SLRTP 2020

23 August 2020

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

NECATI: Does anyone else hear an echo? Right, we're just about to start. Okay. Well, welcome, everyone to our workshop. SLRTP2020. We're happy that you joined us today and hopefully here you'll find it interesting and have lots of good interactions. So, before we go into any details, -- maybe I should introduce myself. I'm Necati Camgoz, a research fellow at the University Of Surrey. And let me go into the details of the workshop. We have translations, interpretations and the captioning of all of the workshop materials. All video materials have been captioned and translated to BSL and ASL, huge thanks to Dr Robert Adam. The material is already available on the ECCV platform and they will be made publicly available on our website after the conference. We also have ASL and BSL interpretations and live captions during the Q&A session. For ASL, we're working with our interpreting and we have Anna Michaels and Brett Best. For BSL we have Akbar and Esther. For captions we have MyClearText and we have Katy ride and Tara Meyer doing the captioning. There's a stream text link as you can see here and I'll post it in the chat. Next, so, for panelists, during the Q&A session we suggest you use side by side mount. At the moment I am, I also have a computer that is an attendee so you should be seeing as attendees, you should be seeing left-hand side the slides and, right-hand side, Anna, Akhbar and me. That should be the default view. If you are having issues, please let us know. Try side by side mode. And we'll try to help. As attendees you can ask questions using the Q&A functionality. You can also, again, use the chat to raise technical issues. Next to the Q&A you'll see a CC, closed caption button if you want to use closed captions or, again, you can use the stream text link that I just post the on the chat. So we have amazing keynotes today. We have five presenters, Lale Akaran, Matt, Christian Oscar. We will play 20 minutes prerecorded and caption videos for each keynote. So video transfer might be slow, as alternative I will post the ECCV platform links so you can watch it with us, not necessarily over Zoom but over ECCV if it works better for you. This will be followed by a 10-minute Q&A session where you can either use Q&A functionality or raise hands if you are a panelist to ask your questions. We have good interest in our workshop. We received 25 high quality submissions and, out of which 18 were accepted. 10 full papers and 8 extended abstracts. You can find the papers online already. But the videos are the papers will be made publicly available after the conference. We hope everybody watched the accepted paper videos already before the workshop. But you can always use the coffee break to watch few if you missed some. We will take the questions for four to six papers at a time. The chair will go through the questions from the Q&A and read them out loud. The presenter is asked to turn their audio and video to answer the question. If another panelist would like to ask a question then, please, again, use the raised hand functionality and then chair will give you the floor and then you can turn your video and audio on to ask a question. Please when the interaction is done, don't forget to close your audio and video so that we can have an another. Also, please specify which paper and presenter you are asking the question, if you are using the Q&A functionality. Here are the organisers. That's me on top left. We have Gul and Samuel who is going to chair some of the workshop. We have Neil, Richard, Andrew and Kearsy as organisers as well. And we would like to give some special thanks to Ben Saunders, Benice, ... for their help and advice. We would also like to thank our sponsors for their support and making it possible to provide American Sign Language and British Sign Language translations and captioning for the workshop, so, thank you Microsoft and Google. So, here is the schedule. We're starting with an amazing talk from Professor Benice Woll titled Processing Sign Languages: Linguistic, Technological and Cultural Challenges. I'll play the video and I'll just share the link. For that talk on the ECCV platform is you can either watch from the ECCV platform or you can watch the video with us on Zoom. Again, thank you very much for attending. I hope you'll find it enjoyable and valuable. I again shared the stream text link, if you missed it at the beginning.

>>: Thank you for an amazing talk, Benice. Can you open your camera? Perfect. Let me share. So we already have a question from the floor which I'm going to share now. By the way, we're recording the workshop which I forgot to start at the beginning. To make it publicly accessible afterwards. If you want to be removed, just let us know and we will do that accordingly.

NECATI: So we have a question from Javier and they ask: text researchers need, a, data, b, human evolution of the systems we develop. Do those two ingredients only some from sign language speakers, raw data somehow align with spoken languages would be enough nowadays. How do you suggest that we establish a long term and enriching relationship between tech researchers and the deaf community from these two resources?

BENICE: Thanks very much for the comment and the question. Of course, this issue of aligning raw data with spoken language is a non-trivial point, since we're not just talking about single sentences, if we're talking about naturalistic text, whether the a monologue or dialogue, there's also issues of pragmatics and structures that extend over sentence. So I still think there's, we're not quite at the stage where -- where the question is alignment is non-trivial. But I think I would like to get to the second point of your question which is, how do we establish a long term and enriching relationship? First of all, we have to build capacity amongst the deaf community to become tech researchers. Are we doing enough to make deaf researchers, deaf people, become researchers so that they can cot work, not just build relationship as outsiders? The point at which the deaf community, whether you've got deaf researchers involved or not, need to be involved is from the point where one begins to think about the research questions. So there needs to be involvement not at the point where one gets funding and then one pays for somebody to provide some data. But where a project is being developed. So we see that model being used increasingly, for example, in health care research. Where you don't get the money unless you can show how patients helped you develop the research questions. We're not talking here about patients but we're talking about a community which must have a say in what the questions are and how to go about answering them. And we need a very different approach to do that. We also need to look at possibilities for further investment and support other deaf communities in projects that they are keen to promote. So, I think it's a case of really needing to work towards a sort of change that we're looking for in other areas of research. We wouldn't consider doing research on members of an ethnic minority community. We wouldn't be satisfied by just saying, we'll ask them to provide some data. We would expect that the questions would be those which they want to see addressed. My particular argument would be, we need to build capacity, we need more deaf researchers to work in teams in order to really get research going in the right direction. Thanks.

NECATI: Thank you for that answer. We had a raised hand from Lale so I'm leaving the floor for her to ask her question.

LALE: Hello. Thank you very much for that. You mentioned that the absence of a reference gram March is an important challenge. There have been enormous efforts in some areas preparing annotating, tagging, sign language data whereas others have discussed resources. If there are many some commonalities between languages, that would help enormously, the resource pool languages. So what do you think is common between the gram Marchs of different languages? For example, can we use some -- is there -- end of sentences, signed sentences? Is this a common thing? Or end of signs, can we use anything that's common in segmentation, for example?

BENICE: Well there are, of course, commonalities between different sign languages. But, of course, there are also differences. As soon as one begins to do detail linguistic analysis, even on sign languages that are closely related one finds differences just as one finds looking at closely related spoken languages. The best source of data that will actually rereflect sign language usage is corp us data. And, in some countries, of course, there are now a corpora of sign language in other countries, there are no such corpora. There's not a very good history of success when using sentence tasks, sentence translation tasks and so on. One of the issues that is of course in most countries, signers are not mono ling wall, they have a greater or lesser degree of access to a written language, if not a spoken language. And there's a lot of evidence that Setential level and often show interference with spoken language. So the best thing at the moment is to collect corpora and get linguistics to work on them. But, also, of course, to look at tools that will speed up that work. So there are automated tools being developed which have real potential to make the creation of that sort of reference grammar faster by, for example, a tool that would segment signed sentences, for example. Or would be able to segment signs would be very useful. But such tools don't yet exist. But they would certainly be a great way to go.

NECATI: Thank you for the answer, Benice. We have some questions from the Q&A. Let me ask one of them. What is the best way to approach deaf researchers or members of the deaf community from a community point of view on the written conversation? Or using interpreters?

BENICE: The answer is autos going to be ask people what they want! I mean, one cannot make assumptions about what is best or that what is best for one person is good for another person. Or that there's only one answer. So the thing to do is to say how would you like me to organise an interpreter? Would you like to correspond in written language? Would you like to lipread? All of these are possible answers for any individual. There's no one answer for everyone and you need to ask.

NECATI: And another question is: Benice, what are the main challenges for enriching the deaf community to become tech researchers themselves? Benice Wednesday I think we need -- there are lots of ways to go about that, as well. The challenges are that young deaf people may not have thought about careers in tech. So there may be a need to go into schools or to find ways of researching young deaf people and to interest them. We have had a lot of success with offering internships and study attachments so that deaf young people can come and sample what it's like doing research in the field. But we're living in a society in which there's still unequal access. And unequal opportunities, and so we can only really fix this by looking more generally at how we can ensure that we actually level the playing field and encourage people to see that this is something that will be a rewarding career.

NECATI: Thank you. One last question, from me: so, from computer region researcher, that works on sign language research, what would be the tool, the one tool, that you would want us to build for easing your life as a linguist?

BENICE: Oh dear! It's bad to be asked for a single tool!

NECATI: Or just the first tool, let's begin with that one.

BENICE: I think myself that something that would help processing at the clausal level would be the best or purpose tool. Because that would be a great stepping stone that would let you look at sign language structure. If you look just at individual lexical signs you sort of lose the structure. Which you ultimately need for coming up with some sensible output. So I would like to see tools for automated processing of clausal information.

NECATI: Do you mean ought mic segmentation, a sign language video, would you like, this is one this is another clause or trying to understand what these clauses mean individually Benice been I think the first thing would be segmentation, taking account not only of manual information but non-manual information. Like, face, head and body. So I think segmentation at clausal level would be really helpful.

NECATI: Thank you very much for your answers. We also have some questions that you can type the answer to. Both I need to move onto Oscar's presentation. By the way, one of the things we didn't realise was the links that I shared with Benice's video, it works on one of my machines; it doesn't work on Mac Os so I would suggest you go to the workshop website if you want to watch the video instead of watching the zoo link. So I'll just move onto Oscar's video, now.

NECATI: Thank you, Oscar, for a great talk. There you go. I'll just share this presentation. So again we have another question from Xavier asking.Ment would you see any benefit on working with sign and spoken language data sets?

OSCAR: All right, thanks Xavier for the question. Can you guys hear me?

NECATI: Perfect.

OSCAR: Okay, cool. So I guess, what you are asking is like would it be beneficial if data sets like Pheonix would as an additional modality and not just a written down, the written down form of it? In short, I don't really think that this is necessary because there's so much complexity in getting the recognition bits right and getting the translation bits right. We also struggle to find data for that. So I think adding more complexity on top, being like a task like spoken language production which is way more mature than any of the other techniques. I think at least for the next few years you would always be better off to just synthesise the written output. So, me, personally, I don't see much of that.

NECATI: Xavier just updated his questions. He also says: or, building sign languages data sets on top of existing spoken language data sets. So would you think it's a good idea to build such data sets?

OSCAR: I think it is crucial to capture nominative real life sign language and I think as Benice also pointed out, this is corp us work, where you need to spend a lot of time and effort into how to get elicitation right such that the data as such is not influenced by the spoke language. And I think this would be really dangerous in such a scenario. So I don't think this is helpful.

NECATI: Okay. Thank you for that. We have another question from Q&A functionality. Ben asks, Oscar, do you think weather is a good domain for future data sets? If not, what could be another good constraint domain for data?

OSCAR: So I personally don't like the weather domain too much! It involves a lot of iconic signs. A lot of pointing just close to the screen. And a lot of difficulty to annotate consistently. When we decided to go for the weather forecast we, I think didn't reach that far and, so, that's why we constrained it to that domain. I think it always depends where you can get a lot of data from. And how you get -- or what are your main concerns for collecting that data. If nativeness is a major concern then you shouldn't go for interpreters data at all and that would certainly be the ultimate goal. We shouldn't want to try to get systems learned on interpreters data but on native, deaf sign language. If we say this is out of scope currently then we could tap into larger resources of interpreted footage. But that is not weather forecast so I would be thinking in other interpreted footages. It can be broadcast news. It could be other YouTube clips, things like that.

NECATI: Perfect. I actually had a question for you but I'll write in the Q&A functionality which you can answer afterwards because we ran out of time to have a coffee break. But thanks a lot for the talk. And now we go into a coffee break for 10 minutes. If you haven't watched accepted papers videos, this is a good chance. And we'll get back in 10 minutes. See you.

(Break).

Welcome back, everyone. From this point on, ... will be chairing. I'll just close my video.

GUL: Okay so knower come back to the next session of the workshop. We can start with our next keynote speaker, Professor Christian, the director of technology access programme.

GUL: Thank you very much for the excellent presentation and for raising these important questions, here, Christian. We can now start the Q&A for 10 minutes. And we already have a lot of questions. We can start by Kearsey, if you want to turn on your camera and ask?

>>: Thank you so much that was a wonderful presentation. At the end of your presentation you made a list of examples of possible community led applications. And language technologies. I agree. We must be asking the deaf community themselves to have this discussion. And discuss with them as to what it is they want. But I'm just curious. Based on your experiences, so far, and your long history with this, what is your personal opinion? Which do you feel is the most important or maybe the most exciting? Is it assigning Alexa or searching to sign language videos. What do you think is the most important application?

>>: Thank you, I'm happy to answer that question. I think there are two different perspectives you can take when answering that question. First, you need to remember in my field of technology as a researcher, we're trying to figure out how to make technology work for deaf and hard of hearing people. And, if you take that perspective, I'm really quite scared of voice assistance such as Alexa. I'm very concerned about that because it's becoming very popular, very widespread, highly utilised, especially here in America and we have over a hundred million home based Alexa systems in use, here in the States. It's very widespread, it's prolific. So what happens in the future if these voice assistances are literally everywhere? We're going to have less access for other types of technology, tablets, typing, any kind of typed devices for communication, et cetera. So deaf people who sign are just basically going to be left behind because they won't have access to that. And I believe that time is coming very soon. Those types of technologies are not accessible at the moment and we need to solve that issue as soon as possible, it needs to become a priority. The other perspective that you can take, when answering this question, is from a language preservation perspective. A sign language preservation perspective. Because that is being utilised less and less for deaf people. There's more and more deaf kids who are maybe focusing on different technical tools to help them hear such as hearing aids, cochlear implants, et cetera. How do we make sure we keep the sign language this we have? And really preserve what is there. I feel it's very important to support the search by video feature for sign languages so we can identify what we do have archived and really keep that. It's really important that we do that before we lose those films that we have and before sign languages, the use of them just diminishes in general.

GUL: Thanks, Kearsey, from the chat. And the next question is from Xavier. The question is, why sign language interpreters are not appropriate for data collection.

>>: It's such a great question. The reason why is because they have a lot of contact between the English language which is -- or any spoken language -- I'm coming from the American context, I'm referencing Spoken English and the sign language. So if an interpreter grows up and they are a hearing language user and they have access to that spoken language they will probably very closely adhere to the structure of that Spoken English. So they are taking in that Spoken English and then producing an interpretation into American sign language which, however still closely follows the English structure of the spoken language. And that's one of the reasons why it's very important to have examples that are used with the native sign language. So sometimes you could have two interpreters working together. One could be a hearing interpreter and they're taking in the spoken message and they're producing an interpretation. And then you have a deaf interpreter who is watching that interpretation and essentially cleaning it up to make it more natural. More natural sign language. And, if you see that, when you see that happen, it looks completely different. This is no offence to hearing interpreters. Hearing interpreters and general -- I don't want to generalise about them -- but this is just basically what we're saying is often you can see a difference between how a hearing interpreter producing something and a native deaf signer would produce something.

GUL: Thank you. I'll move to the next question from Lorna: do you see a lot of possibility in the idea of combining motion capture signing content with AI-powered algorithms to combine and recombine the content? Is this a direction you see that's worth pursuing? What areas do you see this line of work?

>>: That's an interesting question, a difficult question, too. I don't feel completely confident answering that question because I feel like maybe Oscar might have some interesting viewpoints on that. I think if you are using motion capture and you are comparing that to, let's say, video, then the basic features are going to look different. So the big challenge then becomes how you use that for machine learning. How do you combine those two? I'm not really sure it would work. But I'm not an expert in that field any more. So I would really like it hear from Oscar on that one. I see that Lorna says that question was actually meant for Matt!

GUL: Okay, actually, we have more questions. But we have to move to the next keynote. But you can type them on the discussion. We thank Christian again.

CHRISTIAN: Thank you.

GUL: Our next keynote speaker is Matt, he is the director of the school of the information at the Rochester Institute of Technology.

>>: We thank you Matt for the great talk and the great insights. Matt, if you can turn on my camera? So we have a question from Xavier: given the lack of data sets available do you consider that the ASL e-learning tool may be useful for collecting a noisy data set?

MATTHEW: I think there is a lot of the research. There are a lot of resemblers out there looking at methods that can be used for doing distributed collection of sign language that has been contributed by users. So I know that Daniel Bragg at Microsoft has been looking into these projects. In our own project we're not looking at trying to collect data from students. I think there's a really important distinction to be made. So with American Sign Language corporate, as discussed earlier, in the case of using data from interpreters, I think there's also a great risk in using data from student signers. Unless the people who are using the data later appreciate that was the provenance of the data. So, if you have data that are from folks non-native signers and you are using that to train models for applications intended for use in producing fluents signing there's great risk. I think there's still value in collecting some data sets from student signer, especially if your interests specifically is in making robust recognition models that might be useful in certain educational applications. Then you really do need to train to that particular set. But I think that there's going to be a strong responsibility for anyone who is collecting and releasing that type of data set to make sure it's labeled carefully so that folks who make use of it later don't just try to assemble a bunch of data sets to make as big a pot as they can and they mix it up with carefully curated and produced data from native signers. So I think there's a trade off, there.

GUL: Thank you. I'll just repeat the question from Lorna previously: do you see a lot of possibility in the idea of combining motion capture signing content with AI-powered algorithming to combine and recombine the content. Is this a direction you see as worth pursuing? What barriers do you see to this line of work?

MATTHEW: So I guess I'm not sure that I totally understand the question. So what I'm going to take from the question is this: if it's a motion capture data and what you are interested in is using components of that recording, small slices or small pieces of that, in a way to produce and synthesise a fresh animation performance of ASL, I think that is a line of research that has merit and I think there have been researchers that have investigated that type of thing. At our own laboratory we tend not to do that. At our own lab when we're synthesising ASL animations we tend to use more of a key frame animation approach. Where we've carefully produced signs that are in a lexicon collection where human animated the key points of the movement. It gives us the ability to stretch and adjust those a little bit more. However, we do use motion capture data but, for very specifically trained models that tend not to be lexical. Because we don't have quite enough data. So this is why we've investigated things like where do people set up things in space. It's also why we've investigated things like what is this speed and timing and pausing and signing? Because many of those things operate at a level where we think we have enough examples to possibly train those. So the motion capture data is used to train a type of model that is then incorporated into an animation pipeline to make specific choices. But we're not specifically using motion capture to actually produce what you would see as the signing itself. There's trade offs. You could get really beautiful animations if you drive them with motion capture. But you can also have unusual visual artefacts when you combine and manipulate that later. And you can have issues of jitter in the data, shakeyness in the data. Things like this, too. So that's just the particular approach our lab has used.

GUL: Thank you. We have a question from the Q&A tool. Anastasia says: it seems each country uses it's own corpus for sign language. Is there any commonality on how to collect the data. What bias needs for avoided. What methods need to be followed, et cetera.

MATTHEW: There have been some method logical research published for best collection of corpus collection. This follows on from the best practices in field of linguistics when you're trying to collect a high quality data set. In particular some of the practices that folks have published in this space including my lab but many others too, are things like, if you are trying to really get fluent made of signing, it had better only be fluent native signers in the room. And you had better have an environment that locally is really ASL-centric. So when we've collected co-operate in the past that are intended to be for native signers, the person who is the hots, the person operating the camera are all native signers themselves, too, and then engaging in ASL chat before the person sits down to do some recording. So it's really an ASL zone. In my slide show there may have been a picture of me holding a camera. That was posed. I'm never in the room when we're filming because we don't want that English influence. I don't want people doing Englishy signing around because I happen to be in the room. There's also been some method logical research on the best ways to elicit signing. So once again, if you are trying to avoid that sort of English influence and bias, giving someone an English paragraph and asking them to tell this story in ASL is a really bad idea because you've now introduced that English influence. Right? So there's a variety of prompting strategies that you'll see used, sometimes with images, sometimes with someone just sharing personal anecdotes from their life. It depends if you're looking for monologue or some sort of conversational setting.

GUL: Thank you. I think Richard has his hand raised.

RICHARD: In the protocol I'm supposed to switch my camera. Hi Matt. Just a quick question. Historically we've, as lab, we've shied away from production but time to change. We find ourselves now in the situation where we need to do user evaluation. So we come up against the same problem all the time. In production, it's very difficult if I'm asking somebody's opinion to get them to separate the appearance from the AI. And actually what I want them to do is to assess the AI, the quality of the underlying technology. But obviously if you ask a deaf user and the avatar or whatever visual representation you have is substandard, there is an uncanny nightmare they fixate on that. Similarly we did some work with some very similar to what you were talking about where we had an assessment system where it would watch somebody signing, compare that against an envelope of allowable motion and give feedback. When we do the evaluation, people fixate that the colours are wrong, rather than actually -- so maybe you can't answer this question now. But I wonder whether there's certainly some of your papers I should be looking at about how we can do this better.

MATTHEW: So what you are taking about is so the idea that when you are conducting a user study and what you are really trying to evaluate is the quality of sign language that has been synthesised so whether it's an animation or a composite video. We've bounced around and struggled with a lot of this. I am a bit of a methodological geek so I like doing methodological studies to see if we can pick it apart. And it's really hard. We find that the answers are so highly correlated. You can ask about, is this grammatically correct, does it seem understandable, does it look natural? And you think you are asking different things but actually people are giving you this numerical judgment of a gut instinct about whether they think it looks good. If you ask people to share open ended responses about what they might want to comment upon, what looks good and bad in animation, if there's anything encourage whatsoever about the facial expression regardless of what you are trying to ask them about something else, folks will fixate on the part that's flawed. What we've played with in our lab to try to get at this a little bit is other method logical designs where you essentially use maybe side by side presentation of A and B versions of animation and you ask people to basically give you a quality judgment for each one. It means you really cannot do comprehension question testing in that set up because somebody's going to watch the same thing twice, they would have got the answer from the other, they can't answer it so you are stuck with subjective questions. But by letting someone examine side by side sometimes you can get a little bit more of a subtle differences in the two animation. We've also played around with should we tell people what it is we've changed or not? And a lot of times through pilot testing we found people get frustrated if you don't tell them what is slightly different in the two subtly different versions. So in some recent studies we've done with speed and timing and pausing, it's really subtle. If it's a few mili-seconds of time that's different. So we have had to tell people, in one of these, the amount of transition time between two words is little different. Or in one of these, the acceleration curve of the hands during the beginning of signs is a little different. And then we ask them to share their comments. So, why yes, there's no easy answer. There's a variety of methodologies that can be used. If you are just thinking of a single presentation after stimulus and there may be alternatives with simultaneous that could let you get out some of that more subtle stuff.

RICHARD: Thanks.

GUL: Thank you, Matt, for the great discussion. Now we can move to the next session for the first six papers. We will have only question and answer session for them. Unfortunately there was no time to play their presentations. So, if you have any questions for Hannah, Mark, Emily, emit, Katerina or Jessica, can the authors please raise your hands so we make your panelists and when you answer questions please turn on your camera and audio and turn it off afterwards. Thanks. This is only for the first six papers. Do we have everybody? Let's begin with questions already answered, sorry, questions already submitted beforehand. For the first paper, for Hannah, there's a question. The original STGCN model you used to encode the skeleton doesn't seem to include hands and face landmarks in the original paper but you seem to be using them. Could you explain how you obtained these landmarks?

>>: The original model in order to include the hand and the facial key points. And the adaptation follows the same centres so the input for this model, body key points, and the graph structure behind these body key points, and I included the key points and the graph structure behind the key points so the facial key points and the hand key points to the model. And then I trained the model from scratch.

GUL: Thank you. Maybe we can stop the screen share. For the second paper, there's a question for Mark: do you have any plans to integrate others of units such as mouthings and facial expressions to cover non-manual speeches of the sign?

>>: Yes we plan to include other features: we have some information from the body that is obtained from the factorization. The body movements which, for now we're discarding which we're thinking of including it as another modality so that's one avenue of future work.

GUL: Thank you. Oscar has a question.

OSCAR: I have a question for Mark. It's good that you are there. Thanks for the interesting work. I was wondering how do you deal with missing frames? Let's say, like, I think you mentioned that you used track open post on the hands as well. So sometimes these are not recognised and then you have missing frames. How do you deal with them? That's the first question. Attendee second question is: did you happen to run baseline experiments on RWTH Phoenix where you have without pre-training on your subunits data set, just having the full, your full pipeline but just trained in the final training set up.

MARK: Thanks. For the first question, the open post features, we're collecting them into a window of 0.5 seconds as a mate ridiculous to pass to factorize then using matrix completion algorithm so any key points which are missed will be filled in with this matrix completion algorithm. And for the second question, can you remind me please?

OSCAR: I was wondering, if you had run any baseline experiments where you don't pre-train your system on the subunits?do you know what I mean?

MARK: Yes. Within the full system without any subunits just using signer cognition. There, the results were in fear or to the results we obtained when we do the subunits recognition. So we got inferior result, it was around 10% less. So the subunits are somehow with the features it's as if you're just helping the recognition process, the DRN.

OSCAR: Thanks.

GUL: Thank you, Mark. There's a question on the chat from Hannah to Jessica: could you please elaborate on the applications talk about on robots and sign language for military purposes in your article?

JESS: Sure. Hello. So, the reason that this comes up is because the funding I have recently been awarded is military funding. The Australian army is interested in having a gesture based and multimodal interfaces for controlling robotics in the field. So the idea here is that some of the things that Christian was talking about, the idea of like how to we make your Alexa as your Siris and voice interfaces accessible to sign language communities? That that's the space that I see it in. If we've got sign language interfaces for computers. For smart assistances, for robots. My plan is, if we can start from a native sign language first perspective then this is going to result in better technologies for the deaf communities. But, also, better technologies in general for if we have other forms of gestural interfaces. Does that answer the question?

GUL: She says thanks in the chat. Oscar has his hand raised.

OSCAR: Yes, thanks. I have another question, this time for paper 4, from Amit and others. Hi. Thanks for the interesting work. I was wondering, did you happen to compare your methods on this data set that Mark released, I think they also had a very similar task. And the second question: are you considering to making the exact partitioning, I think you mention in the paper how you split the German DJS data set but are you considering making the exact splits publicly available such that other people could reproduce and try out compare against you?

>>: AMIT: Thanks for the question. We did not compare our work on Mark B's data set. The simple reason here is that the DGS data set comes pre-processed with open pose estimation and it was easy to rapidly work on that data set and not run post estimation on the entire large amount of YouTube videos that Mark collected. Regarding the second question, which I'm trying recollect what it was,.

OSCAR: Basically if you are planning on sharing the exact partitioning.

>>: AMIT: We will share everything that we can share publicly about what we're doing with the public DGS corpus, as long as the license allows it. There you have an extensive license that we need to make sure we can share more information of how we did stuff with it. But that's our intention.

OSCAR: Cool, thank you.

>>: AMIT: Thank for the question.

GUL: Thank you. Xavier has a question for Katerina. Did you use the 2D mask skeleton from open pose? Did you consider predicting 3D body skeleton as well.

>>: KATERINA: I'm sorry I'm having some problems with my video, so I'll just connect with audio. I hope this is not problem. Yes, we tried to evaluate our system using the -- we tried to use the 3D human pose key points but it couldn't be easily obtained because there are no available data sets for the training of something like this. We couldn't find something appropriate in order to train our model on upper body videos. For 3D key points estimation. Human pose key points of estimation. And the second question, please?

GUL: The second question was: did you consider predicting 3D body skeleton as well? Did you use the 2d model skeletons from open pose?

>>: KATERINA: Yes, for the mouth region appearance based representation. Not for 3D key point estimation. Mouth key point information.

GUL: Thank you. So another question for Emily, the third paper. The distances between facial landmarks are computed in 2D. Can you comment on the applicability to in the wild videos where the face doesn't necessarily need to be frontal? IS EMILY HERE?: I guess she is not in the participants now. So I will move to another question. Another question from Xavier for Amit: did you run the ablation study of feeding the body land mark coordinates into the RN instead of the optical flow you propose?

>>: AMIT: No, we did not. We used the optical flow in order to mostly track the movement rather than try to detect signs. So we did not see any reason why, using the estimation, the raw pose estimation would work better in such a shallow network. It is a cool idea that we could try. So thanks for the suggestion.

GUL: Thank you. I think Emily is now here. Should I repeat the question?

>>: EMILY: Hi, yes please. I'm sorry. I had a problem.

GUL: The question is: the distances between facial landmarks are computed in 2D. Can you comment on the applicability to in the wild videos where the face doesn't necessarily need to be frontal?

>>: EMILY: Yes. We kind of tried to do the combination automatic. And when the face is over a pose that it's not recognised, then we try to align the face so we can get the landmarks.

GUL: Thank you. Thank you to all the authors. This concludes the Q&A session for the first six papers. Next we would like to announce the best paper of the workshop. We had a large number of papers with great contributions. The organisers singled out one paper based on its significance in sign language research. Could you please share the screen, Necati?

NECATI: I'm on it.

GUL: Okay, the best paper is awarded to the paper entitled Automatic Segmentation of Sign Language Into Subtitled Units: congratulations. We now have our next coffee break for 10 minutes. We'll be back 50 past with the keynote speaker Lale.

(Break).