

A Framework for Independent Hand Tracking in Unconstrained Environments

Imran Achmed¹, Isabella M. Venter¹ and Peter Eisert²

Department of Computer Science

University of the Western Cape¹, Private Bag X17, Bellville 7535, South Africa

Tel: +27 21 959 3010, Fax: +27 21 959 3006

and Image Processing Department

Fraunhofer Heinrich Hertz Institute², 10587 Berlin, Germany

Tel: +49 30 31002 614, Fax: +49 30 31002 190

email: {2507311, iventer}@uwc.ac.za¹ and peter.eisert@hhi.fraunhofer.de²

Abstract—Communication is an essential life skill used to interact with and share information between people. The Deaf, who are not able to participate in normal communication with the hearing community, are therefore marginalised. To address this communication gap, this paper presents a novel learning-based framework that provides sign-language recognition in unconstrained environments by tracking the right and left hands independently. A segmentation algorithm that assists the tracking process of multiple moving objects that share similar characteristics, such as the hands, is suggested and referred to as region reference points. The proposed framework combines region reference points with skin detection, motion detection and the selection of skin-identified clusters. To address the problem of ambiguity caused by occlusion, the hands are learnt while tracking using a support vector machine and predicted when separated from the occlusion state. The framework is evaluated on ten isolated South African Sign Language gestures performed by six individuals. The experimental results show that the framework achieved an average rate of 82% on a hand tracking system.

Index terms: connected components labelling, region reference points, sign language recognition, support vector machine

I. INTRODUCTION

Over the past decade the technological advancement of mobile communications has seen exponential growth, enabling millions of people to interact socially and providing socio-economic opportunities [3]. The benefits arising from this rich form of social communication and information exchange are unfortunately not available for the hearing impaired or Deaf¹, who primarily use sign language [2]. Skilled interpreters are often used to alleviate the communication problem; however, there are not enough of these interpreters to assist the many Deaf individuals in South Africa that require such services. Furthermore, their services are expensive. An automated translation system that would bridge the communication between the two communities would be a valuable tool to the Deaf.

Such an automated translation system encompasses a multidisciplinary research area that involves image

processing, pattern recognition, linguistics and natural language processing. Sign language recognition, a component of the system, is a challenging problem mainly because of the complexities involved in the visual analysis of signed gestures. The visual analysis of sign language includes facial expression recognition, hand shape recognition and hand position detection. Researchers who focused on South African Sign Language (SASL) mostly work on hand tracking and identification without explicitly distinguishing between the right and left hands. Distinguishing between the right and left hands gives rise to three additional challenges: (1) dealing with occlusion factors; (2) continuing to distinguish between the right and left hands after occlusion has occurred; and (3) recovering from the tracking failure of either hand.

In this paper, a novel learning-based approach to independently track the right and left hands is presented. The vision based method is implemented by applying contours to the skin, identified as areas that are likely to be the hands or face, and handling occlusion by distinguishing between the two hands. These contours are clustered together using connected components analysis so that noisy areas are reduced. A novel segmentation algorithm, referred to as region reference points (RRP), divides an image into a number of regions where each region holds information regarding the areas that have motion, skin colour and cluster centres. This information is then used to identify hands in a particular region that is used in the subsequent frame as a reference point, thereby limiting the search area and further reducing noise. The algorithm identifies the features of the hand that are “good” features to learn, based on a previous reference point. Furthermore, the reference point can be used to identify when the tracking of a hand has failed. While the hands are tracked, the “good” features are learnt using a support vector machine (SVM). When occlusion occurs, the learning process is stopped and the trained model is used to predict and re-identify the hands as either the right or left hand. The approach was evaluated on ten isolated SASL gestures performed by six individuals. The results indicate that the hand tracking system achieved an overall accuracy of 82% on average.

This paper is organised as follows: in section II, related work is discussed; section III discusses the framework that combines the selection of contours with RRP; the experimental analysis and results are presented in section IV; and the paper is concluded in section V, where future work is also identified.

¹ ‘Deaf’ in this context refers to people that use South African sign language as their primary language.

II. RELATED WORK

Hand tracking is defined as a process that aims to estimate continuous hand motions in image sequences [5]. Research on hand tracking algorithms varies from those that use auxiliary means and those that are purely passive.

In algorithms that use auxiliary means, devices such as data suits, gloves or position markers are often used to find direct measurements of the joint angles and spatial positions of the hands [11]. These systems are generally faster in terms of real-time processing and provide more accurate information. This is, however, an impractical solution in everyday tasks, as using these devices is inconvenient and often requires some form of calibration. Wang and Popović [14] tracked a hand covered with a glove that was imprinted with a custom pattern. They designed the pattern in order to simplify the estimation of hand postures that would allow a nearest-neighbour approach to be used in tracking the hands. In addition, they used a Hamming distance-based acceleration data structure to obtain interactive speeds and inverse kinematics for accuracy. The results were visually presented and showed a positive result; however, monocular depth ambiguities remain a problem when using this method.

A more practical solution is to use purely passive algorithms that allow the spatial positions of the hands to be found in non-invasive ways by using various image-processing algorithms. In Liu and Zhang [12], the hand-tracking method is based on a particle filter framework combined with local binary patterns and colour cues. An evaluation of this framework shows that the combination of local binary patterns and colour cues contributes to more robust tracking of the hands than with either cue alone. An approach that combines a modified version of the maximally stable extremal region tracker that efficiently calculates colour likelihood maps to detect and track the hands was designed by Donoser and Bischof [8]. By using this combination, detection of the hands is performed efficiently by only searching areas with high likelihood values; however, this approach does not demonstrate any occlusion handling. Asaari and Suandi [4] used a combination of a predictive framework and an appearance model to estimate the state of a hand based on its location in the previous frame. In their predictive framework, adaptive Kalman filters were employed and a combined form of skin colour and motion cues were used to observe the hand state in successive frames. Their appearance model was represented by means of eigenspaces. It provides a compact description of the hands and was able to learn the eigenbases simultaneously while updating the model to account for any appearance changes over time. Their results show an average hand detection rate of 97%. They, however, only evaluated their system on single hand tracking.

Although non-vision based approaches have the advantage of real-time performance, they are expensive, cumbersome in a sign language domain and cannot be used in unconstrained environments. On the other hand, vision based approaches are inexpensive and have the capabilities of achieving near to real-time performance. Therefore, in this paper, a vision-based approach will be followed, as it would be more suitable for hand-tracking in unconstrained environments.

III. INDEPENDENT HAND-TRACKING FRAMEWORK

In the following sub-sections, the proposed framework will be discussed. The discussion will deal with the selection of clusters, the proposed segmentation algorithm, RRP and the handling of occlusion.

A. Cluster Selection

Researchers developing approaches for hand-tracking applications often use a single global feature characteristic such as either colour or shape. In the present research, multiple features (colour, motion and shape) are exploited in order to extract the maximum amount of information that can be collectively used to isolate and track a hand in unconstrained environments. When tracking the limbs of an individual, the most dominant feature would be colour, as it can easily be distinguished from other objects in the background. Furthermore, this feature is robust against scaling, rotations and partial occlusions. The process of identifying these colour features is referred to as skin detection. Identifying skin-coloured pixels, however, is not easy because the appearance of these features varies due to factors such as changes in light, viewing geometry and the characteristics of the camera. Furthermore, scientific studies have shown that skin-colour diversity in South Africa is one of the highest in the world [13]. These studies suggest that skin-colour features would differ from person to person and an adaptive method to retrieve these features among all races and skin tones is therefore required.

Although an optimal colour space for skin detection does not exist, many researchers agree that the hue-saturation-value (HSV) colour space has a restricted range on the human skin colour [6]. They also agree that the hue component of the HSV colour space can be used to effectively isolate skin colour, while discarding the value component, which is directly related to the colour luminance information. In contrast to other researchers [6], who only use the hue component (see Figure 1(a)), the present research uses both the hue and saturation components, where the saturation component provides the degree of the dominant colour of an area in proportion to its brightness (see Figure 1(b)).



Figure 1(a): Back-projected image using hue only (b): Back-projected image using hue and saturation

In contrast to other researchers, who employed a trained model to identify skin-coloured pixels, a more efficient way is proposed that directly identifies the colour distribution of a person's skin colour and adaptively changes the colour distribution according to changes in the scene, such as light. Trained models often rely on the skin-colour range on which it was trained and need to be retrained if small changes occur or simply fail if large changes, such as a burst of light, occur. The proposed method [1] effectively determines the colour distribution instantly by using the area around the nose with a radius of ten pixels to determine the colour

distribution in every frame. The colour distribution in this radius ensures that the optimal skin colour distribution can be extracted without being negatively affected by facial hair, lips or eyes. In the scene, the regions of interest, identified by the skin detection method, are the hands and face. In order to determine these areas, connected-components labelling [7] is used to extract the contour surrounding these areas.

Connected-components labelling is an algorithm that groups a set of pixels into components based on the level of its pixel connectivity. When all the groups have been determined, every pixel is labelled according to the component it was assigned. It is a sequential two-pass algorithm that iterates through the two-dimensional (2D) binary image using either 4-connectivity or 8-connectivity labelling [7].

In the present research, 8-connectivity labelling is used because it searches for connected regions in all directions. In the first pass, moving from the top left of the image to the bottom right; a temporary label is assigned to each skin-coloured pixel based on the values of the neighbouring pixels that have already been processed. When none of the top-left four neighbouring pixels (pixels that have already been passed) is a skin-coloured pixel, then a new label is assigned to the current pixel; however, when only one neighbouring pixel is considered to be a skin-coloured pixel, then its label is assigned to the current pixel. When a skin-coloured pixel is found containing two or more skin-coloured neighbouring pixels that have different labels, then the labels associated with these neighbouring pixels are stored as being equivalent. These equivalences are then used to determine equivalence classes after the first pass, where a unique label is assigned to each class. In the second pass, the label of its corresponding equivalence class replaces each temporary label.

The connected components are then used to extract the contour clusters that define the skin-coloured areas. This is followed by the analysis of each cluster, where clusters larger than the face and clusters smaller than the fist (half the size of the face) are discarded. This process of eliminating unwanted clusters reduces the amount of noise in a frame, as seen in Figure 2.

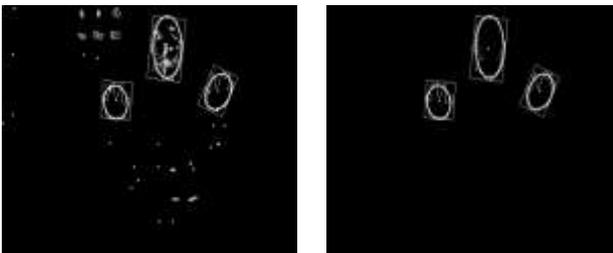


Figure 2(a): Extracted contour clusters from the skin-detected image

(b): Filtered contour clusters of the skin-detected image

B. Region reference points

Tracking both the right and left hands of an individual is a complex task, because hands share many similarities that cannot easily be distinguished. In addition, the colour similarity between the hands and face is identical, which further complicates the task. Therefore, when tracking a single hand, it is possible that tracking may fail when the tracked hand comes near the face or the opposite hand.

In this paper, the segmentation algorithm RRP is

proposed. This algorithm assists in the tracking process of multiple moving objects that share similar characteristics, such as the hands, while simultaneously reducing the search area to regions that are most likely to contain the tracked object.

RRP operates by dividing a 2D frame into a group of regions and by creating a 10x10 grid that results in 100 regions that are defined and that are used as reference points in the tracking process (see Figure 3). The number of regions was selected based on the assumption that an individual that signs would be at a common distance from the camera such that only the upper half of the body would be visible.

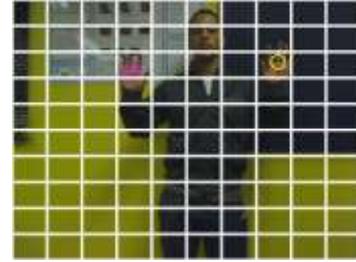


Figure 3: A grid of RRP applied to an image

Each region is assigned to a sub-image that allows features relating to that region to be extracted when needed. It is linked to a neighbouring region, similar to the 8-connectivity labelling described in the previous section. This association allows the immediate neighbourhood to be searched when tracking, thereby improving tracking speed and not searching the entire image. By searching the neighbouring region only it is possible to track multiple objects, even though they share similar characteristics. When comparing RRP to the tracking process that uses Bayesian filters, RRP differs in that it does not assume a constant velocity of the tracked object and allows the tracked object to change direction at any instance. Furthermore, each region contains an “activated” flag that identifies the regions that most likely would contain the tracked object. In addition, when the system is used to track the hands, each region has a flag that identifies the region in which the right hand has been detected and one in which the left hand has been detected. The algorithm also allows for more flags to be added when three or more objects need to be tracked.

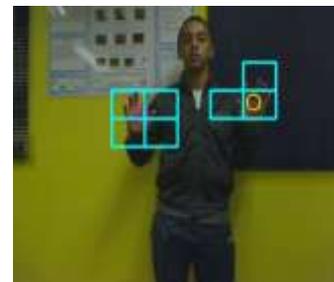


Figure 4: The “activated” regions are highlighted in blue squares. The pink circle identifies the right hand and the yellow circle identifies the left hand

The idea behind the method is to use RRP to segment areas of high likelihood in which the hands can be found (see Figure 4). The tracking procedure would operate by searching for available skin-identified clusters in the immediate neighbourhood. This procedure would allow the

hands to be tracked from one region to the next. If a cluster is not present in the immediate neighbourhood, then regions in the immediate neighbourhood would be searched for one containing the highest density of skin-coloured pixels. The pseudo code of the state transitions using the algorithm is shown in Figure 5.

```

Step 1: Identify pixels that have moved in the frame.
    • If a region contains any pixels that have moved,
        ◦ Then "activate" the region.
    • Else,
        ◦ "De-activate" the region.
Step 2: Identify skin-coloured pixels in a frame.
    • If a region contains any skin-coloured pixels, and is
      "activated"
        ◦ Then leave the region "activated".
    • Else,
        ◦ "De-activate" the region
Step 3: Track the hands in each frame
    • If a skin-cluster is present in the current region and the
      region is "activated"
        ◦ Then select the region.
    • Else if a skin-cluster exists in one of the neighbouring
      regions and the region is "activated",
        ◦ Then select the identified region.
    • Else if the current region does not contain a skin-cluster,
      nor any of the regions in the neighbourhood,
        ◦ Then check whether the current region or
          regions in the neighbourhood has the highest
          skin-colour pixel density, and is "activated",
        ◦ Select the region with the highest skin-colour
          pixel density.
    • Else, if the current region and the regions in the
      neighbourhood are not "activated",
        ◦ Then the tracked hand has not moved and the
          current region is reselected, provided that it
          contains skin-coloured pixels.

```

Figure 5: The pseudo code of the state transitions using RRP

C. Dealing with occlusion

Occlusion is a common problem when tracking body parts due to the high dimensionality of the various body parts. In order to deal with occlusion, the shape of the hand is used to distinguish between the right and left hand. While tracking the hands in each frame, the regions containing the hands are converted to grayscale, noise is reduced by Gaussian blur and edge detection is applied to identify the shape of the hand. The edge features of the shape of hand are extracted and placed into a feature vector. For each feature vector, a label "0" is given if the features belong to the right hand and a label "1" if the features belong to the left hand. The feature vectors are then placed into a data file, and the model is subsequently trained using the radial basis function kernel in an SVM. In the framework, when both the right and left hand share the same region it is detected as occlusion due to the quantisation into larger regions, as shown in Figure 6(a). While the hands are in a state of occlusion, the SVM model is not trained. When the hands separate from the state of occlusion and are found to be in more than one region, the neighbourhood of the occluded region is "activated", and if it contains a contour, then these regions are predicted using the trained model, as shown in Figure 6(b).



Figure 6(a): The occlusion state



(b): The separated state after occlusion

After the regions have been predicted, the tracked regions are separated and the tracking procedure continues tracking the right and left hands independently.

IV. EXPERIMENTAL ANALYSIS AND RESULTS

This section describes experiments that were carried out to evaluate the independent hand-tracking framework in a sign-language recognition prototype. In the experimental setup, a single Logitech webcam connected to a notebook was used. This setup allowed for portability and the capturing of video sequences in unconstrained environments with varying levels of illumination. The video sequences were captured at approximately 12–14 fps with an image resolution of 640x480 pixels, and an average of 77 frames per video sequence.

The evaluation was based on ten SASL isolated gestures, carefully selected from the *Fulton School for the Deaf SASL Dictionary* [10]. The set of selected isolated gestures includes signs that involve the use of a single hand, both hands and the intersection between two hands. It was important to use isolated gestures in these experiments and not continuous sign language gestures so that the proposed framework could be evaluated. In isolated gestures, gestures begin and end in the neutral pose, with arms and hands at the side of the body. Continuous sign language gestures consist of more than one isolated gesture with a single start and end pose. The recognition of continuous sign language gestures will require that the framework should include a tracking failure and recovery mechanism, as well as the ability to detect transitions between isolated gestures. The recognition of continuous sign language gestures will thus not be considered, but will be dealt with in future work.

The ten SASL isolated gestures were performed by six individuals (three males and three females) with different body types and skin-colour tones, ranging from a very light to a very dark skin-colour tone. Each individual performed the ten gestures twice, resulting in two sets of 60 video sequences that were used for testing. The aim of capturing each gesture twice was to compare the two sets, named Test1 and Test2 in the following experiment.

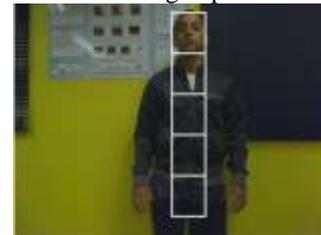


Figure 7: An example of the body proportions theory proposed by Da Vinci

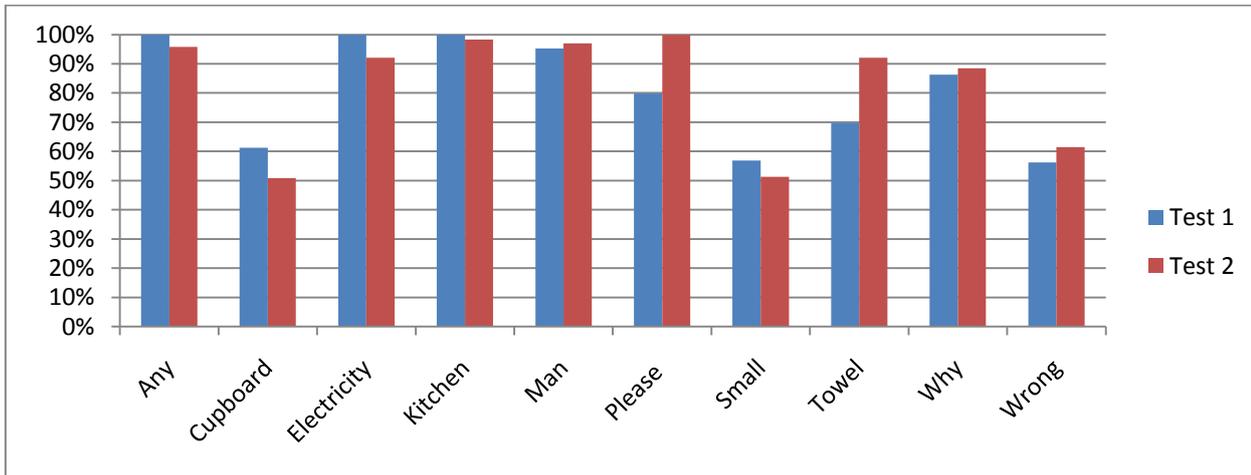


Figure 8: Comparison between the average tracking success rate between the first and second test on the ten isolated SASL gestures

Table 1: Results of the tracking evaluation of the framework in the first test set (%)

	Signer A	Signer B	Signer C	Signer D	Signer E	Signer F	Avg.
Any	100	100	100	100	100	100	100
Cupboard	48.4	45	46.9	97.4	65.2	64.6	61.2
Electricity	100	100	100	100	100	100	100
Kitchen	100	100	100	100	100	100	100
Man	100	100	71.2	100	100	100	95.2
Please	56.6	100	54.1	68	100	100	79.8
Small	39.4	59.4	37.5	88	48.3	68.2	56.9
Towel	31.8	81.5	38.2	77.7	100	89.0	69.7
Why	57.7	100	89.0	71.2	100	100	86.3
Wrong	69.0	57.3	33.3	53.4	54.4	69.4	56.1
Avg.	70.2	84.3	67.0	85.6	86.8	89.1	80.5

Table 2: Results of the tracking evaluation of the framework in the second test set (%)

	Signer A	Signer B	Signer C	Signer D	Signer E	Signer F	Avg.
Any	87.2	100	100	100	87.4	100	95.8
Cupboard	36.8	37.7	36.4	39.1	98.6	56.4	50.8
Electricity	57.1	100	100	95.0	100	100	92.0
Kitchen	89.9	100	100	100	100	100	98.3
Man	98.8	100	82.9	100	100	100	97.0
Please	100	100	100	100	100	100	100
Small	46.1	38.7	32.5	50.7	93.9	46.1	51.3
Towel	76.5	100	100	91.8	90.2	93.9	92.0
Why	60.9	100	100	69.9	100	100	88.4
Wrong	62.5	61.4	72.4	62.3	55.9	54.2	61.4
Avg.	71.6	83.8	82.4	80.9	92.6	85.0	82.73

To evaluate the framework, each video sequence was fed into the system. In the initial frames of each signed gesture, the hands are located next to the body of the signer (neutral pose). According to the body proportions proposed by Da Vinci [9], the hands of an average person in the neutral pose can be found at a position five times the length of the head, starting from the tip of the head, as seen in Figure 7. The right hand would be the hand on the right side of the body (from the perspective of the person whose body it is) and the left hand would be the hand on the left side of the body.

After finding the hands, they are tracked using the prototype of the proposed framework. For each tracked location, a blue square identifies the tracking of the right hand and a green square identifies the tracking of the left hand, shown in Figure 6(b). When the hands intersect and self-occlusion occurs, only a green square should be seen (see Figure 6(a)). However, when the hands are separated from the occlusion state, the blue and green square should track the right and left hands, respectively, again.

After feeding the video sequence to the system, the outcome was analysed by someone not related to the research. Similar to other researchers in this field [8, 12], subjective evaluation was used to perform the analysis. Thus, a frame was deemed correct if the right hand was enclosed by a blue square and a left hand was enclosed by a green square. If this were not true, the frame would be labelled as incorrect.

The number of times a success was obtained was recorded and divided by the number of frames in the video sequence. The result was the average tracking success rates per signed gesture. A complete list of the tracking success rates for

each signed gesture in the first and second evaluation is presented in Table 1 and Table 2, respectively. The results in Figure 8 show that every sign obtained an average (avg.) success rate greater than 50% in both evaluations. In addition, three signs in the first evaluation – “any”, “electricity” and “kitchen” – and one sign in the second evaluation – “small” – obtained the highest possible success rate: 100%. This suggests that the hands were tracked independently throughout the video sequence. Furthermore, four signs in the first evaluation – “man”, “please”, “towel” and “why” – and six signs in the second evaluation – “any”, “electricity”, “kitchen”, “man”, “towel” and “why” – obtained an average success rate of between 70% and 90%. In these signs, the hands were tracked independently, but were lost due to rapid hand motions and were later re-tracked when the paths crossed between the hand and the assumed location. Three signs in both evaluations – “cupboard”, “small” and “wrong” – obtained an average success rate below 70%. These video sequences contained gestures where the hands intersected and self-occlusion was more prominent. This suggests that the hands were either not tracked correctly or not predicted correctly after self-occlusion occurred.

When analysing the accuracy according to each individual signer, the majority of the subjects obtained an average success rate greater than 80%, with Signer E obtaining the highest average success rate of 92.60%. The results indicate that the hand-tracking framework performed equally well on each body type, as well as on the different skin-colour tones.

When comparing the results between the two evaluations, the results were comparable. However, there is a distinct

difference when tracking between some signers on some signs, which suggest that the framework is not negatively affected by certain signs, but rather affected by the rapid movements of the signer, tracking failure or incorrect predictions by the SVM. In half of the total number of signed gestures, the system obtained a tracking success rate of 100%, thus being able to fully track both hands independently throughout the video sequence. Overall, the system achieved an average accuracy of 82%.

V. CONCLUSION

In this paper, a novel learning-based framework to independently track the right and left hands was presented. Connected components analysis was used to reduce noisy areas and cluster together skin-identified contours. In addition, a novel segmentation algorithm referred to as region reference points was introduced. The algorithm divides an image into a number of regions and holds information regarding motion, skin colour and cluster centres for each region. The information was collectively used to identify potential regions where the hand may be found. The identified region was subsequently used as a reference point, thereby limiting the search area and assisting in the tracking of the hands. The algorithm furthermore assists in identifying which features of the hand are “good” features to learn. An SVM was used to learn these features and generate a model that was later used to distinguish between the right and left hands when they intersect and occlusion occurs. The framework was evaluated on ten SASL isolated signed gestures performed by six individuals, where each individual performed each gesture twice, so that a comparison could be made between the gesture sets. The results between the two evaluations were comparable and resulted in an overall tracking success rate of 82%.

Although positive results were obtained, the framework requires several improvements. Tracking often fails when light is inconsistent throughout a frame, thus leading to weak back projection. A means to identify this condition is therefore required. Furthermore, should tracking fail, it should be immediately detected and a recovery mechanism should be employed. This would allow the hands to be re-identified and tracked accordingly. It is furthermore believed that velocity would play an essential part in continuously tracking the hands and should be added to the information retrieved from each region when using RRP.

VI. ACKNOWLEDGEMENTS

This research is funded by Telkom and the Telkom/CISCO/Aria Technologies/THRIP Centre of Excellence.

VII. REFERENCES

- 1) I. Achmed, “Upper Body Pose Recognition and Estimation towards the Translation of South African Sign Language”, unpublished Masters Thesis, University of the Western Cape, South Africa, 2010.
- 2) I. Achmed and J. Connan, “Upper body pose estimation towards the translation of South African sign language”, in *Proceedings of the Southern Africa Telecommunication Networks and Applications Conference*, 2010, pp. 427—432.
- 3) J.C. Aker and I.M. Mbiti, “Mobile phones and economic development in Africa”, *The Journal of Economic Perspectives*, pp. 207—232, 2010.
- 4) M.S.M. Asaari and S.A. Suandi, “Real Time Hand Tracking System Using Predictive Eigenhand Tracker”, *Australian Journal of Basic and Applied Sciences*, vol. 6, no. 2, pp. 190—198, 2012.
- 5) C.S. Chua, H. Guan and Y.K. Ho, “Model-based 3D hand posture estimation from a single 2D image”, *Image and Vision Computing Journal*, vol. 20, no. 3, pp. 191—202, 2002.
- 6) F. Dadgostar and A. Sarrafzadeh, “An adaptive real-time skin detector based on Hue thresholding: A comparison on two motion tracking methods”, *Pattern Recognition Letters*, vol. 27, no. 12, pp. 1342—1352, 2006.
- 7) L. Di Stefano and A. Bulgarelli, “A simple and efficient connected components labeling algorithm”, in *Proceedings of the International Conference on Image Analysis and Processing*, 1999, pp. 322—327.
- 8) M. Donoser and H. Bischof, “Real time appearance based hand tracking”, in *Proceedings of International Conference on Pattern Recognition (ICPR)*, 2008, pp. 1—4.
- 9) R.D. Green and L. Guan, “Quantifying and recognizing human movement patterns from monocular video images-part ii: applications to biometrics”, in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 2, pp. 191—198, 2004.
- 10) S. Howard, *Finger talk-South African sign language dictionary*. South Africa: Mondri, 2008.
- 11) M. Lee and I. Cohen, “Human upper body pose estimation in static images,” in *Proceedings of the European Conference on Computer Vision*, 2004, pp. 126—138.
- 12) Y. Liu and P. Zhang, “Hand Gesture Tracking using Particle Filter with Multiple Features”, in *Proceedings of the International Symposium on Intelligent Information Systems and Applications (IISA'09)*, Qingdao, P. R. China, Oct. 28-30, 2009, pp. 264—267.
- 13) J. Relethford, “Human skin color diversity is highest in sub-Saharan African populations.” In *Human biology: An International Record of Research*, vol. 72, no. 5, pp. 773—780, October 2000.
- 14) R.Y. Wang and J. Popović, “Real-time hand-tracking with a color glove”, in *ACM Transactions on Graphics (TOG)*, vol. 28, no. 3, pp. 63—71, 2009.

Imran Achmed is a Telkom/Cisco/Aria Technologies/THRIP Centre of Excellence PhD student at the University of the Western Cape. His research deals with sign language synthesis and novel communication applications for the Deaf and hearing impaired.

Isabella M. Venter is an associate professor and Chair of the Department of Computer Science at the University of the Western Cape.

Peter Eisert is a professor for Visual Computing at the Humboldt University, Berlin, and is heading the Computer Vision and Graphics Group at the Fraunhofer Heinrich-Hertz Institute.