

Gaussian Multiscale Aggregation oriented to Hand Biometric Segmentation in Mobile Devices

Alberto de Santos Sierra
Carmen Sánchez Ávila

*Group of Biometrics, Biosignals and Security, GB2S
Centro de Domótica Integral
Polytechnical University of Madrid
28223, Pozuelo de Alarcón, Madrid
{alberto,csa}@cedint.upm.es*

Gonzalo Bailador del Pozo
Javier Guerra Casanova

*Group of Biometrics, Biosignals and Security, GB2S
Centro de Domótica Integral
Polytechnical University of Madrid
28223, Pozuelo de Alarcón, Madrid
{gbailador,jguerra}@cedint.upm.es*

Abstract—New trends in biometrics are oriented to mobile devices in order to increase the overall security in daily actions like bank account access, e-commerce or even document protection within the mobile.

However, applying biometrics to mobile devices imply challenging aspects in biometric data acquisition, feature extraction or private data storage. Concretely, this paper attempts to deal with the problem of hand segmentation given a picture of the hand in an unknown background, requiring an accurate result in terms of hand isolation.

For the sake of user acceptability, no restrictions are done on background, and therefore, hand images can be taken without any constraint, resulting segmentation in an exigent task.

Multiscale aggregation strategies are proposed in order to solve this problem due to their accurate results in unconstrained and complicated scenarios, together with their properties in time performance. This method is evaluated with a public synthetic database with 480000 images considering different backgrounds and illumination environments.

The results obtained in terms of accuracy and time performance highlight their capability of being a suitable solution for the problem of hand segmentation in contact-less environments, outperforming competitive methods in literature like Lossy Data Compression image segmentation (LDC).

Keywords—Multiscale Aggregation, Hand Biometrics, Mobile Devices, Security, Biometrics, Graphs

I. INTRODUCTION

Hand geometry biometrics can provide mobile devices with an increase in security in daily applications like mobile secure access, data protection or electronic commerce and bank operations. The characteristics of this biometric technique in terms of non-invasiveness and acceptability highlight the fact that hand biometrics could be a proper and adequate biometric method for verification and identification in mobile devices.

However, hand biometrics oriented to mobile devices imply several challenges which require additional efforts, being hand segmentation the most significant difficult task, provided that hands are acquired contact-less with an embedded camera. Previous approaches barely consider hand

segmentation, since backgrounds were extremely controlled, providing an easy and straightforward hand isolation.

In fact, the segmentation step is essential in hand biometrics, given that a subsequent feature extraction depends enormously on an accurate and precise hand isolation, otherwise template features could be inappropriately extracted, resulting in a reduction in individual identification.

The proposed segmentation method is based on multiscale aggregation, gathering pixels along scales according to a given similarity gaussian function. This method produces an iterative clustering aggregation, providing a solution for hand image segmentation with a quasi-linear computational cost and an adequate accuracy for biometric applications.

The method has been tested with a synthetic image database, with around 408000 images considering different backgrounds and illumination environments, and compared to a competitive approach in literature in terms of image segmentation. This approach is named Lossy Data Compression (LDC), [18].

Finally, the structure of the paper is as follows: Section II provides an overview on the current literature, describing the proposed method under Section III. Results will be presented in Section IV, providing conclusions and future work in Section V.

II. STATE OF THE ART

Several approaches have been proposed to solve the problem of hand segmentation. As an overview, segmentation methods are more complicated, as the background increases in difficulty, in terms of texture or illumination.

In early works, hand segmentation received little attention, given the fact that isolating hand from background, given a monochromatic or apriori known background was straightforward [1], [2].

However, more challenging problems in segmentation appeared when hand biometrics required no constraint on background, presenting new trends in contact-less biometrics [3], [4], [5], [6].

An adequate solution for contact-less hand biometrics avoiding hard and time-consuming segmentation algorithms is based on infrared camera acquisition [4], [5], where the acquisition procedure facilitates a posterior hand extraction, given the fact that infrared illumination only highlights those objects closer to the sensor. In other words, infrared illumination avoids background acquisition.

At present, hand biometrics are oriented to mobile applications in order to provide more security to mobile devices [7], [8], with the single requirement of using the embedded camera within the mobilephone, involving no additional hardware. Therefore, more complicated algorithms are proposed given both the complexity of the segmentation aim and that mobile devices are increasing their capability to carry out complex operation, together with the fact that increasing acceptability by final user involves less restrictions in backgrounds.

The selection of algorithms based on multi-scale aggregation strategies [9], [10], [11], [12] is then justified, since these methods allow fast and precise segmentation, requiring no supervision in learning and providing, therefore, a segmentation result independent from background. This family of algorithms has been extensively used in image segmentation with important and remarkable results in terms of accuracy and memory efficiency [13].

III. METHODOLOGY: GAUSSIAN MULTISCALE AGGREGATION

Every hand acquisition will be represented by an image I , which can be defined as a graph $G = (V, E, W)$, considering V as the set of nodes in graph corresponding to pixels in image I , E as the edges connecting pairs of nodes in V and W the weight associated to previous edges, denoting the similarity between a pair of nodes in V .

The idea is to divide graph G into two subgraphs: one containing pixels corresponding to hand, associating the other subgraph with pixels related to background. In addition, nodes in V contain three important parameters: intensity (represented by μ), deviation (represented by σ) and position (represented by π). This intensity corresponds in the first scale to the intensity in terms of grayscale images [14], 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. Finally, position π describes the location of the corresponding node within image I , being of importance for aggregation and neighbourhood definition in subsequent scales.

Concerning the former two parameters, they 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, $\phi_{v_i}^{[s]}(\mu_{v_i}^{[s]}, \sigma_{v_i}^{[s]}) = \phi_{v_i}^{[s]}$.

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

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

where ζ makes reference to the complete color space. This color space is obtained by transforming RGB space into CIELAB (CIE 1976 L*,a*,b*) due to its ability to describe all visible colors by the human eye [14], [15], [16].

The more similarity between nodes, 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 according to their weights W , clustering these pairs under the same subgraph in case at least one of both has no previous subgraph already assigned. In other words, a pair of nodes are assigned under the same subgraph if $\nexists i, j, G_k^{[s]}, G_m^{[s]} / v_i \in G_k \vee v_j \in G_m, \forall i, j, k, m, s$, together with the condition that the dispersion of the possible subgraph is within a certain bound, Equation 2:

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

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 forcing that uniformity within subgraph is bounded and could not get disperse.

This method iterates along all weights sorted in W so that every node is associated with a subgraph. Next step consists of 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.

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

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 procedure, and therefore a new structure must be provided efficiently to these scattered nodes. When obtaining $G_k^{[s+1]}$, the position of those gathered nodes is calculated ($\pi_k^{[s+1]}$), 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 [17].

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 beginning of this section. However, due to the constraints provided to aggregate (Equation 2), it could be possible that 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:

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

being $k^{[s]}$ a factor able to avoid aggregation method from being stuck in the loop. This factor can be dynamically increased or decreased, according to previous method necessities. However, initial value is set to $k^{[s]} = 0.01$, for each scale s . The capability of $k^{[s]}$ to adapt the necessities of the algorithm remains as future work.

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 [9]. 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 approximately comparable to two times the processing time to aggregate first scale.

IV. RESULTS

This section provides both the description of the database involved for a proper testing of the proposed algorithm, together with an explanation on the evaluation scenario and mechanisms to assess the accuracy of the algorithm. In addition, the algorithm performance is compared to Lossy Data Compression segmentation (LDC), a competitive existing segmentation approach in the state of the art [18]. This comparison is carried out in terms of temporal performance and segmentation accuracy.

A. Database Collection

Since the segmentation algorithm must isolate completely hand from background, provided no prior knowledge on such background, the collected database must consider a wide range of different scenarios where hands can be taken with a mobile device.

Therefore, a total of 17 different textures were proposed to create the database, containing samples (specifically 5 different samples per texture) of scenarios like carpets, fabric, glass, grass, mud, different objects, paper, parquet, pavement, plastic, skin and fur, sky, soil, stones, tiles, tree, wall and wood. These textures were obtained from the website <http://mayang.com/textures/>.

Hands were obtained from a database (called HTC database, since it was collected with a HTC mobile) already collected of 120 individuals of both hands with 20

acquisitions per hand. This database was carried out with a controlled background, so that segmentation can be easily performed, and it is publicly available at <http://www.gb2s.es>.

The creation of the synthetic database (named synData) considers the hands extracted in former database and the set of the aforementioned different textures. First of all, a straightforward segmentation was carried out with a colour-based segmentation [6], obtaining two binary masks: M_h , corresponding to those pixels representing hand, and M_b with pixels corresponding to background.

Afterwards, both masks are laid one over each other, with M_b containing pixels associated to a specific texture, resulting in an image with the hand over a desired background (grass, water, wood and so forth).

In order to ensure there is no considerable differences in illumination between hand and background, each image is converted from RGB to YCbCr color space [14] carrying out a histogram equalization in terms of illumination (Y), performing afterwards the inverse transform from YCbCr to RGB color space. Finally, a morphological operation consisting on an opening operator with a structural element of a disk of small size (5 pixels radio) is considered to fade the boundary between hand and background, so that hand is integrated within background.

All these former operations attempt to ensure a fair scenario, simulating the conditions provided in real situations.

Lastly, this operation was carried out for each image in the HTC database and considering all possible textures. Then, the synthetic database comes up with a total of $2 \times 20 \times 120 \times 5 \times 17 = 408000$ images, also publicly available at <http://www.gb2s.es>.

B. Evaluation and Results

The evaluation of the proposed segmentation algorithm is based on a supervised ground-truth evaluation, considering as ground-truth the segmentation result provided by the colour-based segmentation applied to HTC database (120 users, 2 hands, 20 samples per hand).

This ground-truth is easy to obtain and reliable in terms of performance due to the distinctive characteristics of the hand texture in comparison to the blue background provided on that database.

Thus, the segmentation obtained by the proposed method in the synthetic database will be compared to the result extracted from HTC database.

Despite of existing several methods to evaluate segmentation in literature, provided a ground-truth, a suitable measure is the F-measure, because of its attributes and properties in terms of sensitivity and specificity [10].

The selected F-measure is defined as the harmonic mean between Precision (P) and Recall (R) as follows in Equation 5:

$$F = \frac{2RP}{R+P} \quad (5)$$

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-measure is within $[0, 1]$ interval, so that 0 states a bad segmentation, while on the contrary 1 represents the best segmentation result.

Based on this criteria, segmentation is evaluated in terms of performance, considering F-measure (Equation 5) as the main criterion. In addition, the results are compared to those obtained with an existing approach in literature, namely Lossy Data Compression (LDC) [18], very competitive in terms of image segmentation.

The obtained results are summarized in Table I.

Reader may notice that those environments where hand could be camouflaged (like mud, soil, parquet or wood) slightly decrease the performance of the algorithm. Moreover, the inclusion of illumination correction between hand and background hardly affect the results of segmentation when compared to not using such an illumination correction [14].

Furthermore, Table I presents the results provided by LDC approach [18], so that the performance of the proposed algorithm can be fairly compared to an existing approach. Table I highlights the fact that the results obtained by the proposed method overcomes the performance of LDC algorithm in terms of F-measure.

In addition, visual segmentation results with different backgrounds and hands are provided in Figure 1. Despite of being the images horizontally oriented, it must be stated that hands in both databases (HTC and synData) present different degrees of rotation. The fact of presenting images horizontally regards a better use of the paper space.

Concerning time performance, the algorithm requires 18 seconds to segment images of 640×340 pixels, in a MATLAB implementation running in a PC computer 2.4 GHz Intel Core 2 Duo with 4GB 1067 MHz DDR3 of memory.

In addition, Table II gathers the time results obtained for different image dimensions, for both the proposed method and the approach used for comparison, Lossy Data Compression (LDC). Reader may notice the considerable difference in time performance between approaches.

A more refined implementation remains as future work. Nonetheless, both temporal result and their memory usage are very competitive if compared to approaches in literature, [19], where approaches in literature could take several minutes with previous images sizes and present some problems of memory management [13], [19].

V. CONCLUSIONS

This paper has proposed a segmentation algorithm based on multiscale aggregation oriented to hand segmentation with the aim of contact-less and non-invasive biometrics. The method aggregates pixels according to gaussian accumulative functions, describing the similarity between pixels.

The results highlight the fact that hand is isolated with a competitive accuracy (F-measure $\geq 85\%$), providing a good result for a posterior feature extraction, independently on the background of the hand image, notoriously improving current approaches in the state of the art.

The assessment of the method has been carried out with a supervised evaluation approach comparing the obtained result with a ground-truth segmentation. In addition, this paper comes up with a synthetic database of 408000 hand images in different positions, orientation and with a total of 17 varying scenarios and backgrounds like mood, grass, carpets or wood.

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.

The results are very competitive when compared to existing and competitive approaches in literatures like Lossy Data Compression (LDC) [18], not only in terms of time performance, but also in accuracy and image segmentation.

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.

REFERENCES

- [1] Y. P. Lew, A. R. Ramli, S. Y. Koay, R. Ali, and V. Prakash, "A hand segmentation scheme using clustering technique in homogeneous background," in *Research and Development, 2002. SCORED 2002. Student Conference on*, 2002, pp. 305–308, iD: 1.
- [2] R. Sanchez-Reillo, C. Sanchez-Avila, and A. Gonzalez-Marcos, "Biometric identification through hand geometry measurements," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 10, pp. 1168–1171, oct 2000.
- [3] M. Arif, T. Brouard, and N. Vincent, "Personal identification and verification by hand recognition," in *Engineering of Intelligent Systems, 2006 IEEE International Conference on*, 0-0 2006, pp. 1–6.

Table I: Segmentation evaluation by means of measure F in database synData with 17 different background textures, together with the corresponding standard deviation. In addition, the results for LDC is also provided for comparison.

Texture	Proposed, $F(\%)$	LDC, $F(\%)$	Texture	Proposed, $F(\%)$	LDC, $F(\%)$	Texture	Proposed, $F(\%)$	LDC, $F(\%)$
Fabric	88.4±0.3	65.3±0.1	Parquet	88.3±0.2	66.1±0.2	Tiles	90.1±0.2	71.5±0.3
Glass	94.1±0.1	75.8±0.1	Pavement	88.9±0.2	67.8±0.1	Tree	96.0±0.2	73.4±0.2
Grass	93.3±0.2	70.1±0.1	Skin and Fur	95.3±0.3	82.3±0.2	Wall	94.1±0.1	70.9±0.2
Mud	89.5±0.2	68.3±0.1	Sky	96.1±0.1	77.2±0.2	Wood	93.5±0.1	82.5±0.2
Objects	92.0±0.1	70.1±0.1	Soil	89.0±0.2	67.2±0.3			

Table II: Relation between time performance (in seconds), the dimension of the image, and the size in number of pixels, comparing both the proposed method and LDC approach.

Image Dimensions	Number of Pixels	Proposed (seconds)	LDC (seconds)
450x600	270000	19.8	63.4
300x400	120000	9.4	52.1
150x200	30000	3.1	32.8

- [4] M. Ferrer, J. Fabregas, M. Faundez, J. Alonso, and C. Travieso, "Hand geometry identification system performance," in *Security Technology, 2009. 43rd Annual 2009 International Carnahan Conference on*, 5-8 2009, pp. 167–171.
- [5] A. Morales, M. Ferrer, J. Alonso, and C. Travieso, "Comparing infrared and visible illumination for contactless hand based biometric scheme," in *Security Technology, 2008. ICCST 2008. 42nd Annual IEEE International Carnahan Conference on*, 2008, pp. 191–197.
- [6] A. de Santos Sierra, J. Guerra Casanova, C. Sánchez Ávila, and V. Jara Vera, "Silhouette-based hand recognition on mobile devices," in *43rd Annual 2009 International Carnahan Conference on Security Technology, 2009.*, oct. 2009, pp. 160–166.
- [7] A.-C. Muñoz, A. de Santos Sierra, C. Ávila, J. Casanova, G. del Pozo, and V. Vera, "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*, 2010, pp. 1–6.
- [8] T. Alhussain, S. Drew, and O. Alfarraj, "Biometric authentication for mobile government security," in *Intelligent Computing and Intelligent Systems (ICIS), 2010 IEEE International Conference on*, vol. 2, 2010, pp. 114–118, iD: 1.
- [9] A. G.-C. Muñoz, C. S. Ávila, A. de Santos Sierra, and J. G. Casanova, "A mobile-oriented hand segmentation algorithm based on fuzzy multiscale aggregation," in *Proceedings of the 6th international conference on Advances in visual computing - Volume Part I*, ser. ISVC'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 479–488.
- [10] S. Alpert, M. Galun, R. Basri, and A. Brandt, "Image segmentation by probabilistic bottom-up aggregation and cue integration," in *IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR '07.*, june 2007, pp. 1–8.
- [11] E. Yoruk, E. Konukoglu, B. Sankur, and J. Darbon, "Shape-based hand recognition," *IEEE Transactions on Image Processing*, vol. 15, no. 7, pp. 1803–1815, july 2006.
- [12] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 888–905, 1997.
- [13] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation."
- [14] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1992.
- [15] W. Tan, C. Wu, S. Zhao, and S. Chen, "Hand extraction using geometric moments based on active skin color model," in *Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE International Conference on*, vol. 4, 20-22 2009, pp. 468–471.
- [16] A. Mojsilovic, H. Hu, and E. Soljanin, "Extraction of perceptually important colors and similarity measurement for image matching, retrieval and analysis," *Image Processing, IEEE Transactions on*, vol. 11, no. 11, pp. 1238 – 1248, Nov. 2002.
- [17] M. de Berg, M. van Kreveld, M. Overmars, and O. Schwarzkopf, *Computational Geometry: Algorithms and Applications*, 3rd ed. Springer, April 2008.
- [18] Y. Ma, H. Derksen, W. Hong, and J. Wright, "Segmentation of multivariate mixed data via lossy data coding and compression," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 9, pp. 1546 –1562, sept. 2007.
- [19] S. Chen, L. Cao, Y. Wang, J. Liu, and X. Tang, "Image segmentation by map-ml estimations," *Trans. Img. Proc.*, vol. 19, pp. 2254–2264, September 2010. [Online]. Available: <http://dx.doi.org/10.1109/TIP.2010.2047164>

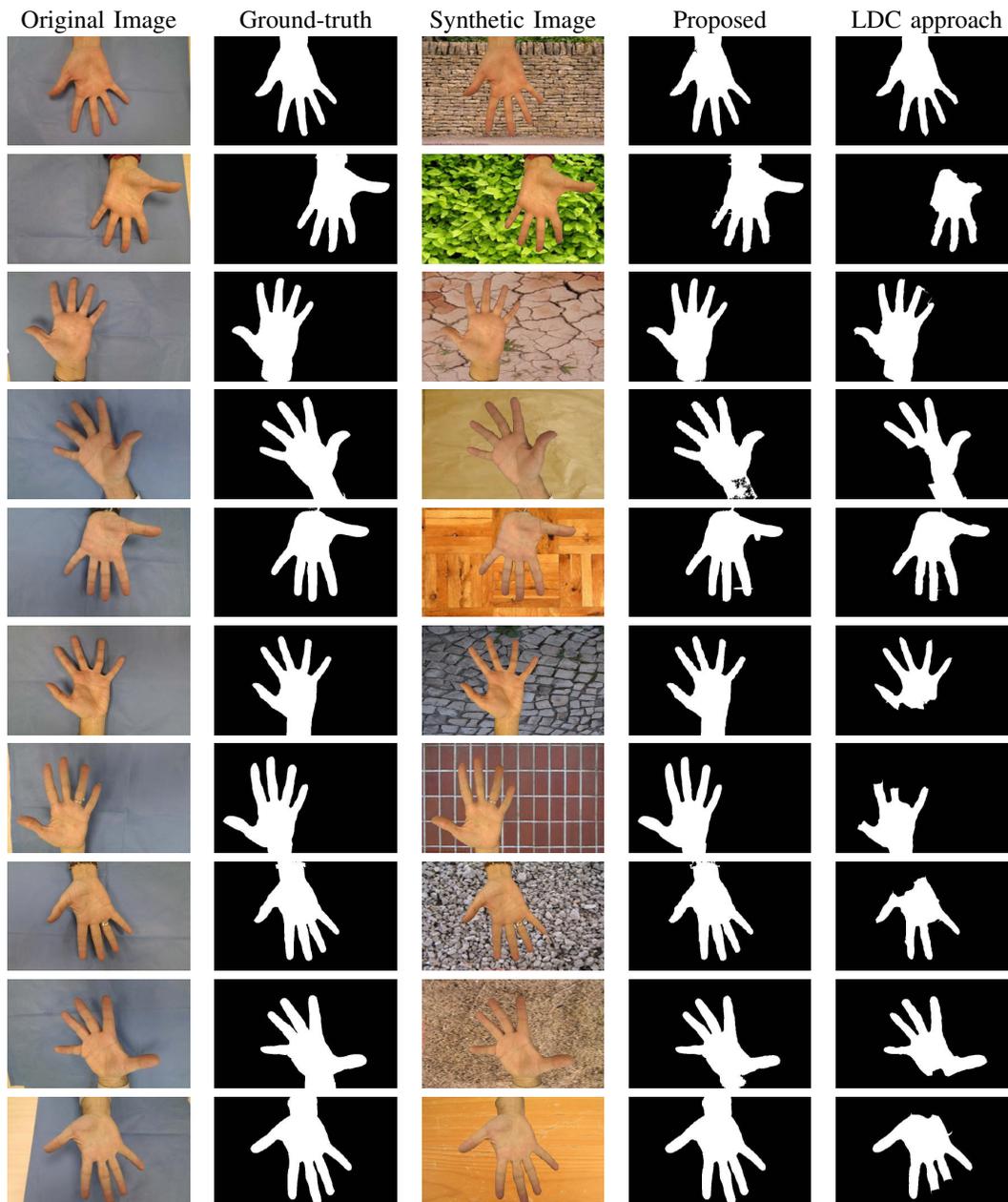


Figure 1: A comparative study of results provided by segmentation algorithm in comparison to ground-truth. First column gathers examples from first database, together with their segmentation on second column, considered as ground truth. Third column presents synthetic images based on first column images, providing on the fourth column the final segmentation result. Last column presents the segmentation result provided by the Lossy Data Compression (LDC) [18].