Comparison of hand segmentation approaches in grasp images obtained with an omnidirectional vision system


* USP, São Carlos School of Engineering, Department of Mechanical Engineering, ivanperissin@gmail.com, Av. Trab. São Carlense, 400 - Parque Arnold Schimidt, São Carlos - SP, 13566-590, Brazil;
** USP, São Carlos School of Engineering, Department of Mechanical Engineering, vivianecroma@gmail.com, matheusbrezende7@gmail.com, lmpedro@ufscar.br, gcaurin@sc.usp.br, Av. Trab. São Carlense, 400 - Parque Arnold Schimidt, São Carlos - SP, 13566-590, Brazil;

Special Issue: 6th Brazilian Meeting of Biomechanical Engineering (ENEBI 2018-Part I)

Abstract

Hand grasp evaluation is a great challenge in clinical and therapeutic applications. Conventional acquisition devices present inherent and restrictive difficulties for their use. As an alternative, an instrumented cup with an omnidirectional vision system is proposed, capable of capturing hand images during the grasp movement, as well as acquiring kinematic data. The process of hand segmentation from the background scene is a critical step for future analysis, aiming to quantify the grasp in unprecedented manners. This article presents a comparison of 24 different hand segmentation methods using 102 images acquired in experiments performed with 51 children. The final results showed that the methods, color distance in red and a modification in the HSV color space, obtained a segmentation performance index higher than 95.5% Youden score, a promising result for our application and future studies. DOI: https://doi.org/10.24243/JMEB/4.2.210

1 Introduction

Hand grasp is one of the main functions for performing tasks that handle objects or tools. Several diseases (inherited, congenital or acquired) can lead to partial or complete loss of this function, compromising the accomplishment of activities of daily living (ADLs), and consequently, the life quality of the affected people. Surgical interventions and physiotherapeutic treatments are recommended to recover, even partially, impaired functions [1][1]. In this context, it is necessary to evaluate the grasp quality to, in a first moment, quantify the motor deficit and, in a second moment, evaluate the progress during and after the clinical intervention [2].

Regarding the grasp evaluation, there are two main aspects: i) the position of the hand, that is, the recruitment of the fingers and their joint angles; ii) the force applied and the pressure distribution along the contact area. In the first

*Corresponding Author: Ivan C. Perissini
Email Address: ivanperissin@gmail.com
case, there are devices that allow the measurement of the angular position of the finger joints during grasping, e.g. motion trackers, which capture with the help of cameras the spatial position of markers strategically placed on the hand, and instrumented gloves, whose most widespread technology is composed of angular deflection sensors positioned on the joints. Both technologies mentioned present disadvantages and difficulties in clinical use [3]. Positioning the markers on the patient's hand requires time and experience from the therapist, and gloves cannot be washed, which represents contamination risks.

As an alternative to the current technologies available for grasp evaluation, and in an attempt to solve its application problems, we have proposed a cylindrical format object mimicking a glass cup, instrumented with an omnidirectional vision system and an inertial measurement unit (IMU) [4]. The image acquisition system, composed of a high-definition camera and a hyperbolic mirror, captures images of the hand grip during the manipulation of the device, to obtain hand posture and contact area, and correlate this information with kinematic data such as acceleration and orientation from the IMU. For the image processing, in order to extract the cited information, it is necessary a proper image segmentation, i.e. separation of the object of interest, the hand, from the background scene.

This article presents a comparative study of 24 image segmentation methods, advancing the preliminary work [5], in which color-based segmentation approaches were used, specifically, skin detection techniques. It is important to highlight that several factors can influence the skin appearance in images, such as lighting, background color, shadows, ethnicity, individual characteristics such as age and gender, individual posture and positioning, and camera sensor characteristics. Therefore, the proper restrain of the environment conditions represent a crucial role in the experiments.

There is a large number of studies that describe the significance of the color space choice for the segmentation process [6]-[8], [9]-[10] published relevant reviews related to skin pixel detection methods and techniques, while [11] demonstrated that better segmentation rates were obtained with new proposed color spaces, based on the combination of conventional models’ channels.

In this context, the objective of this work is to make a deeper analysis of hand segmentation methods, using a superior number of color spaces and a larger image dataset than our previous study. A comparison between 24 different approaches is presented to determine which ones provide the best performances for hand segmentation. A dataset of 102 images of 51 children were used in the experiment, in order to determine which classification algorithms obtained the best performances, and, therefore, identify the conditions that led to the most accurate detection. For the analysis of the results, the Youden Index was used [12], a performance index, that correlates the recall rate and specificity of each approach to determine the overall efficiency of the analyzed method.

2. Materials and methods

2.1. Ethical procedures, participants and dataset

The images used in this study are part of the observational study approved by the Human Research Ethics Committee of the Federal University of São Carlos (document number 1.337.492). The data collection was carried out in a private school in the city of São Carlos. All of the guardians signed the Legal Consent Form, and the children signed the Consent Assent. A total of 51 healthy, full-term born, and typically developing children (aged 5–10 years) participated in this study.

A total of 102 images of this database were selected in accordance with some intended characteristics. The grip position had to be properly and fully made at the beginning of the of the water-drinking protocol movement, as described in [4], moment in which, the grasp is considered stable. The main aspect of the protocol is to capture the natural movement of the water drink task and as such no further recommendations were addressed to the participants. Gender and age of the participating children were not considered.

2.2. Color models

According [13], a color model also called color space or color system, is a specification of a coordinate system and a subspace contained in this system where each color is represented by a single point to standardize the color
A wide variety of color models has been considered for this study experiments, aiming to increase the discrimination between skin and non-skin classes. In general, all these color models are obtained by transformations of the RGB model. Among them, we selected the ones used in computer vision applications for: skin detection [10], [14]-[16], agriculture [6], some novel models that combine different color channels and also a color distance function. In total, seven color models have been assessed in addition to RGB: rgb, HSV, XYZ, YCrCb, Lab, Luv, TLS [7], [13].

Derived from the HSV model, it was proposed a modification where the hue axis was rotated by 30º clockwise, since the scale starts and ends in the color red that correspond to the skin tone. And, as an additional proposal, the color distance function was used. In this case, the Euclidean distance between each pixel of the image was calculated with a red color reference based on the Cielab-76.

Besides the traditional color models, other segmentation strategies based on manipulations of their color channels and alternatives segmentation approaches, have been proposed and tested. A new model based on the YCrCb space formulation, called CrCgCb, was used. The achromatic channel of the original space Y was replaced by an additional chromatic channel associated with the green color (Cg), applying the equation Cg = 1.211×(G - Y) + 0.5.

Other proposals were inspired by vegetative indexes, as the application of the Excess Red index (ExR) [17], and two adaptations of the Excess Green index (ExG) [18] to highlight the red color instead of the original green color, using the RGB and CrCgCb color models, instead of the intended rgb. At the end, the operations performed to obtain the modified indexes were, for Excess Green with red in RGB, 2×R-G-B and for Excess Green with red in CrCgCb, 2×Cr-Cg-Cb.

3. Proposed System

The proposed system (Figure 1) was developed in the software Wolfram Mathematica 10.4, which prepares and pre-processes the original image for the experimental campaign. The software generates an unwrapped version of the image (test samples), and calculates and evaluates the segmentation performance. For each test sample all the different segmentation approaches are applied and the final results are them evaluated by comparison with an ideal reference image (ground truth). In sequence, a brief description of the system and image processing steps are covered, in equivalence to previous studies [5].

The original omnidirectional images are initially unwrapped into panoramic images, according to the algorithm described by [19]. The lower part of this image is then removed by a 30% crop, considering that the central portion of the omnidirectional image, relative to the mirror center, is represented by a small number of pixels that generate slow resolution at the bottom of the panoramic image.

In sequence, the 16 proposed color transformations are applied for each sample and the respective classification is then performed. Some cases a manually defined threshold value is fixed, while in others the Otsu method [20] or the K-means method is used. In manual and Otsu thresholding, the channel or combinations of channels with the highest discriminative power for skin detection is selected, while the k-means clustering algorithm is applied to all channels and was configured for partitioning the pixels into two groups (k = 2). After these segmentations, a specific classification rule is applied resulting in binary images of skin (1) and background (0). This image then passes through a post-processing stage that contains: median filter, morphological operation of erosion, labeling of regions, removal of small areas, concluding with dilation with the same structural element used in erosion, thus preserving the original dimensions.

At the end of this stage the test image is obtained, which is equivalent to the result of each classification approach. The last step evaluates the segmentation approaches, comparing the resulting image with its respective ground truth image, generated by the manual selection of specialists, defining the ideal separation of the hand and the background. The performance metric adopted in this study was the Youden index [11], a performance index that makes
Ivan C. Perissini et al.

a balance between the recall rate and specificity. This index varies between -1 and +1 (equivalently, -100% and +100%), and as higher the result, better is the performance of the evaluated method.

Fig. 1 Proposed system framework

4. Results and discussion

In total, 24 approaches for hand segmentation were used in this study, with 16 color transformations and 3 classification methods (manual thresholding, Otsu and k-means). The proposed evaluation method was applied for each approach to all the 102 images of the study. The average values of the results, as well as the standard deviations of the error and the Youden index, were computed, and can be seen in Table 1, organized by the average Youden index in descending order. First thing noticed, is that there is a high correlation between Youden's Index and the error (as higher the Youden, smaller is the error, and vice versa). It is also observed a large variation in the Youden index results (Figure 2), showing relevance to the proposed study, with responses varying from unsatisfactory segmentation values close to 56% to a near complete segmentation with 96%.

The Otsu binarization in the gray scale (approach 1) establishes an important reference for the analysis of a segmentation based only on the achromatic aspects of the images. Therefore, results that are better than this reference would demonstrate the contribution of the chromatic aspects for the segmentation. In contrast, given the control characteristics established for data collection, the authors consider that the approaches with results lower than this reference would represent invalid attempts of hand segmentation.

The classic HSV model, commonly used in color segmentation, did not present good results, however, the modified HSV showed good values independent of the classification method, surpassed only by the approaches related to color distance with proven statistical difference (p < 0.05).

A pertinent analysis is the comparison of the current data with the best result of previous study [5], the YCrCb color model. This result was surpassed in this study by some new adopted approaches, being them: the transformations based on color distance, the modified HSV, the proposed Excess Green with red in RGB and TLS with K-means, all with significantly better results (p < 0.05).

It can be observed that for certain color spaces the classifier presented a considerable difference in the results, as in TLS, where the k-means method presented better performance due to the use of all the channels for the classification. However, for some approaches the classifier did not show a significant influence, as seen in RGB, HSV modified and color distance. In the latter case, the classification approaches were applied in only one channel, which justifies this response. In this context, it is recommended to use Otsu because it has a lower computational cost.
Expanding the analysis, the average error and Youden index of the 24 approaches were also computed for each of the 102 images. For most images the methods in average performed well, but it can be observed that both, color transformation and classification method have a great impact on some particular conditions. These image results were ranked allowing a visual identification of the image characteristics of the best and worst average results. In general, the best results were accompanied by a lower standard deviation, which shows that most of the methods were successful in the proposed task of segmenting the hand. In contrast, the images with worse results had high standard deviation.

When evaluating the worst performing images, it is noted that in most cases the face is connected to the hand and the fingers appear separated, which worsens the response of the post-processing, besides that, in many of these images the hand occupies a large part of the image, which may impair the performance of the classifier. For some approaches the noises led to large concentrations of holes that caused the post-processing to remove large portions of the hand with extreme cases of entire non-detection. Fig. 3 (a), which obtained the lowest average Youden index and the highest mean standard deviation, demonstrates the discussed effects. The images with the best average results showed opposed characteristics, such as face separated from the hand, fingers close to each other, uniform background,
similar skin tone and balanced proportion between background and hand. These conditions can be observed in the image that obtained the best performance, showed in Fig. 3 (b).

(a) Original image RGB image Application of Otsu Post-processing
(b) Original image Color distance image Application of Otsu Post-processing

Fig.3 (a) worst performance; (b) best performance

5. Conclusions

The results showed that some proposed approaches presented an excellent performance for our application needs. The methods associated with color distance, modified HSV, proposed method Excess Green with red in RGB and TLS with K-means presented promising and robust results, and regardless of the image conditions, obtained a high Youden index and low error rate. From the classical color segmentation approaches, only HSV was considered suitable for the application, provided that the beginning of the hue scale is properly changed to different values from the skin tones (red).

These results meet the expectations of the GULM sensor by presenting robustness and efficiency in the hand segmentation task, with superior results than the ones obtained in previous studies. In addition, the expanded number of experimental samples and hand segmentation approaches proposed in this study provide greater confidence in the results. With respect to the scene analysis, some particular conditions resulted in unacceptable segmentation results for certain approaches, creating a significant performance variability, highlighting the importance of choosing the appropriate color space. These challenging scene features show the importance of environmental control at the data collection procedure.

The unwrapping method of the omnidirectional images along with the particular scene conditions, and the need to perform hand only segmentation, fostered the investigations that were the basis for this study. The results showed relevant advances to the GULM sensor project and consolidates a standard approach for the grasp evaluation, assisting upper limb rehabilitation process. Future works with image processing of the GULM sensor can use as standard procedure the best identified segmentation approaches.

Acknowledgements

The authors would like to thank CAPES, the Royal Society (Project Newton Advanced Fellowship, process NA140231) and FAPESP CEPID Process 2013/07276-1.

Funding

The authors are solely responsible for the information included in this work.

References


