

TOWARDS LOW-COST ROBUST AND STABLE HAND TRACKING FOR EXERCISE MONITORING

Rui Liu and Burkhard Wünsche

*Graphics Group, Department of Computer Science
University of Auckland, Private Bag 92019, Auckland, New Zealand*

Keywords: Hand tracking, Hand segmentation, Feature detection, Perceptual-based colour space.

Abstract: Applications for home-based care are rapidly increasing in importance due to spiraling health care and elderly care costs. An important aspect of home-based care is exercises for rehabilitation and improving general health. However, without caregivers supervising these exercises it is difficult to monitor them, i.e., to determine whether the exercises have been performed correctly and for the prescribed duration. In this paper we present the first steps toward a computer-based tool for monitoring hand exercises. Hand exercises are important for various diseases such as Parkinson disease. While many algorithms exist for gesture recognition, most of them do require special set-ups and are difficult to use for very inexperienced users in home-based environments. In this paper we present a robust hand region segmentation method which represents the first step toward a hand-tracking algorithm. Our solution requires no calibration and is easily set-up. We evaluate its robustness with regard to complex backgrounds, changes in illuminations, and different hand colours. Our results indicate that the robust hand region segmentation provides a solid foundation for monitoring hand exercises.

1 INTRODUCTION

Injuries and diseases such as inflammatory and autoimmune diseases (arthritis), degenerative muscle diseases (Welander distal myopathy), overuse syndromes and neurological damage and diseases (stroke, Parkinson's disease) may cause reduced or complete loss of control of hand and fingers which would be a catastrophic event for patients. For many of the above diseases surgical or drug treatment does not exist, is expensive, or only partially effective. Finger and hand exercises have been shown to be a very effective alternative or complementary treatment (Wessel, 2004).

A major problem with exercise training is the lack of supervision by qualified instructors over time. While patients are encouraged to exercise at home, many patients lack motivation without monitoring and evaluation of their performance. The popularization of personal computers and web-cams offers a solution to this problem. We can use existing hardware in combination with animation technology to teach patients the correct exercises and to use web-cam based gesture recognition technology to evaluate the correctness of the exercises, to give instructions

for improvement, and to monitor the success of the exercise program over time. Such home-based care applications are rapidly increasing in importance due to spiraling healthcare and elderly care costs and are becoming an integral part of government health policies (UK Department of Health, 2009).

Our goal is to provide an affordable platform for novice users and patients to set up their own exercise environment easily. With the help of a region-growing approach and a perception-based colour space, a calibration-free application is presented based on web-cam input.

2 LITERATURE REVIEW

A large variety of hand tracking algorithms have been proposed. A good survey is given in (Mahmoudi and Parviz, 2006). Two important categories are marker-based and marker-less methods.

2.1 Marker-based Tracking Methods

Marker-based hand tracking algorithms require the user to wear point or area markers such as LED-

gloves and colour coded markers. An example is the inexpensive 3D articulated hand tracking system by (Wang and Popović, 2009). The hand position and motion of the users are tracked by wearing a glove with a custom colour pattern. Robustness and interactive speed are illustrated by applications such as driving an animated character using finger motion.

Marker-based tracking applications can provide highly accurate results, but the need for auxiliary devices (markers, gloves) can be inconvenient for the user and often requires some type of calibration.

Our preliminary studies and interviews with healthcare and geriatrics specialists suggest that patients and elderly prefer a marker-less application. In order to make current marker-based methods acceptable the marker devices must become cheaper, more readily available, and more flexible and easier to use.

2.2 Marker-less Tracking Methods

Without the use of markers, alternative techniques must be employed in order to identify a hand on the camera image and determine its 3D position.

The easiest way to identify (potential) hand shapes is by using a skin colour classifier. A large amount of literature exist on this topic. Some of the more recent surveys and comparative studies include (Kakumanu et al., 2007; Vassili et al., 2003). Recently a perception-based colour space has been proposed (Chong et al., 2008) which can be used to make image processing techniques more robust to different lighting conditions.

By using a 3D hand model and searching for a mapping (Stenger et al., 2001), the (potential) hand shape can be verified and its 3D position and orientation can be determined. An opposite approach can also be taken by matching a hand image to a set of hand templates (Stenger et al., 2006).

Due to its robustness to noise and lighting changes, Haar-like features are also widely used to achieve fast matching and rapid elimination of wrong candidates (Chen et al., 2007).

3 DESIGN

The results of the review in the previous section demonstrate that with the current limitations a marker-less method using a monocular web cam is the most appropriate set-up for home-based healthcare applications. Designing such a system requires three steps:

1. A robust way for detecting the hand image (it should work for different backgrounds, different illuminations and different skin colours)

2. Pose estimation

3. Hand tracking

In this paper we only describe the first step, i.e., a robust way for detecting hand images using a low-cost web-cam.

3.1 Overview

Our hand region segmentation algorithm utilizes a pixel-based region-growing technique. A seed point is obtained by displaying a target (e.g., hand template) in the center of the screen and requiring the user to move their hand until it matches the target. This type of initialization does not require technical know-how and can be accomplished by anyone able to do hand exercises. The seed point is added to the currently detected hand region and all its neighbours are stored in a queue data structure. In each iteration we now remove a pixel from the queue, test whether it belongs to the hand region, and if yes add its not yet tested neighbours to the queue. In order to decide whether a pixel belongs to the hand region we use a perception-based colour space and compute the current pixel's colour distance to the mean value of the hand region detected so far. Only pixels with a distance below a given threshold are added. The algorithm continues until the queue is empty.

3.2 Classification Criteria

Two pixels are considered to belong to the same region if their colour distance is within a given threshold. The colour distance is computed using the ℓ_2 -norm and transformation function F which converts a pixel's colour into the perception-based colour space from (Chong et al., 2008). The perceptual colour distance d between two colours \vec{x} and \vec{x}' is hence:

$$d(\vec{x}, \vec{x}') = \|F(\vec{x}) - F(\vec{x}')\| \quad (1)$$

Therefore, if

$$d' \leq d(\vec{x}, \vec{x}') \quad (2)$$

the pixel is classified as a point within the region in which the seed point belongs to, otherwise the next pixel will be tested.

Finding a suitable value for thresholding is difficult. Making the inclusion of pixels only depends on comparison with its direct neighbours is error prone because of "colour leaking" (a continuous change of colours from hand regions to other regions due to shadows, highlights and other influences). Gradient based methods are common in other segmentation applications, but work unsatisfactory because of strong local illumination changes over the hand

surface. We used instead a mixed approach combining a colour distance criteria with edge detection. The colour distance is computed by using the mean colour value of the current hand region for comparison. We apply thresholding to the l_2 -norm of colour differences in the perception-based colour space. We also experimented with applying thresholding to each channel separately, but we were unable to find a consistently superior combination and hence decided to keep using the l_2 -norm. In order to avoid ‘‘colour leaking’’ over object boundaries we compute image edges using a Canny edge detector. New pixels are only accepted if they do not lie on an edge.

4 IMPLEMENTATION

We implemented our algorithm in C/C++ using the Open Source Computer Vision (OpenCV) library. The following results were obtained using an iMac (Intel(R) Core(TM)2 Duo CPU E8335 2.66GHz, 3.00GB RAM) with an inbuilt 640×480-pixel VGA resolution *iSight* camera and a laptop (Intel(R) Centrino CPU 1.60GHz, 1.00GB RAM) with a low quality external web-cam of unknown brand and with 320×240 resolution.

5 RESULTS

We tested our algorithm using web cam input and AVI video clips and in all experiments processing was performed in real time (i.e., 20 frames/second). We have not yet determined the maximum speed possible.

We compared our algorithm with segmentation results obtained by using different skin classifiers using different colour spaces (RGB, YCbCr and CIE XYZ). In all cases the results were inferior to our method as illustrated in Figure 1.

Figure 2 shows the results of testing our segmentation method under different lighting conditions. The identification of the hand region works similarly well in both examples. However, under dim light the skin region around the wrist is missing.

We tested different backgrounds including a textured desk chair, a computer lab, and the user’s face (figure 3). Surprisingly the algorithm works well even when the hand covers the face, which seems to be a much harder segmentation task. The results indicate that the algorithm is robust, but requires some fine tuning and modifications might be necessary for highly saturated image regions.

Figure 3 (right) demonstrates the robustness of our algorithm when applied to different skin colours.



Figure 1: Comparison of the hand detection and segmentation results using our region growing technique (bottom-right) and the skin classifiers from (Hsu et al., 2002) (top-left), (Yang et al., 1998; Stauffer and Grimson, 1999) (top-right) and (Kovac et al., 2003) (bottom-left).



Figure 2: Segmentation results under indoor conditions with daylight and fluorescent ceiling lamps (first row) and dim environment without fluorescent ceiling lamps (second row).

We used users from three different ethnic groups and found that the algorithm worked similarly well in all cases.

Section 3 described our region growing algorithm and explained that the testing of candidate pixels is performed by computing their colour distance to the mean colour of the current hand segment. An alternative application is to only consider the neighbours of the candidate pixels which are within the segmented region. We found that using the mean constraint results in better segmentation for the wrist region, but causes one artifact in the finger region. Overall we



Figure 3: Evaluation of the sensitivity of our hand segmentation algorithm to different backgrounds and skin colours.

found that using the mean value gives visually more pleasing and consistent results. A quantitative study is necessary to confirm this observation.

6 CONCLUSIONS

We have presented a novel region-growing method for hand segmentation. The main difference to previous methods is the use of a perception-based colour space and a classification function using edge detection information in combination with mean colour values rather than neighborhood information.

We have compared our method with traditional skin classifiers and demonstrated that it is superior. While the comparison has to be treated with caution due to the lack of edge information when using the skin classifiers, we believe that the results still demonstrate the usefulness of our method as a starting point for hand tracking applications.

We evaluated our method for different illumination conditions, backgrounds and skin colours, and found that it is sufficiently stable and forms a suitable foundation for low-cost hand tracking applications. We have started to experiment with active contour models, but found less improvement than expected due to the high curvature of the hand silhouette.

7 FUTURE WORK

The next steps for our application are to estimate 3D motion of the hand by using a 3D hand model with kinematic constraints. In order to resolve ambiguities in the mapping process we plan to use simplified markers, e.g., stickers, coloured rubber bands or lipstick marks. This type of markers gives less reliable results than traditional markers and we can not expect users to place them correctly - however, such markers are easy to use, cost effective, do not constrain mobility, and can be used even if the patient suffers from conditions such as swellings and sores on the hand.

REFERENCES

- Chen, Q., Georganas, N., and Petriu, E. (2007). Real-time vision-based hand gesture recognition using haar-like features. In *Instrumentation and Measurement Technology Conference Proceedings, 2007. IMTC 2007. IEEE*, pages 1–6.
- Chong, H. Y., Gortler, S. J., and Zickler, T. (2008). A perception-based color space for illumination-invariant image processing. *ACM Trans. Graph.*, 27(3):1–7.
- Hsu, R.-L., Abdel-Mottaleb, M., and Jain, A. K. (2002). Face detection in color images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:696–706.
- Kakumanu, P., Makrogiannis, S., and Bourbakis, N. (2007). A survey of skin-color modeling and detection methods. *Pattern Recogn.*, 40(3):1106–1122.
- Kovac, J., Peer, P., and Solina, F. (2003). Human skin color clustering for face detection. In *EUROCON 2003. Computer as a Tool.*, volume 2, pages 144–148.
- Mahmoudi, F. and Parviz, M. (2006). Visual hand tracking algorithms. In *GMAI*, pages 228–232.
- Stauffer, C. and Grimson, W. (1999). Adaptive background mixture models for real-time tracking. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, page 252.
- Stenger, B., Mendona, P. R. S., and Cipolla, R. (2001). Model-based 3d tracking of an articulated hand. *Computer Society Conference on Computer Vision and Pattern Recognition*, 2:310.
- Stenger, B., Thayananthan, A., Torr, P. H. S., and Cipolla, R. (2006). Model-based hand tracking using a hierarchical bayesian filter. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1372–1384.
- UK Department of Health (2009). UK telecare policy and strategy. <http://www.pasa.nhs.uk/PASAWeb/Productsandservices/Telecare/Governmentpriorities.htm>.
- Vassili, V. V., Sazonov, V., and Andreeva, A. (2003). A survey on pixel-based skin color detection techniques. In *Proc. Graphicon-2003*, pages 85–92.
- Wang, R. Y. and Popović, J. (2009). Real-time hand-tracking with a color glove. In *SIGGRAPH '09: ACM SIGGRAPH 2009 papers*, pages 1–8, New York, NY, USA. ACM.
- Wessel, J. (2004). The effectiveness of hand exercises for persons with rheumatoid arthritis: A systematic review. *Journal of Hand Therapy*, 17(2):174–180.
- Yang, J., Lu, W., and Waibel, A. (1998). Skin-color modeling and adaptation. In *ACCV '98: Proceedings of the Third Asian Conference on Computer Vision-Volume II*, pages 687–694, London, UK. Springer-Verlag.