, \hat{\mathbf{y}}) = \|\mathbf{y} - \hat{\mathbf{y}}\|^2. \quad (12)$$

For the lateral position in the world, the distance $d_p(\mathcal{Y}, \mathcal{X})$ is computed with

$$d_p(\mathcal{Y}, \mathcal{X}) = \|\mathcal{Y} - \mathcal{X}\|^2. \quad (13)$$

Using the distances between the observed and predicted data, the observation likelihood $p(\mathbf{y}_t, \mathcal{Y}_t | \phi_t^{(s)}) \propto w_t^{(s)}$ is approximated using

$$w_t^{(s)} = \exp \left(-\frac{d_f(\mathbf{y}_t, \hat{\mathbf{y}}_t^{(s)})^2}{2\sigma_f^2} - \frac{d_p(\mathcal{Y}_t, \hat{\mathcal{X}}_t^{(s)})^2}{2\sigma_p^2} \right) \quad (14)$$

with an empirically estimated $\sigma_f = 7$ for the feature similarity and $\sigma_p = 0.06$ for the deviation of the lateral position. The

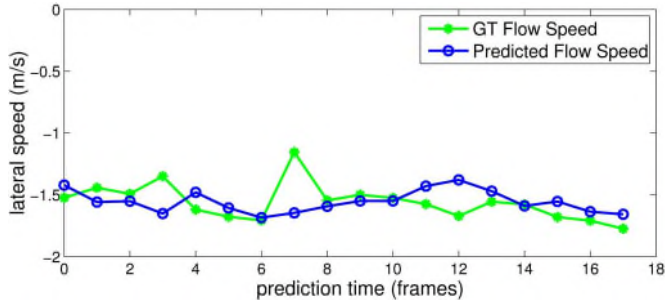


Fig. 6. Predicted speed derived from (o) predicted optical flow and corresponding (+) measured optical flow speed for different prediction horizons.

updated $\phi_t^{(s)}$ is obtained from $\hat{\phi}_t^{(s)}$ by reweighting the particle set.

For efficiency, an estimated state ϕ_T representing the pedestrian state in the future $T = t + \Delta T$ is derived from the weighted mean $\mathbf{x}^*$ of the particle set $\{\phi_t^{(s)}\}$ and iteratively predicted using $\mu_{\mathbf{x}}(\mathbf{x}^*)$. From the reconstructed predicted scene flow data (see Fig. 5), the pedestrian velocity in the future is computed (see Fig. 6). Integrating over the velocity predictions results in the predicted pedestrian position.

As mentioned in Section III-A2c, we trained two models for the different pedestrian motions. Models are combined using an interacting multiple model particle filter (*MM-PF*) similar to [33]. For each model a fixed number of particles ($S = 200$) is used to represent the state. From the set of particles M_i in model i the model probability is derived using

$$\gamma_i(t) = \frac{\sum_{\phi_t^{(s)} \in M_i} w^{(s)}}{\sum_{\phi_t^{(s)} \in M_1} w^{(s)} + \sum_{\phi_t^{(s)} \in M_2} w^{(s)}}. \quad (15)$$

Model probabilities are updated similar to the *IMM-KF* scheme, described in Section III-C. The conditional probability γ_{ij} of a transition from model i to j is computed using

$$\gamma_{ij}(t) = \frac{\Psi_{ij} \cdot \gamma_i(t)}{\sum_{k=1}^2 \Psi_{kj} \cdot \gamma_k(t)} \quad (16)$$

with the state transition matrix Ψ .

We assume the lateral and longitudinal pedestrian dynamics to be weakly dependent. Longitudinal state estimation is decoupled, and to each of the lateral models (GPDM), a KF with a corresponding constant velocity (CV) or constant position (CP) model is assigned to track the position. Longitudinal positions $s_z^i(\Delta t)$ are linearly predicted with the estimated velocity of each filter. Mixing the lateral model predictions $s_x^i(\Delta t)$ and the longitudinal KF prediction $s_z^i(\Delta t)$ with the state transition probabilities at t results in the pedestrian position

$$s_{x,z}(\Delta t) = \sum_{i=1}^2 \gamma_i \cdot \begin{pmatrix} s_x^i(\Delta t) \\ s_z^i(\Delta t) \end{pmatrix}. \quad (17)$$

In the following, the approach using the GPDMs in combination with scene flow features is abbreviated with *SFlowX/GPDM*.

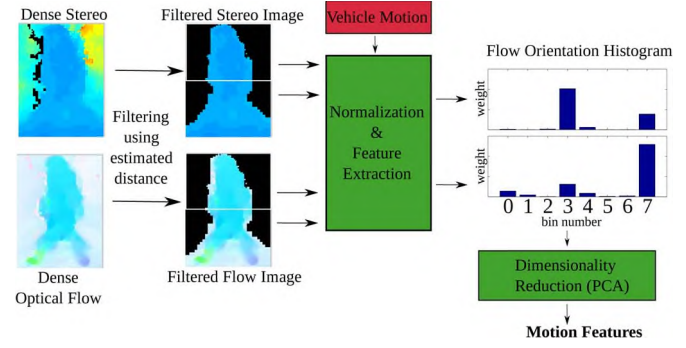


Fig. 7. Motion feature extraction in the PHTM-based system.

B. PHTM System

The second approach uses motion features involving a low-dimensional histogram representation of optical flow. Measured pedestrian positions and motion features are subsequently used in a trajectory matching and filtering framework. From the filter state, a future pedestrian position is derived by looking ahead on matched trajectories of the training set.

1) *Motion Features*: The low-dimensional feature captures flow variations on the pedestrian legs and upper body. In order to operate from a moving vehicle, additional invariance to pedestrian distance and vehicle motion is important. Features are designed to allow bounding box localization errors from a pedestrian detection system. Fig. 7 illustrates the feature extraction steps. Flow vectors are normalized with the camera cycle time to account for asynchronous capture and frame drops. Flow vectors are further normalized with measurements from dense stereo for invariance to different pedestrian distances. The resulting normalized motion field is used to extract features given a bounding box detection and distance estimation z_{ped} from a pedestrian detection system. To ensure that the pedestrian is located in the box for all possible limb extensions and slight localization errors, a bounding box aspect ratio of 4:3 is used. Motion vectors not belonging to the pedestrian body are suppressed by using only values at a depth similar to the estimated pedestrian distance. Remaining values in the motion field are used to compute the median object motion and extract orientation histograms. To capture motion differences between torso and legs, the bounding box is split into upper and lower subboxes. For each subbox, the median motion is removed to compensate the pedestrian egomotion. Resulting orientation vectors $v = [v_x, v_y]^T$ are assigned to bins $b \in [0, 7]$ using their 360° orientation $\theta = \text{atan2}(v_y, v_x)$ and bin index $b = \lfloor \theta/\pi/4 \rfloor$. Bin contributions are weighted by their magnitude, and resulting histograms are normalized with the number of contributions. A feature vector is formed by concatenating the histogram values and the median flow for the lower and upper boxes. Dimensionality reduction of the feature vector is achieved by applying PCA. The first three PCA dimensions with the largest eigenvalue are used as final histogram of orientation motion (*HoM*) features.

2) *Trajectory Matching*: A pedestrian trajectory Ω is represented using the ordered tuples $\Omega = ((\omega_1, t_1), \dots, (\omega_N, t_N))$. For every time stamp t_i , the pedestrian state ω_i consists of the lateral and longitudinal positions of the pedestrian and

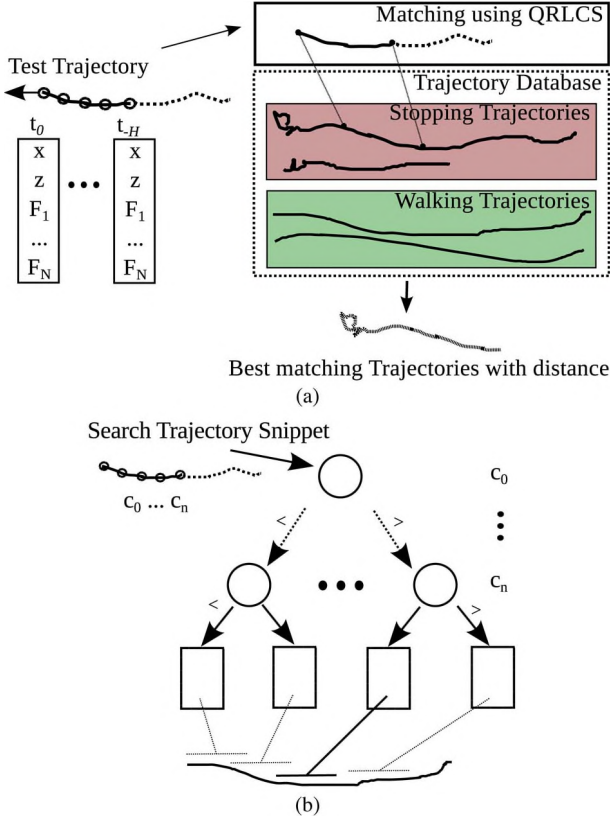


Fig. 8. (a) Test trajectory with a history of length H containing position and feature information for every entry is matched to the training database. Resulting matching position and similarity distance to trajectories in the training database describe a possible trajectory course and class label. (b) Tree representation of the trajectory training database. Leaf nodes represent trajectory snippets of fixed length. Similar trajectories are searched by traversing the tree using the trajectory descriptors for every level.

additional features extracted from optical flow [see Fig. 8(a)]. For path prediction, it is possible to compare an observed test trajectory with a history of H pedestrian states to each trajectory in a training database using a similarity measure. With the QRLCS metric [16], the optimal translation and rotation parameters to superimpose two trajectories are derived. The distance $\text{dist}_{\text{QRLCS}}(\Omega_i, \Omega_j) \in [0, 1]$ between two trajectories is given by the number of possible assignments determined by an ε area around each pedestrian state, normalized by the number of pedestrian states. Fig. 8(a) illustrates this comparison process.

We replace this exhaustive search by a probabilistic search framework [15], [16]. A set of overlapping subtrajectories (snippets, e.g., [34]) with a fixed history of pedestrian states is created from a training database. Information of the snippet position in the origin trajectory and successor snippets is kept for later use. By piling the features for each state in a snippet into a description vector and applying the PCA method to these vectors, their principal dimensions can be ordered according to the largest eigenvalue. The resulting transformed description vector $\mathbf{c}$ is used to build a binary search tree. For each level l , the snippet is assigned to the left or right subtree depending on the sign of c_l . Given N training snippets, the depth of the search tree, i.e., n , is $O(\log(N))$. Fig. 8(b) illustrates this search tree.

Given a trajectory $\Omega_{1:t}$, the probability of the state ϕ_t is computed by

$$p(\phi_t|\Omega_{1:t}) = \eta p(\Omega_{1:t}|\phi_t) \int p(\phi_t|\phi_{t-1})p(\phi_{t-1}|\Omega_{1:t-1})d\phi_{t-1} \quad (18)$$

with a normalization constant η . The distribution $p(\phi_t|\Omega_{1:t})$ is represented by a set of samples or particles $\{\phi_t^{(s)}\}$, which are propagated in time using a particle filter [14]. Each particle $\phi_t^{(s)}$ represents a snippet describing a pedestrian state with a history and an assigned likelihood. Our transition model $p(\phi_t|\phi_{t-1})$ is determined by a probabilistic search in the binary tree. Particle prediction is performed by a probabilistic search in the constructed binary tree and a lookup for the successor snippet in the training database. The distribution $p(\Omega_{1:t}|\phi_t)$ represents the likelihood that the measurement trajectory $\Omega_{1:t}$ can be observed given the current state. In the context of particle filters, this value corresponds to the weight of a particle and is approximated using $w^{(s)} = 1 - \text{dist}_{\text{QRLCS}}$ for each particle $\phi_t^{(s)}$.

An estimated state $\phi_T^{(s)}$ representing the pedestrian state in the future $T = t + \Delta T$ can be derived by looking ahead on the associated origin trajectory for the current state $\phi_t^{(s)}$. This results in many hypotheses, which are compensated using a weighted mean shift algorithm [28] with a Gaussian kernel and weights $w^{(s)} \sim p(\phi_T^{(s)}|\Omega_{1:t})$. At the final predicted state ϕ_T^* , the cluster center with the highest accumulated weight is selected.

The trajectory database contains two classes of trajectory snippets: the class $\mathcal{C}_s$, in which the pedestrian is stopping, and the class $\mathcal{C}_w$, in which the pedestrian continues walking. For the predicted object state ϕ_T^* derived using cluster members $L = \{\phi_t^{(l)}\}$ and the corresponding weight $w^{(l)}$, the stopping probability can be approximated using

$$p(\mathcal{C}_s|L) \approx \frac{\sum_{\phi_t^{(l)} \in \mathcal{C}_s} w^{(l)}}{\sum_{\phi_t^{(l)} \in \mathcal{C}_s} w^{(l)} + \sum_{\phi_t^{(l)} \in \mathcal{C}_w} w^{(l)}}. \quad (19)$$

In the following, the PHTM approach using HoM features is abbreviated with *HoM/Traj*.

C. KF-Based Systems

1) *KF*: As a third approach, a linear KF [4] is used. The state $\hat{\mathbf{X}}$ of the filter is modeled as

$$\hat{\mathbf{X}} = [z \quad x \quad v_z \quad v_x]^T$$

with z/x being the longitudinal/lateral position of the pedestrian to the vehicle and v_z/v_x being its absolute longitudinal/lateral velocity in the world. Pedestrian positions are pseudomeasurements provided by the stereo pedestrian detection component, as described at the beginning of this section. A CV model is assumed as a pedestrian motion model. Using this model means that all deviations from a constant pedestrian motion have to be captured as process noise. With the assumption that a pedestrian moving at 1.8 m/s can stop in 1 s, we select a process noise parameter $q = 1.8$ for the filter.

2) *IMM-KF*: The fourth approach extends the previous KF with an additional CP model; this way, the *IMM-KF* [4] is realized. The basic idea is to maintain a KF for each possible

motion model with state $\hat{\mathbf{x}}_j(t)$ and model probability $\gamma_j(t)$. This means that a steady walking pace is represented using a filter with the CV model with process noise parameter q_{CV} . For nonmoving pedestrians, the CP model with q_{CP} applies. Each iteration consists of three steps: interaction, filtering, and mixing. The interaction step computes the mixing probability γ_{ij} from the current model probability γ_j and the state transition probability Ψ_{ij} [see (16)]. From the mixing probability, the mixed state mean $\hat{\mathbf{x}}_{0j}(t)$ and the covariance matrix $\hat{\mathbf{P}}_{0j}(t)$ are computed as initial input for each filter in the filtering step using

$$\hat{\mathbf{x}}_{0j}(t) = \sum_{i=1}^2 \hat{\mathbf{x}}_i(t) \gamma_{ij}(t). \quad (20)$$

For the computation of $\hat{\mathbf{P}}_{0j}(t)$, see [4]. In the filtering step, a *KF* predict/update step is done using the mixed state mean $\hat{\mathbf{x}}_{0j}(t)$ and covariance matrix $\hat{\mathbf{P}}_{0j}(t)$ derived in the interaction step. Given the likelihood function $\Lambda_j(t+1) = \mathcal{N}(r_j(t+1), \mathbf{S}_j(t+1))$ with residuum $r_j(t+1)$ and residual covariance $\mathbf{S}_j(t+1)$, the updated probabilities $\gamma_j(t+1)$ are computed using

$$\gamma_j(t+1) = \frac{1}{c} \Lambda_j(t+1) \sum_{i=1}^2 \Psi_{ij} \gamma_i(t) \quad (21)$$

with normalization factor c . An approximation of the resulting mixture model is then computed in the mixing step using

$$\hat{\mathbf{x}}(t+1) = \sum_{i=1}^2 \hat{\mathbf{x}}_i(t+1) \gamma_i(t+1). \quad (22)$$

For the following evaluation, $q_{CV} = 0.21$ and $q_{CP} = 0.41$ have been derived from the set of training trajectories, with respect to the positions minimum root-mean-square error (RMSE). The matrix Ψ describing the transition probabilities between the CV and CP models has been experimentally derived from the available training data $\Psi = [0.999, 0.001, 0.001, 0.999]$. Choosing larger values for the model transitions results in more frequent undesired switches, particularly with noisy measurements. The *IMM-KF* is said to be nonsensitive to improperly selected transition probabilities [35].

IV. EXPERIMENTS

Video data of two scenarios (see Fig. 1) were recorded using a stereo camera system (baseline 30 cm, 22 fps) mounted behind the windshield of a vehicle. The first scenario features the stopping of a pedestrian at the curbstone. In the second scenario, the pedestrian crosses the street. In both scenarios, the pedestrian was not occluded. In some test runs, the vehicle is stationary, whereas in others, the vehicle is moving at speeds of 20–30 km/h. The data set involved four different pedestrians in three different locations at a distance range of 5–34 m to the vehicle. Table I provides some further statistics on the data set. Fig. 9 illustrates some test images.

The GT locations of the pedestrians in the world were obtained by manual labeling the pedestrian shapes in the images. The median disparity value on the pedestrian upper body and

TABLE I
NUMBER OF SEQUENCES WITH DIFFERENT PEDESTRIAN
AND VEHICLE ACTIONS IN OUR DATA SET

Sequences	vehicle standing	vehicle moving	vehicle standing+moving
ped. stopping	11	5	16
ped. walking	9	4	13



Fig. 9. Example images from the data set showing the pedestrian action. Images show the labeled (left) stopping or (right) walking moment.

the center foot point of the shape is used to obtain the longitudinal and lateral positions on the ground plane. In terms of alignment along the time axis, for each trajectory in which the pedestrian is stopping, the moment of the last placement of the foot is labeled as the stopping moment. The time-to-stop (TTS) value counts the number of frames until this event; frames earlier to the stopping event have positive TTS values; frames after the stopping event have negative TTS values. In sequences in which the pedestrian continues walking, the closest point to the curbstone (with closed legs) is labeled. Analogous to the TTS definition, the latter is called the time-to-curb (TTC) value.

Performance evaluation is done using input data with different noise characteristics with regard to the image bounding box positions. Two-dimensional bounding boxes derived from manually labeled pedestrian shapes (termed *label box*) are used as the most accurate input data for feature extraction and localization; it reflects the case of an “ideal” pedestrian detector. We further consider the case in which these ideal 2-D bounding boxes are perturbed by uniform noise; we add up to 10% of the original height of the bounding boxes to their height and center (the resulting bounding boxes are termed *jittered*). Finally, we consider 2-D bounding boxes provided by a state-of-the-art HOG/linSVM pedestrian detector [36] (termed *system detections*). Hereby, detection “gaps” are filled in by means of a standard correlation tracker. Considering data with artificial noise allows abstracting away from the noise bias of a particular pedestrian detector. As we will see shortly, the overall noise level artificially added is realistic, in the sense that it is similar to that of a state-of-the-art detector.

The lateral and longitudinal position errors on the ground plane for different input data are summarized in Table II. In these experiments, we compared with a smoothed version of

TABLE II
MEAN DEVIATION (IN METERS) OF THE PEDESTRIAN POSITION
ON THE GROUND PLANE (LATERAL AND LONGITUDINAL)
COMPARED WITH THE SMOOTHED GT DATA

	veh. standing+moving	
	lat.	long.
GT	0.03	0.10
label box	0.05	0.22
jittered box	0.13	0.68
sys. detections	0.06	0.64

the GT ground plane positions. GT positions from walking trajectories, in which we are certain that the pedestrian is moving with an approximately constant velocity ($-40 \leq \text{TTC} \leq 40$), were fitted with a curvilinear model, minimizing pedestrian velocity and yaw changes by nonlinear least squares. For the stopping trajectories, smoothing was only applied to the cases in which the pedestrian is standing ($\text{TTS} \leq 0$), by simple averaging. Note that smoothed GT was only used for the purpose of Table II. In the path prediction experiments, comparisons involved the nonsmoothed GT. Following observations can be made from Table II. As expected, the longitudinal error is larger than the lateral error due to stereo vision characteristics. Adding aforementioned uniform noise on 2-D bounding boxes results in degradation of positional accuracy of about 10 and 50 cm in lateral and longitudinal directions, respectively. Positional errors are similar for the artificial noise and real detector case.

A. Parameter Settings and Evaluation Setup

We compare the four approaches using equal parameter settings, whenever possible. Lateral and longitudinal noise parameters for the *KF* and *IMM-KF* and longitudinal noise parameters for tracking the distance of the *SFlowX/GPDM* are selected from Table II. Process noise parameters and state transition matrix were heuristically derived (see Section III-C). The same state transition matrix is used for the *MM-PF* of the *SFlowX/GPDM* system and the *IMM-KF*.

The analysis of walking trajectories showed an average gait cycle of 10–14 frames for different pedestrians. The trajectory database for the *HoM/Traj* contains subtrajectories, generated in a sliding window fashion, with a fixed length of ten frames. For test trajectories, a history of 14 frames is used to capture gait cycle variations. Approximating the current probability density is done with $S = 400$ particles and a tree search deviation parameter of $\beta = 0.05$. The mean shift position procedure operates with a kernel width value $h = 0.1$.

Training and testing data have been processed using *leave-one-out* cross-validation. This means that one sequence is used for testing and the remaining training sequences are used to learn the GPDM models *SFlowX/GPDM* or for search tree generation (*HoM/Traj*).

B. Pedestrian Path Prediction

We are interested in the ability of each system to predict future pedestrian positions accurately. Tables III and IV list ground plane localization accuracy (i.e., longitudinal and lateral dimensions combined) at different prediction horizons, for each

system. Localization accuracy is measured in terms of the mean and standard deviation of the per-sequence RMSE. Per-sequence RMSE is determined by comparing system predictions at various time horizons with the GT, when the pedestrian is inside the frame range $[20, -10]$, when frame 0 denotes the manually labeled TTS/TTC moment. This corresponds to an evaluation time range of $[0.91, -0.45]$ s around the TTS/TTC event. Pedestrian positions are predicted up to 17 frames (0.77 s) into the future. Tables III and IV list the results for walking and stopping trajectories, respectively. Results are further differentiated based on whether the own vehicle is standing or moving or whether all data are used (cf., Table I).

On walking scenarios (see Table III), all approaches show a similar prediction performance when pedestrian bounding boxes are set precise (*label box*). In the more realistic case of inaccurate image localization (*jittered box*) and moving vehicle, we see that *HoM/Traj*, unlike the other approaches, shows no performance degradation and thus gains a slight edge. We attribute this to the robustness of trajectory matching to outliers in the longitudinal dimension.

On stopping scenarios (see Table IV), in which the constant velocity assumption is violated, incorporation of motion features leads to a path prediction performance advantage of up to a factor of 2 for *HoM/Traj* and *SFlowX/GPDM* compared with the *KF*-based variants (e.g., *jittered* data and vehicle standing and moving cases). As before, the trajectory matching of *HoM/Traj* shows added robustness to noise caused by bounding box position errors and vehicle egomotion.

In the intelligent vehicle pedestrian safety context, the *lateral* component of the localization error is particularly relevant; it determines whether the pedestrian enters the vehicle driving corridor and a collision potentially occurs. Fig. 10 lists the mean lateral localization error at various time offsets to the labeled TTS/TTC moment (*jittered* data and vehicle standing and moving cases). Separate plots are shown depending on the prediction horizon (0 or 17 frames) and whether the pedestrian is walking or stopping. We observe no significant performance difference between walking trajectories [see Fig. 10(a) and (c)]. For stopping scenarios [see Fig. 10(b) and (d)], the advantage of the additional motion model of the *IMM-KF* versus the *KF* becomes visible (in Table IV, this advantage was averaged away over the frame range $[20, -10]$, due to the inclusion of time instants still involving walking). Stopping of the pedestrian leads to a switch to the CP model and a lower localization error compared with the *KF* with CV model. Fig. 10 also shows that *HoM/Traj* and *SFlowX/GPDM* are more quickly able to adjust to the change in the pedestrian motion, resulting in a lower lateral localization error than the *KF*-based approaches. Fig. 11 illustrates the distribution of the lateral prediction error *difference* between *IMM-KF* and *HoM/Traj* for stopping trajectories. Performance differences are clearly visible close to the stopping event $\text{TTS} = 0$.

Results using tracked detections from a state-of-the-art HOG/linSVM pedestrian detector [36] are listed in Table V. For this experiment, we used a subset of 7 walking and 13 stopping trajectories (five trajectories with a moving vehicle) in which the pedestrian detector had a decent performance in the first place (detection gaps no longer than ten frames consecutively).

TABLE III
MEAN COMBINED LONGITUDINAL AND LATERAL RMSE (IN METERS) FOR WALKING TRAJECTORIES
AND DIFFERENT PREDICTION HORIZONS (FRAMES)

		veh. standing						veh. standing + moving						veh. moving			
		label box				jittered box		label box		jittered box		label box		jittered box		label box	
		0	5	11	17	0	17	0	17	0	17	0	17	0	17	0	17
KF	Mean	0.15	0.19	0.22	0.28	0.24	0.35	0.17	0.38	0.24	0.45	0.21	0.62	0.24	0.69		
	$\pm Std$	0.07	0.07	0.09	0.12	0.2	0.2	0.07	0.25	0.17	0.28	0.05	0.29	0.06	0.29		
IMM-KF	Mean	0.14	0.17	0.2	0.25	0.23	0.34	0.19	0.41	0.26	0.5	0.29	0.77	0.33	0.86		
	$\pm Std$	0.07	0.08	0.1	0.12	0.21	0.22	0.1	0.3	0.19	0.34	0.09	0.26	0.09	0.27		
HoM/Traj	Mean	0.13	0.17	0.23	0.29	0.14	0.30	0.14	0.33	0.15	0.35	0.15	0.43	0.15	0.44		
	$\pm Std$	0.03	0.03	0.05	0.07	0.03	0.07	0.03	0.13	0.03	0.11	0.02	0.17	0.02	0.12		
SFlowX/GPDM	Mean	0.15	0.2	0.26	0.34	0.17	0.5	0.17	0.41	0.19	0.52	0.21	0.62	0.25	0.69		
	$\pm Std$	0.04	0.08	0.13	0.18	0.06	0.32	0.05	0.24	0.08	0.31	0.06	0.25	0.08	0.22		

TABLE IV
MEAN COMBINED LONGITUDINAL AND LATERAL RMSE (IN METERS) FOR STOPPING TRAJECTORIES
AND DIFFERENT PREDICTION HORIZONS (FRAMES)

		veh. standing						veh. standing + moving						veh. moving			
		label box				jittered box		label box		jittered box		label box		jittered box		label box	
		0	5	11	17	0	17	0	17	0	17	0	17	0	17	0	17
KF	Mean	0.20	0.36	0.61	0.93	0.27	0.81	0.24	1.04	0.33	1.14	0.32	1.25	0.44	1.39		
	$\pm Std$	0.04	0.06	0.1	0.15	0.14	0.23	0.08	0.26	0.15	0.37	0.1	0.33	0.15	0.51		
IMM-KF	Mean	0.18	0.31	0.55	0.87	0.27	0.77	0.22	0.98	0.32	1.08	0.32	1.19	0.47	1.26		
	$\pm Std$	0.03	0.04	0.07	0.12	0.13	0.24	0.08	0.19	0.16	0.2	0.1	0.17	0.17	0.15		
HoM/Traj	Mean	0.12	0.19	0.34	0.58	0.12	0.56	0.13	0.63	0.12	0.62	0.15	0.74	0.15	0.76		
	$\pm Std$	0.02	0.04	0.11	0.17	0.02	0.10	0.02	0.21	0.02	0.18	0.02	0.23	0.03	0.19		
SFlowX/GPDM	Mean	0.23	0.27	0.35	0.51	0.29	0.62	0.16	0.53	0.23	0.64	0.21	0.66	0.41	0.89		
	$\pm Std$	0.07	0.07	0.06	0.07	0.08	0.22	0.05	0.23	0.26	0.29	0.05	0.32	0.42	0.36		

We observe that using actual system detections, rather than simulated detections, does not change the performance ranking of the approaches considered (compare Table V with the entries in Tables III and IV, in which the vehicle is standing and moving). In fact, performance with actual system detections is similar to that obtained with noise-perturbed GT *jittered* data; this is not surprising given similar per-frame localization measurement error (cf., Table II).

C. Pedestrian Action Classification

We also tested the ability of various systems to classify pedestrian actions, i.e., whether the pedestrian will cross or not. Fig. 12 illustrates the performance of each system on stopping and walking test trajectories; depicted is the estimated probability of stopping, as a function of TTS or TTC. For the *SFlowX/GPDM* and *HoM/Traj* systems, this was achieved by means of (15) and (19), respectively. For *IMM-KF*, stopping was estimated by means of the probability of the CP model, following (21).

To put the performance of the systems in context, we also evaluated human performance. Video data were presented to several test subjects using graphical user interfaces, where playback was automatically stopped at five different TTC or TTS moments (20, 11, 8, 5, and 3). For each run, the test subjects had to decide whether the pedestrian will stop at the curbstone or cross the street and provide a probability (i.e., confidence) using a slider ranging from 0 to 1. Sequence and playback stopping point were randomly selected before being presented to the test subjects to avoid the effect of reidentification.

In Fig. 12, on walking trajectories, all systems show a low and relatively constant stopping probability. On stopping trajectories, all systems initially start with a low stopping probability, since stopping is preceded by walking. However,

within a dozen frames before the stopping event, the stopping probability increases more markedly.

Class membership is determined at each time instant of an input trajectory assigned by thresholding the estimated stopping probability (cf., Fig. 12). Based on the training set, we selected for each system and for the human group a threshold that minimizes its classification error (i.e., stopping classified as walking and vice versa) over all sequences and time instants. Fig. 13 illustrates the resulting classification accuracy over time using these “optimal” thresholds. As can be seen, the humans outperform the various automatic systems at this action classification task. The humans reach accuracy of 0.8 in classifying the correct pedestrian action about 570 ms before the event. This accuracy is only reached about 230 ms before the event by the newly developed *SFlowX/GPDM* and *HoM/Traj* systems, which use augmented visual features. The baseline *IMM-KF* system does worst, reaching the corresponding accuracy only about 90 ms before the event.

V. DISCUSSION

Table V indicates that the proposed more advanced methods for pedestrian path prediction (*SFlowX/GPDM* or *HoM/Traj*) can achieve more accurate path prediction than basic approaches (linear KF or IMM extension thereof). The associated benefit, in terms of reduction of the combined lateral and longitudinal position error, is 10–50 cm at a time horizon of 0–17 frames (up to 0.77 s) around the stopping event. Fig. 10(d) indicates that a 50-cm improvement in lateral position estimation is reached at several time instants. Tables III and IV also suggest that the vehicle egomotion compensation is done reasonably but not perfectly. Further benefits can be obtained when localizing the pedestrian more accurately and improving upon the vehicle egomotion compensation. Comparing the columns

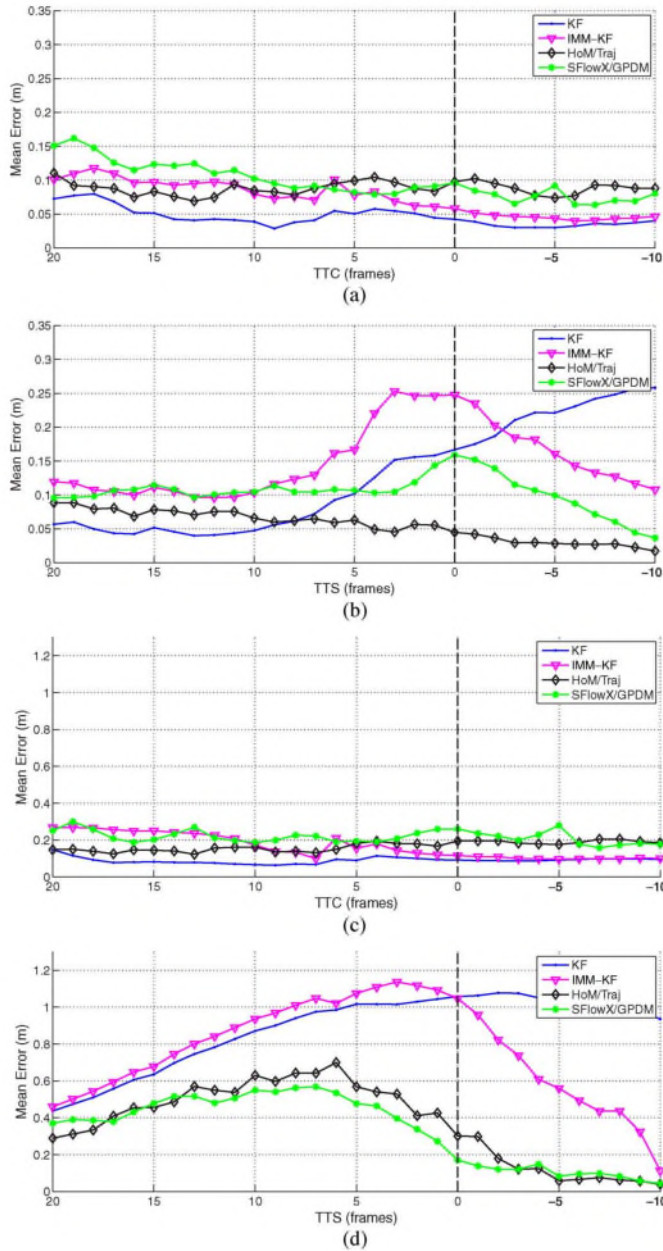


Fig. 10. Mean lateral localization error at each time step for *jittered* data and vehicle standing and moving (walking versus stopping trajectories, prediction horizon 0 versus 17 frames). (a) Pedestrian walking, prediction time 0 frames. (b) Pedestrian stopping, prediction time 0 frames. (c) Pedestrian walking, prediction time 17 frames. (d) Pedestrian stopping, prediction time 17 frames.

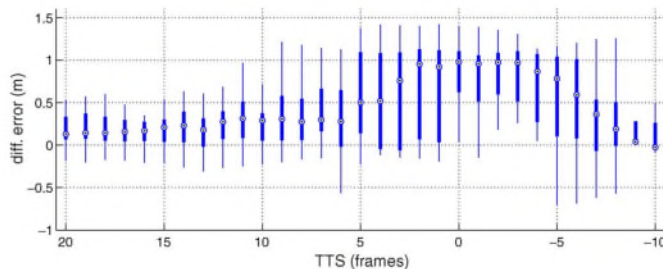


Fig. 11. Distribution of the lateral prediction error difference (IMM-KF-HoM/Traj). Results for the *jittered* data, prediction horizon of 17 frames and stopping trajectories.

TABLE V
MEAN COMBINED LONGITUDINAL AND LATERAL RMSE (IN METERS)
FOR STOPPING AND WALKING TRAJECTORIES USING SYSTEM DETECTIONS
WITH DIFFERENT PREDICTION HORIZONS (FRAMES)

		system detections			
		walking		stopping	
KF	Mean	0.2	0.55	0.27	1.08
	$\pm$ Std	0.05	0.28	0.08	0.29
IMM-KF	Mean	0.21	0.55	0.29	1.04
	$\pm$ Std	0.06	0.3	0.14	0.25
HoM/Traj	Mean	0.14	0.39	0.14	0.63
	$\pm$ Std	0.03	0.12	0.04	0.22
SFlowX/GPDM	Mean	0.15	0.43	0.21	0.52
	$\pm$ Std	0.06	0.27	0.06	0.19

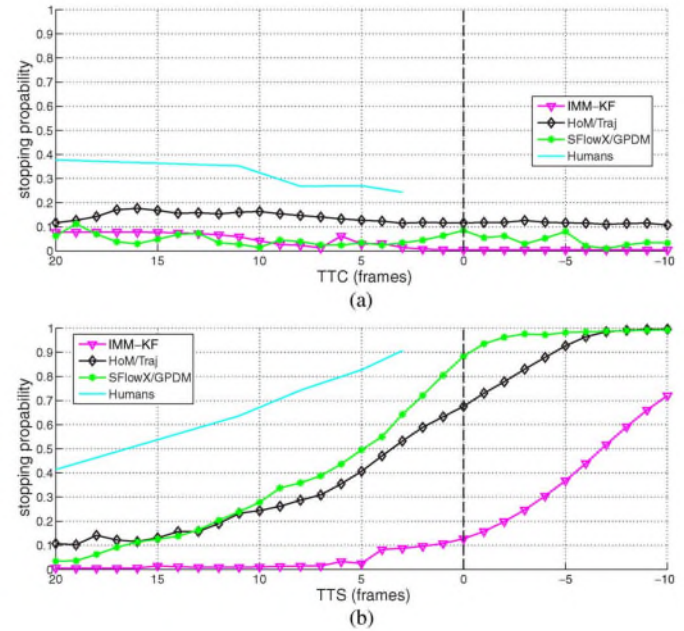


Fig. 12. Estimated probability of stopping over time for (a) walking and (b) stopping test trajectories (averaged over all respective sequences). (a) Pedestrian walking. (b) Pedestrian stopping.

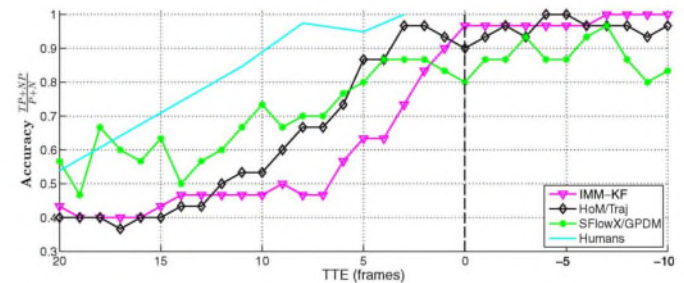


Fig. 13. Classification accuracy of the different systems over time. Results for the *jittered* data.

“vehicle standing, label box, 17” (ideal situation) and “vehicle moving, jittered box, 17” (currently achievable situation) shows that position prediction can be improved by approximately 15–81 cm for the various systems.

These findings are encouraging in terms of the expected benefits that can be achieved, when integrating more sophisticated path planning in pedestrian safety systems that perform emergency vehicle maneuvers (braking, steering).

We now turn to computational cost issues. The popularity of the simple linear KF can be explained due to its relative effectiveness and its low computational requirements. Although the computational cost doubles for the two-process model *IMM-KF*, it remains moderate compared with the *HoM/Traj* and *SFlowX/GPDM* approaches. For the latter, the cost of motion feature extraction needs to be accounted for first. Furthermore, for the *HoM/Traj* approach, a prediction step requires traversing the search tree for each particle and looking up the successor snippet. Computational costs to predict a snippet is linear in the depth of the search tree. To incorporate new measurements, the *QRCLS* distance to each particle has to be computed to update the particle weights. Looking ahead pedestrian position requires applying the mean shift procedure to the predicted particle positions to find the main mode. Main computational costs of the *SFlowX/GPDM* can be subdivided into the costs of predicting a GPDM latent space position and reconstructing the feature to apply the particle weight update. To predict a single particle, the mean prediction on the latent space has to be applied [see (9)]. Because the first part of the formula ($\mathbf{X}_{2:N}^T \mathbf{K}_{\mathbf{X}}^{-1}$) can be precomputed, the online costs for a latent space prediction result from evaluating the kernel function $k_{\mathbf{X}}(\mathbf{x})$ between the particle latent position $\mathbf{x}$ and all inducing variables. Similarly, reconstructing the feature requires evaluation of (10) with a precomputed $\mathbf{Y}^T \mathbf{K}_{\mathbf{Y}}^{-1}$ and evaluation of the kernel function $k_{\mathbf{Y}}(\mathbf{x})$. Costs for an update and prediction of a particle are limited by the number of inducing variables.

Using an unoptimized MATLAB implementation on a 2.53-GHz central processing unit, the path prediction 17 frames into the future requires, on average, 0.003 s for *KF* and 0.017 s for *IMM-KF*. The MATLAB version of the *HoM/Traj* approach with an optimized version of the trajectory matching and mean shift procedure in *C* requires 0.6 s. Without code optimization, the *SFlowX/GPDM* approach requires, on average, 5.4 s for the prediction. Processing times for both *HoM/Traj* and *SFlowX/GPDM* can be much improved by special hardware (i.e., graphics processing unit, digital signal processor, and field-programmable gate array) by parallelizing the particle computation.

In terms of scalability, learning a GPDM quickly becomes unfeasible for larger data sets (for example, ≥ 1000 samples) without an approximation method. The fully independent training conditional (FITC) [37] method reduces the complexity from $O(N^3)$ for the SCG method [31] (cf., Section III-A3) to $O(k^2N)$, when k is the number of data points that remain in the computation of the covariance matrix. Our full data set contains approximately 1700 training samples, and we set $k = 100$. When using the FITC approximation with a fixed number of inducing variables k , the online computational costs do not increase when extending the size of the training set. Without an approximation method, kernel evaluations between all samples in the training set have to be applied. Regarding scalability with the number of pedestrian motion patterns considered, training a single model containing different motion patterns leads to degenerated models on our data set. Degenerated models showed an insufficient latent space prediction performance. Although methods exist to prevent model degeneration [32] when using sequences with a large variety of motion patterns, the

computational complexity during training increases. Extending the *SFlowX/GPDM* system with additional motion patterns requires training separate GPDMs for each motion pattern. In the online case, the computational costs linearly increase in the number of models.

Since the *HoM/Traj* system is an instance-based learning approach using a probabilistic search tree, different motion patterns can be added to the training set without complication. Adding additional snippets to the training set leads to an increase in the depth of the binary search tree. Online costs to predict the state of the particle filter are thus sublinear (logarithmic) in the number of training samples.

VI. CONCLUSION

We have considered four approaches (*SFlowX/GPDM*, *HoM/Traj*, *KF*, and *IMM-KF*) for stereo-vision-based pedestrian path prediction from a vehicle. Two scenarios were considered: in one, the pedestrian walking toward the curbside, lateral to the vehicle driving direction, would stop; whereas in the other, the pedestrian would continue walking.

Experiments indicated similar path prediction performance of the four approaches on walking motion, with near-linear dynamics. During stopping, however, the newly proposed approaches (*SFlowX/GPDM* or *HoM/Traj*), with nonlinear and/or higher order models and augmented motion features, achieved a more accurate (longitudinal and lateral) position prediction of 10–50 cm at a time horizon of 0–0.77 s around the stopping event. During stopping, a 50-cm improvement in lateral position prediction was reached at several time instants. Further benefits are possible when localizing the pedestrian more accurately and improving upon the vehicle egomotion compensation: We obtained improvements in lateral position prediction of 15–81 cm for the various systems.

These are encouraging results, indicating that more advanced pedestrian path prediction approaches can make a real difference, when integrated in the next-generation active pedestrian safety systems that perform emergency vehicle maneuvers (braking, steering). However, more work is necessary on improving pedestrian localization, enlarging the set of pedestrian motion patterns considered and increasing the size of the data set, before these benefits can materialize.

ACKNOWLEDGMENT

The authors would like to thank C. Schnörr (University of Heidelberg) and M. Enzweiler (Daimler Research and Development) for valuable discussions and feedback. They would also like to thank N. Lawrence (University of Sheffield) for making his GPDM software publicly available.

REFERENCES

- [1] M.-M. Meinecke, M. Obojski, D. M. Gavrila, E. Marc, R. Morris, M. Töns, and L. Lettelier, "Strategies in terms of vulnerable road user protection," EU Project SAVE-U, Deliverable D6, 2003.
- [2] S. Schmidt and B. Färber, "Pedestrians at the kerb—Recognizing the action intentions of humans," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 12, no. 4, pp. 300–310, Jul. 2009.

- [3] J. M. Wang, D. J. Fleet, and A. Hertzmann, "Gaussian process dynamical models for human motion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 283–298, Feb. 2008.
- [4] Y. Bar-Shalom, X. Li, and T. Kirubarajan, *Estimation With Applications to Tracking and Navigation*. Hoboken, NJ, USA: Wiley, 2001.
- [5] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: A benchmark," in *Proc. IEEE Conf. CVPR*, 2009, pp. 304–311.
- [6] M. Enzweiler and D. M. Gavrila, "Monocular pedestrian detection: Survey and experiments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2179–2195, Dec. 2009.
- [7] M. Meuter, U. Iurgel, S.-B. Park, and A. Kummert, "The unscented Kalman filter for pedestrian tracking from a moving host," in *Proc. IEEE Intell. Veh. Symp.*, 2008, pp. 37–42.
- [8] J. Tao and R. Klette, "Tracking of 2D or 3D irregular movement by a family of unscented Kalman filters," *J. Inf. Convergence Commun. Eng.*, vol. 10, no. 3, pp. 307–314, Jul. 2012.
- [9] C. Keller, T. Dang, A. Joos, C. Rabe, H. Fritz, and D. M. Gavrila, "Active pedestrian safety by automatic braking and evasive steering," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1292–1304, Dec. 2011.
- [10] Y. Abramson and B. Steux, "Hardware-friendly pedestrian detection and impact prediction," in *Proc. IEEE Intell. Veh. Symp.*, 2004, pp. 590–595.
- [11] G. De Nicolao, A. Ferrara, and L. Giacomini, "A collision risk assessment approach as a basis for the on-board warning generation in cars," in *Proc. IEEE Intell. Veh. Symp.*, 2002, pp. 436–441.
- [12] C. Wakim, S. Capperon, and J. Oksman, "A Markovian model of pedestrian behavior," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 2004, pp. 4028–4033.
- [13] G. Antonini, S. V. Martinez, M. Bierlaire, and J. Thiran, "Behavioral priors for detection and tracking of pedestrians in video sequences," *Int. J. Comput. Vis.*, vol. 69, no. 2, pp. 159–180, Aug. 2006.
- [14] M. J. Black and A. D. Jepson, "A probabilistic framework for matching temporal trajectories," in *Proc. ECCV*, 1998, pp. 909–924.
- [15] H. Sidenbladh, M. J. Black, and L. Sigal, "Implicit probabilistic models of human motion for synthesis and tracking," in *Proc. ECCV*, 2002, vol. 2350, pp. 784–800.
- [16] E. Käfer, C. Hermes, C. Wöhler, H. Ritter, and F. Kummert, "Recognition of situation classes at road intersections," in *Proc. IEEE ICRA*, 2010, pp. 3960–3965.
- [17] C. G. Keller, C. Hermes, and D. M. Gavrila, "Will the pedestrian cross? Probabilistic path prediction based on learned motion features," in *Proc. DAGM Symp. Pattern Recognit.*, 2011, pp. 386–395.
- [18] S. Köhler, M. Goldhammer, S. Bauer, K. Doll, U. Brunsmann, and K. Dietmayer, "Early detection of the pedestrian's intention to cross the street," in *Proc. IEEE Intell. Transp. Syst. Conf.*, 2012, pp. 1759–1764.
- [19] Z. Chen, D. Ngai, and N. Yung, "Pedestrian behavior prediction based on motion patterns for vehicle-to-pedestrian collision avoidance," in *Proc. IEEE ITSC*, 2008, pp. 316–321.
- [20] L. van der Maaten, E. Postma, and H. van den Herik, "Dimensionality reduction: A comparative review," Tilburg Univ., Tilburg, The Netherlands, Tech. Rep. TiCC-TR 2009-005, 2009.
- [21] N. Lawrence, "Gaussian process latent variable models for visualization of high dimensional data," in *Proc. Adv. NIPS*, 2004, vol. 16, pp. 329–336.
- [22] M. Tipping and C. Bishop, "Probabilistic principal component analysis," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 61, no. 3, pp. 611–622, 1999.
- [23] R. Urtasun, D. J. Fleet, and P. Fua, "3D people tracking with Gaussian process dynamical models," in *Proc. IEEE Conf. CVPR*, 2006, pp. 238–245.
- [24] M. Andriluka, S. Roth, and B. Schiele, "People-tracking-by-detection and people-detection-by-tracking," in *Proc. IEEE Conf. CVPR*, 2008, pp. 1–8.
- [25] L. Raskin, M. Rudzsky, and E. Rivlin, "Dimensionality reduction using a Gaussian process annealed particle filter for tracking and classification of articulated body motions," *Comput. Vis. Image Understanding*, vol. 115, no. 4, pp. 503–519, Apr. 2011.
- [26] C. Keller, M. Enzweiler, C. Schnörr, M. Rohrbach, D.-F. Llorca, and D. M. Gavrila, "The benefits of dense stereo for pedestrian detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1096–1106, Dec. 2011.
- [27] H. Hirschmüller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.
- [28] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.
- [29] P. Riekert and T. E. Schunck, "Zur fahrmechanik des gummibereiften kraftfahrzeugs," *Arch. Appl. Mech.*, vol. 11, no. 3, pp. 210–224, Jun. 1940.
- [30] A. Wedel, T. Pock, J. Braun, U. Franke, and D. Cremers, "Duality TV-L1 flow with fundamental matrix prior," in *Proc. Image Vis. Comput. New Zealand*, 2008, pp. 1–6.
- [31] J. M. Wang, D. J. Fleet, and A. Hertzmann, "Gaussian process dynamical models," in *Proc. Adv. NIPS*, 2006, pp. 1441–1448.
- [32] R. Urtasun, D. Fleet, and N. Lawrence, "Modeling human locomotion with topologically constrained latent variable models," in *Proc. Human Motion—Understanding, Modeling, Capture Animation*, 2007, pp. 104–118.
- [33] Y. Boers and J. Driessen, "Interacting multiple model particle filter," *IET Radar, Sonar Navig.*, vol. 150, no. 5, pp. 344–349, Oct. 2003.
- [34] N. Howe, M. Leventon, and W. Freeman, "Bayesian reconstruction of 3D human motion from single-camera video," in *Proc. NIPS*, 2000, pp. 820–826.
- [35] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Norwood, MA, USA: Artech House, 1999.
- [36] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. CVPR*, 2005, pp. 886–893.
- [37] N. Lawrence, "Learning for larger datasets with the Gaussian process latent variable model," in *Proc. Workshop Artif. Intell. Stat.*, 2007, pp. 21–24.



Christoph G. Keller received the M.Sc. degree in computer science from the University of Freiburg, Freiburg, Germany, in 2007. He is currently working toward the Ph.D. degree with the University of Heidelberg, Heidelberg, Germany.

In 2006 and 2007, he was a visiting Student Researcher with Siemens Corporate Research, Princeton, NJ, USA. Since 2012, he has been with Daimler Research and Development, Böblingen-Hulb, Germany. His research interests include video analysis for intelligent vehicles, particularly

pedestrian tracking and vehicle localization.



Dariu M. Gavrila received the Ph.D. degree in computer science from the University of Maryland at College Park, MD, USA, in 1996.

In 1996, he was a Visiting Researcher with the Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA. Since 1997, he has been a Senior Research Scientist with Daimler Research and Development, Ulm, Germany. Since 2003, he has additionally been a part-time Professor with the University of Amsterdam, Amsterdam, The Netherlands, in the area of intelligent perception systems. Over the past 15 years, he has focused on visual systems for detecting human presence and activity, with application to intelligent vehicles, smart surveillance, and social robotics. He led the multiyear pedestrian detection research effort at Daimler, which materialized in the Mercedes-Benz E- and S-Class models (2013).

Prof. Gavrila was the recipient of the I/O Award 2007 from the Dutch Science Foundation (NWO). His personal Web site is www.gavrila.net.