HAND IMAGE SEGMENTATION BY MEANS OF GAUSSIAN MULTISCALE AGGREGATION FOR BIOMETRIC APPLICATIONS

Alberto de Santos Sierra, Carmen Sánchez Ávila
Group of Biometrics, Biosignals and Security, GB2S
Centro de Domótica Integral, Universidad Politécnica de Madrid, Madrid, Spain

Javier Guerra Casanova, Gonzalo Bailador del Pozo
Group of Biometrics, Biosignals and Security, GB2S
Centro de Domótica Integral, Universidad Politécnica de Madrid, Madrid, Spain

Keywords: Hand segmentation, Fuzzy multiscale aggregation, Biometrics, Hand geometry, Synthetic database.

Abstract: Applying biometrics to daily scenarios involves demanding requirements in terms of software and hardware. On the contrary, current biometric techniques are also being adapted to present-day devices, like mobile phones, laptops and the like, which are far from meeting the previous stated requirements. In fact, achieving a combination of both necessities is one of the most difficult problems at present in biometrics. Therefore, this paper presents a segmentation algorithm able to provide suitable solutions in terms of precision for hand biometric recognition, considering a wide range of backgrounds like carpets, glass, grass, mud, pavement, plastic, tiles or wood. Results highlight that segmentation accuracy is carried out with high rates of precision (F-measure$\geq$ 88%)), presenting competitive time results when compared to state-of-the-art segmentation algorithms time performance.

1 INTRODUCTION

Biometrics based on hand recognition have received an increasing attention in latter years due to their huge applicability in daily scenarios and the relation between user acceptance and identification/verification rates (Kukula and Elliott, 2005; Kukula and Elliott, 2006).

In addition, hand biometrics system requirements are easily met with a standard camera and hardware processor, so that these systems can be easily adapted to devices like PC, mobile phones and the like.

This paper presents a segmentation method for isolating hand from background in real environments oriented to mobile devices. However, biometrics systems in general strongly rely on the environment conditions such as illumination, background, proximity to sensor and so forth, and therefore, applying biometrics to mobile devices requires to solve more demanding problems in terms of segmentation and invariance to changes in feature extraction with less resources.

The proposed method is based on Gaussian Multiscale Aggregation (GMA) (García-Casarrubios Muñoz et al., 2010), gathering those pixels with similar characteristics under a same cluster, providing a hierarchichal structure along scales to find a proper segmentation where segments correspond to objects within image.

Segmentation involves a database to properly evaluate to what extent the method can isolate objects within an image. In order to assess the proposed method with real scenarios, a synthetic database has been collected with a total of 408000 images, containing samples with different backgrounds like soil, skins/fur, carpets, walls, grass and the like, corresponding to those possible scenarios where hand images can be taken.

The layout of this paper remains as follows: First of all, a literature review is presented under Section 2. The proposed approach is described in Section 3, together with the database used to test and evaluate the algorithm (Section 4) and the evaluation criteria (Section 5). Finally, results in Section 6 and conclusions in Section 7 will end this paper.

de Santos Sierra A., Sánchez Ávila C., Guerra Casanova J. and Bailador del Pozo G..
HAND IMAGE SEGMENTATION BY MEANS OF GAUSSIAN MULTISCALE AGGREGATION FOR BIOMETRIC APPLICATIONS.
DOI: 10.5220/0003462500400046
In *Proceedings of the International Conference on Signal Processing and Multimedia Applications* (SIGMAP-2011), pages 40-46
ISBN: 978-989-8425-72-0

2 LITERATURE REVIEW

Segmentation problem has been cope with from different mathematical approaches with a wide number of different applications (Kang et al., 2009; Shirakawa and Nagao, 2009).

Concretely, one approach which has experienced a great development in recent years is based on multiscale aggregation (Sharon et al., 2006). This procedure is based on processing the image according to a set of mathematical operations, so that pixels with similar properties are gathered in a same segment. The main characteristic of this approach relies on repeating this procedure through subsequent scales, in which the number of pixels is reduced for each scale, due to former aggregation (Sharon et al., 2006). Moreover, recent results obtained by these algorithms have shown improvements when compared to other methods (Alpert et al., 2007), like the Normalized Cuts (Shi and Malik, 2000) and Mean-Shift method (Comaniciu et al., 2002).

Multiscale aggregation methods gather a wide range of algorithms involving different mathematical operations applied to pixels in image.

In fact, several approaches have been proposed based on Segmentation by Weighted Aggregation (SWA (Sharon et al., 2006)), providing accurate results by means of similarities between intensities of neighboring pixels (Sharon et al., 2000), measurements of texture differences and boundary integrity (Sharon et al., 2001), more complicated operations, such as Gradient Orientations Histograms (GOH (Rory Tait Neilson and McDonald, 2007)), or more straightforward grouping methods based on the intensity contrast between two segments boundary and each segment inner (Felzenszwalb and Huttenlocher, 2004).

3 GAUSSIAN MULTISCALE AGGREGATION

Let define an image I as a graph $G = (V, E, W)$ were V represents nodes in graph corresponding to pixels in image, E stands for the edges connecting pairs of previous nodes V and W the weight of previous edges, measuring the similarity between two nodes (pixels) in V.

The idea is to divide graph G into two subgraphs $G = G_h \cup G_b$, so that subgraph G_h contains pixels corresponding to hand and the subgraph G_b gathers pixels corresponding to background. In addition, nodes in V contains two parameters: intensity (represented by μ) and deviation (represented by σ). This intensity corresponds in the first scale to the intensity in terms of grayscale images (Gonzalez and Woods, 1992), and to average intensities in subsequent scales. However, despite of existing deviation intensity value in subsequent scales, this parameter lacks of sense within first scale. The deviation in first scale will be set based on their neighbourhood of each pixel, which is a 4-neighbourhood structure for the first scale. These two parameters are gathered into a single function $\phi_{v_i}^{[s]}(\mu_{v_i}^{[s]}, \sigma_{v_i}^{[s]})$ representing the degree of being similar to node v_i, where s represents the scale. For simplicity sake, $\phi_{v_i}^{[s]}(\mu_{v_i}^{[s]}, \sigma_{v_i}^{[s]}) = \phi_{v_i}^{[s]}$, to avoid and excessive complicated notation.

Thus, the weight between two neighbour nodes v_i and v_j at scale s is defined as in Equation 1.

$$w_{ij}^{[s]} = \int \phi_{v_i}^{[s]} \phi_{v_j}^{[s]} d\zeta \tag{1}$$

where ζ makes reference to the complete color space. The more similarity between membership functions, the higher weight $w_{ij}^{[s]}$. Moreover, functions $\phi_{v_i^{[s]}}$ are normalized by definition so that $\int \phi_{v_i}^{[s]} d\zeta = 1$, for every scale s. Notice that w_{ij} is only calculated for neighbour pixels, according to the neighbourhood provided by each scale s.

The algorithm firstly sorts pairs of nodes in V based on their weights W, grouping these pairs under the same subgraph in case at least one has no previous subgraph already assigned ($\nexists\, i, j, G_k^{[s]}, G_m^{[s]} / v_i \in G_k \vee v_j \in G_m,\ \forall i, j, k, m, s$), and the dispersion of the possible subgraph is within a certain bound, given by the following relation presented in Equation 2:

$$\sigma_{i,j}^{[s+1]} \leq \sqrt{\sigma_i^{[s]} \sigma_j^{[s]}} \tag{2}$$

where i and j represent either subgraphs or nodes. In other words, the relation in Equation 2 states that a subgraph can gather new elements provided that uniformity within subgraph is bounded and could not get disperse. Same case happens when there are only two nodes to be gathered.

This method iterates along all weights sorted in W so that every node is associated with a subgraph. Next step consists on extracting the new membership functions for each subgraph, based on the functions associated with the nodes within such a subgraph.

For a given subgraph in the subsequent scale, $G_k^{[s+1]}$, the membership function is defined as follows in Equation 3.

$$\phi_{G_k}^{[s+1]} = \frac{\bigcup_j^N \phi_{G_j}^{[s]}}{\int \bigcup_j^N \phi_{G_j}^{[s]} d\zeta} \tag{3}$$

where N represents the number of nodes gathered by subgraph $G_k^{[s+1]}$. Notice $\phi_{G_k}^{[s+1]}$ is normalized according to definition, so that $\int \phi_{G_k}^{[s+1]} d\zeta = 1$.

However, the initial structure of 4-neighbourhood grid is lost with this aggregation, and therefore a new structure must be provided efficiently to these scattered nodes. In addition to function $\phi_{v_j}^{[s]}$, every node v_j is provided with their location within image I in terms of vertical and horizontal cartesian coordinates. When obtaining $G_k^{[s+1]}$, the centroid of those gathered nodes is calculated, so that each subgraph on subsequent scales have a position within image. This centroid, ξ, allows to provide a structure in successive scales by means of Delaunay triangularization (de Berg et al., 2008).

This operation represents the final step in the loop, since at this moment, there exist a new subgraph $G^{[s+1]} = \bigcup_k G_k^{[s+1]}$ at scale $s+1$ where each $G_k^{[s+1]}$ represents a node, and edges $E^{[s+1]}$ are provided by Delaunay triangulation, and weights $W^{[s+1]}$ are obtained based on Equations 1 and 3.

The whole loop is repeated until only two subgraphs remain, as stated at the begining of this section ($G = G_h \cup G_b$). However, due to the constraints provided to aggregate, the method could not aggregate more segments, without achieving the goal of dividing image into two subgraphs. Therefore, Equation 2 is in practice relaxed and stated as follows in Equation 4:

$$\sigma_{i,j}^{[s+1]} \leq \sqrt{\sigma_i^{[s]}\sigma_j^{[s]}} + k^{[s]} \tag{4}$$

being $k^{[s]}$ a factor able to avoid aggregation method from being stuck in the loop. This factor is dinamically increased, according to previous method necessities. However, initial value is set to $k^{[s]} = 0.1$, for each scale s.

The computational cost of this algorithm is quasi-linear with the number of pixels, since each scale gathers nodes in the sense that nodes in subsequent scales are reduced by (in practice) a three times factor. Therefore, time to process the first scale (which contains the highest number of nodes) is greater than the rest of times to process subsequent scales, and the total time is comparable to two times the processing time to aggregate first scale.

Finally, image I is based on a transform from RGB space to CIELAB (CIE 1976 L*,a*,b*) due to its ability to describe all visible colors by the human eye (Gonzalez and Woods, 1992; Tan et al., 2009; Mojsilovic et al., 2002). More specifically, layer a is considered as image I.

4 DATABASE

With the aim of evaluating the segmentation method a synthetic database has been created, gathering different hand positions, rotation degrees and environments, being possible to assess to what extent the segmentation algorithm can satisfactory perform a hand isolation from background on real scenarios.

Many different backgrounds are considered so that all possible scenarios are selected, containing textures from carpets, fabric, glass, grass, mud, different objects, paper, parquet, pavement, plastic, skin and fur, sky, soil, stones, tiles, tree, wall and wood. In addition, five different samples from every texture were collected to provide a more realistic evaluation scenario.

Initially, hands were taken with a blue-coloured background, so that hand can be easily extracted, being this prior segmentation result considered as ground-truth for posterior segmentation evaluation. This database contains a total of 120 individuals, with their both hands and 20 acquisitions per hand. Some acquisition examples of this database can be seen in Figure 1.

Figure 1: Samples of first database, with blue-coloured background. Synthetic database is based on this database, considering different backgrounds. In addition, the segmentation result of this database, will be considered as ground truth in a posterior evaluation.

Hand is then isolated and superposed to former backgrounds, carrying out an opening morphological operation (with a disk structural element of radius 5) for colour images (Gonzalez and Woods, 1992) to avoid possible edges separating hand and underlying texture, ensuring a more realistic image.

For each image, a total of 5×17 (five images and 17 textures) images are created. Therefore, second database collects a total of $120 \times 2 \times 20 \times 5 \times 17 = 408000$ images (120 individuals, two hands, 20 acquisitions per hand, five images and 17 textures) to properly evaluate segmentation on real scenarios. A visual example of this database is provided in Figure 2.

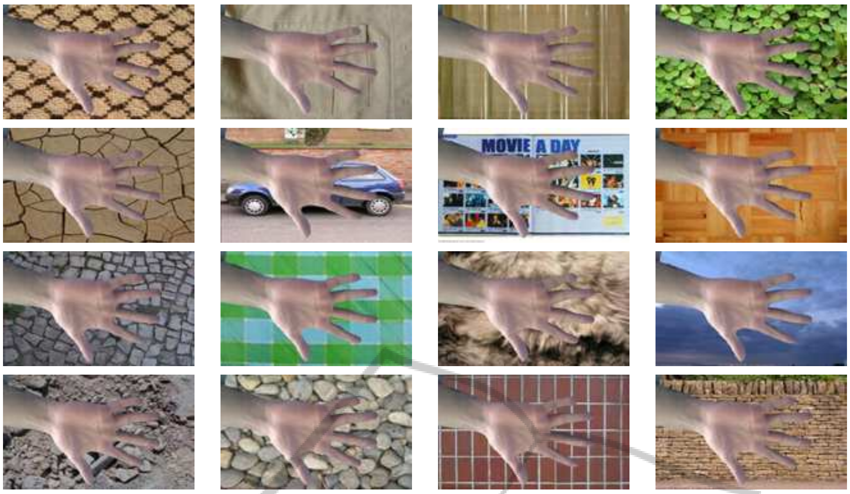

Figure 2: Samples from the synthetic database in different backgrounds for a given acquisition taken from the first database.

5 EVALUATION CRITERIA

The proposed segmentation algorithm must be evaluated according to criteria able to assess to what extent the algorithm is able to isolate hand from background. There exist several methods to evaluate segmentation in literature (Chen et al., 2010; Unnikrishnan et al., 2007; Meilă, 2005), but most of them consider several manual/human segmentations carried out by different individuals.

The presented evaluation criteria is based on a ground-truth segmentation, automatically obtained, but on the contrary, very reliable since the background is very easily distinguishable from hand (Figure 1). Therefore, the proposed method is based on F-Factor, (Alpert et al., 2007), defined as follows:

$$F = \frac{2RP}{R+P} \tag{5}$$

where P (Precision, Confidence) stands for the number of true positives (true segmentation, i.e. classify a hand pixel as hand) in relation to the number of true positives and false negatives (false hand segmentation), and R (Recall, Sensitivity) represents the number of true positives in relation to the number of true positives and false positives (false background segmentation, i.e. consider background as hand). F-Factor is within $[0,1]$ interval, so that 0 states a bad segmentation, while on the contrary 1 represents the best segmentation result.

In addition, a very important aspect of segmentation algorithm regards the required time to perform the aim of isolating objects on an image. This time depends strongly on the image size, the computer where experiments take place and the implementation, among other characteristics.

These former criteria will permit to assess to what extent the proposed algorithm meet their goals in an adequate time.

6 RESULTS

Under this section, results are presented according to evaluation criteria presented in previous Section 5.

First of all, segmentation is evaluated in terms of performance, considering F-Factor (Equation 5) as the main criterion. The obtained results are summerized in Table 1.

Reader can notice that those environments where hand could be camouflaged (like mud, soil, parquet, wood, ...) slightly decrease the performance of the algorithm. In addition, visual examples of segmentation results with different backgrounds and hands are provided in Figure 3.

The temporal performance for images of 640×340 pixels is 18 seconds, in a MATLAB implementation to be run in a PC computer @2.4 GHz Intel Core 2 Duo with 4GB 1067 MHz DDR3 of memory. A more refined implementation remains as future work. Nonetheless, this temporal result is very competitive if compared to approaches in literature, (Chen et al., 2010; Alpert et al., 2007).

7 CONCLUSIONS

This paper has presented an approach for hand bio-

Table 1: Segmentation evaluation by means of factor F in a synthetic database with 17 different background textures.

Texture	$F(\%)$	Texture	$F(\%)$	Texture	$F(\%)$
Carpets	92.3±0.2	Paper	91.3±0.2	Stones	91.4±0.1
Fabric	89.1±0.1	Parquet	88.4±0.3	Tiles	90.3±0.2
Glass	94.3±0.1	Pavement	89.1±0.2	Tree	96.3±0.2
Grass	93.7±0.1	Skin and Fur	95.7±0.2	Wall	94.2±0.1
Mud	89.8±0.1	Sky	96.4±0.2	Wood	93.8±0.1
Objects	92.1±0.2	Soil	89.4±0.1		

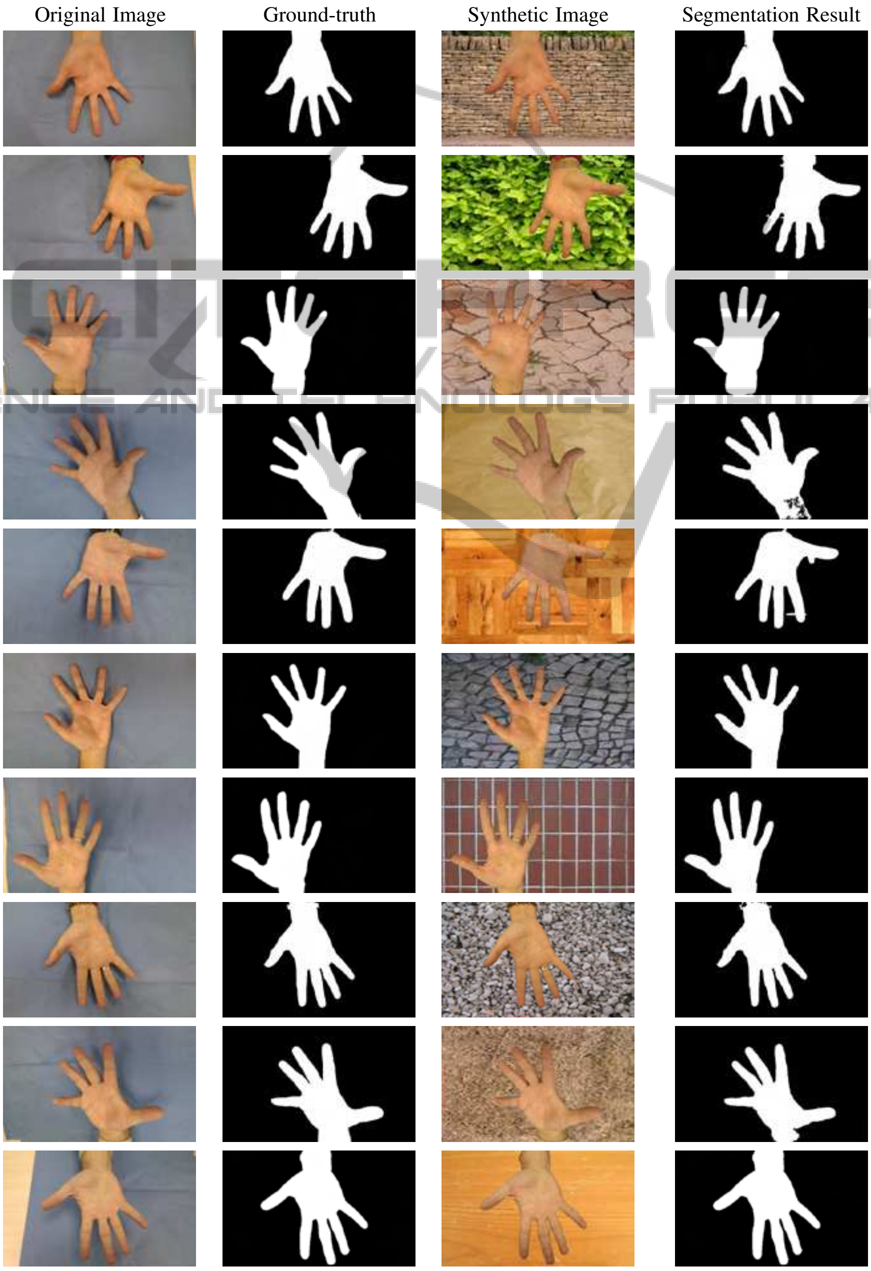

Figure 3: A comparative study of results provided by segmentation algorithm in comparison to ground-truth. First column gathers examples from first database, together with their segmentation on second column, considered as ground truth. Third column presents synthetic images based on first column images, providing on the fourth column the final segmentation result.

metric segmentation based on gaussian multiscale aggregation. This method is able to isolate hand from background in different situations, simulated by an own synthetic public database, with a total of 408000 images.

The results highlight the fact that hand is isolated with a competitive accuracy, providing a good result for a posterior feature extraction, independently on the background of the hand image.

Applications of this method are very suitable for mobile applications, since hand mobile biometrics must be able to identify individuals everywhere, without no constrains on the background. However, more efforts must be done to adapt this approach for mobile biometrics, since its temporal performance is far at present from being adequate for real-time applications. In addition, the time performance is still low (18 seconds), when compared to other similar approaches in literature, and considering the challenging backgrounds to segment.

Future work regards an improvement and refinement in implementation, together with a mobile orientation, so that mobile hand biometrics could benefit of a reliable segmentation algorithm, and therefore, increase their identification accuracy.

ACKNOWLEDGEMENTS

This research has been supported by the Ministry of Industry, Tourism and Trade of Spain, in the framework of the project CENIT-Segur@, reference CENIT-2007 2004.

REFERENCES

Alpert, S., Galun, M., Basri, R., and Brandt, A. (2007). Image segmentation by probabilistic bottom-up aggregation and cue integration. In *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR '07.*, pages 1–8.

Chen, S., Cao, L., Wang, Y., Liu, J., and Tang, X. (2010). Image segmentation by map-ml estimations. *Image Processing, IEEE Transactions on*, 19(9):2254 –2264.

Comaniciu, D., Meer, P., and Member, S. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:603–619.

de Berg, M., van Kreveld, M., Overmars, M., and Schwarzkopf, O. (2008). *Computational Geometry: Algorithms and Applications*. Springer, 3rd edition.

Felzenszwalb, P. F. and Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. *Int. J. Comput. Vision*, 59:167–181.

García-Casarrubios Muñoz, A., de Santos-Sierra, A., Sánchez-Ávila, C., Guerra-Casanova, J., Bailador-del Pozo, G., and Jara-Vera, V. (2010). Hand biometric segmentation by means of fuzzy multiscale aggregation for mobile devices. In *Emerging Techniques and Challenges for Hand-Based Biometrics (ETCHB), 2010 International Workshop on*, pages 1 –6.

Gonzalez, R. C. and Woods, R. E. (1992). *Digital Image Processing*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

Kang, W.-X., Yang, Q.-Q., and Liang, R.-P. (2009). The comparative research on image segmentation algorithms. In *ETCS '09: Proceedings of the 2009 First International Workshop on Education Technology and Computer Science*, pages 703–707, Washington, DC, USA. IEEE Computer Society.

Kukula, E. and Elliott, S. (2005). Implementation of hand geometry at purdue university's recreational center: an analysis of user perspectives and system performance. In *Security Technology, 2005. CCST '05. 39th Annual 2005 International Carnahan Conference on*, pages 83–88.

Kukula, E. and Elliott, S. (2006). Implementation of hand geometry: an analysis of user perspectives and system performance. *Aerospace and Electronic Systems Magazine, IEEE*, 21(3):3–9.

Meilă, M. (2005). Comparing clusterings: an axiomatic view. In *Proceedings of the 22nd international conference on Machine learning*, ICML '05, pages 577–584, New York, NY, USA. ACM.

Mojsilovic, A., Hu, H., and Soljanin, E. (2002). Extraction of perceptually important colors and similarity measurement for image matching, retrieval and analysis. *Image Processing, IEEE Transactions on*, 11(11):1238 – 1248.

Rory Tait Neilson, B. N. and McDonald, S. (2007). Image segmentation by weighted aggregation with gradient orientation histograms. *Southern African Telecommunication Networks and Applications Conference (SATNAC)*.

Sharon, E., Brandt, A., and Basri, R. (2000). Fast multiscale image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition, 2000. Proceedings.*, volume 1, pages 70 –77 vol.1.

Sharon, E., Brandt, A., and Basri, R. (2001). Segmentation and boundary detection using multiscale intensity measurements. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001.*, volume 1, pages I–469 – I–476 vol.1.

Sharon, E., Galun, M., Sharon, D., Basri, R., and Brandt, A. (2006). Hierarchy and adaptivity in segmenting visual scenes. Macmillan Publishing Ltd.

Shi, J. and Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:888–905.

Shirakawa, S. and Nagao, T. (2009). Evolutionary image segmentation based on multiobjective clustering. In

CEC'09: Proceedings of the Eleventh conference on Congress on Evolutionary Computation, pages 2466–2473, Piscataway, NJ, USA. IEEE Press.

Tan, W., Wu, C., Zhao, S., and Chen, S. (2009). Hand extraction using geometric moments based on active skin color model. In *Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE International Conference on*, volume 4, pages 468–471.

Unnikrishnan, R., Pantofaru, C., and Hebert, M. (2007). Toward objective evaluation of image segmentation algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29:929–944.