The Negative Selection Algorithm: a Supervised Learning Approach for Skin Detection and Classification.

Esma Bendiab 1† and Mohamed Kheireddine Kholladi 2††,

University of Mentouri, MISC Laboratory, Constantine, ALGERIA

Summary

Artificial immune systems (AIS) are relatively new class of meta-heuristics that mimics aspects of the human immune system to solve computational problems. They consist of three typical intelligent computational algorithms termed clonale selection algorithm, immune network theory and negative selection algorithm. The negative selection algorithm is a supervised learning algorithm based population. It has been successfully applied to change and anomaly detection. As AIS emerged, classification has been become an important application area of AIS. Classification systems that are based on AIS have attractive features inherited form biological immune system.

On the other hand, skin color detection is used as a preliminary step in numerous computer vision applications like: face detection, hand gesture detection, person identification and others. Two main issues of the skin detection are: the selection of the best skin features space and the skin pixel classification algorithm.

This paper describes initial framework of pixels skin color classification approach based on the negative selection algorithm from AIS. We present a skin classification, for detecting skin pixels and non skin pixels in color images, using both color and texture features space and the negative selection algorithm as a classifier. The proposed approach is able to detect skin regions from images taken from different imaging conditions. The results are promising and suggest a new approach for adapting human skin color using negative selection algorithm.

Key words:

Artificial Immune System, Negative Selection Algorithm, Classification, Supervised Learning, Skin Detection.

1. Introduction

The classification aims to identify classes and the objects belonging by using some descriptive features. It is applied to many human activities; in addition it is especially appropriate for automated decision making problems such: diagnostics aid, pattern recognition and others [1]. One of the approaches to solve this problem is the automatic extraction of the classification process from a set of examples. In short, an example is the description of a case with the corresponding classification. This is called a learning system.

Methods that use learning systems are numerous. They can be classified into two methods: statistical methods, often known as classical methods, and artificial intelligence methods. Classical methods require a large data volume and several tools, which result complexity and low accuracy in their use. Hence, to get over these flaws, a remarkable convergence to artificial intelligence methods has emerged. So, researchers have shifted their interests towards the natural inspiration to solve complex problems. Hence, significant attention has been given to artificial immune system.

Artificial immune systems are relatively new class of meta-heuristics that mimics aspects of the human immune system to solve computational problems [2-4]. They are massively distributed and parallel, highly adaptive and reactive and evolutionary where learning is native. Growing interests are surrounding those systems due to the fact that natural mechanisms such as: recognition, identification, and intruders' elimination, which allow the human body to reach its immunity; suggest new ideas for computational problems. Artificial immune systems consist of three typical intelligent computational algorithms termed immune network theory, clone selection and negative selection [2][5].

Artificial negative selection is a computational imitation of self non self discrimination. It was introduced by Forrest and all [6] as a supervised learning algorithm based population in order to classify a bit or string representations of real-world data, termed antigen, as normal or anomalous. It has a successful outcome to computer security, network security and anomalies detection problems.

Otherwise, the recognition of human skin can be of great interest for face recognition, hand gestures recognition, facial expressions analysis, searching and filtering image contents on the web, and so on. Skin color detection is a challenging task as the skin color in an image is sensitive to various factors like: illumination, camera characteristics, ethnicity, and individual characteristics such as: age, sex and body parts… [7-10]. All these factors affect skin color appearance. A significant overlap between the skin and non-skin pixels is another problem to be considered [8].For the above reasons combining the skin texture features with its color feature will increase the accuracy of skin classification [7,9].

Manuscript received November 5, 2010
Manuscript revised November 20, 2010

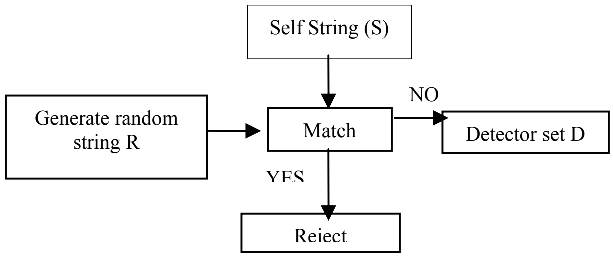
In this paper, artificial immune systems, exhibiting at better their learning aspect by the negative selection algorithm, are applied on two class classification problem. The classification is done to detect skin pixels and non skin pixels in color images, using both color and texture features space and the negative selection algorithm as a classifier.

The rest of the paper is organized as follows. Section 2 contains relevant background information and motivation regarding the negative selection algorithm. Section 3 describes the problem of the classification and learning. This is followed by the proposed approach in section 5. Section 6 includes experimentations and parameters analysis. The paper ends with a conclusion and future works.

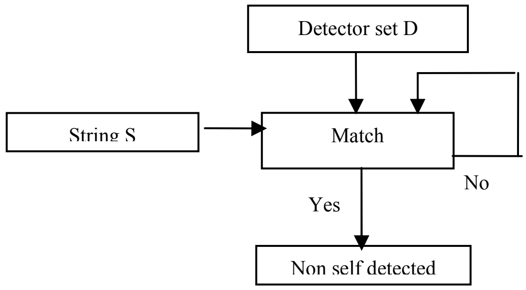
2. The Negative Selection Algorithm

The negative selection algorithm is a supervised learning algorithm introduced by Forrest and al [6] to computer security, network security and anomalies detection problems. It is based on the discriminatory mechanism of the natural system. The aim of the negative selection algorithm is to classify a bit or string representations of real-world data, termed antigen, as normal or anomalous. In nature, Antigen is anything which is not part of the body itself [6][11,12]. The algorithm processes in two steps: learning and testing.
The basic idea of the negative selection algorithm is to generate a number of detectors in the complementary space. Then, apply these detectors to classify new, unseen, data as self or non self.

The algorithm can be summarized in the following steps [6]:
• Define self as a set of S elements of length l over a finite alphabet: a collection that needs to be protected or monitored.
• Then generate a set of D detectors, which does not match any element in S. Instead of exact or perfect matching, the method uses a partial matching rule, in which two strings match if and only if they are identical at least at r contiguous positions, where r is a chosen parameter.
• Monitor S form changes by continually matching the detectors in D against S.
A schematically representation of the algorithm can be found in fig. 1 [6].
In the initial description of the algorithm, candidate detectors are generated randomly and then tested to see if they match any self string. If a match is found, the candidate is rejected .This process is repeated until a desired number of detectors is generated.



(a) Censoring phase



(b) Monitoring phase

Fig.1 The Negative Selection Algorithm [6].

The fundamental difference between the negative selection algorithm and another artificial intelligence paradigm, when applied to pattern recognition, is in the nature of the identification itself. In the negative selection algorithm, identification consists in finding several schemes that don't correspond to self in the population. Table 1 resumes AIS principles. We describe algorithm elements, their corresponding elements in real world problem and their corresponding elements in the natural immune system.

Table 1: Table resuming AIS principles.

Algorithm elements	***Real world problem***	***Natural immune system***
Detector	Data	Antibody
Self samples	Self set1, training data	Self Cells
Incoming data instance	New data sample, data item	Antigen
Distance measure	Affinity measure, similarity measure	Affinity measure in the shape space.
r-contiguous matching rule	r-contiguous bits rule, rcb matching rule	
Matching rule	Matching rule	

3. Classification and Learning

Machine learning refers to the development, testing and implementation of methods that allow a machine to evolve through a learning process [13], and so to fulfill tasks that are difficult or impossible to accomplish by classical methods. Learning algorithms are of three forms: supervised, semi supervised or unsupervised [14,15]. In the first form, we need a set of examples that is labeled by an expert.

In this case, the trainee must then find or approximate the function which assigns the correct label to these examples. This machine learning technique allows the automatic generation of rules from a training data database containing examples of previously treated cases [14].

More specifically, the training set is an input-output pairs (x,y) in the formula (1), which is drawn by an unknown law on . For example, x_n follows a uniform law and w_n is centered noise.

$$y_n = f(x_n) + w_n \quad (x_n, y_n)_{1\leq n\leq N} \qquad (1)$$
$$\text{with} \quad x_n \in X \quad \text{and} \quad y_n \in Y$$

The objective of supervised learning method is to use this training set to determine a compact representation of f noted g and called a prediction function, which associates to a new input x an output g (x). So, the purpose of a supervised learning algorithm is to generalize unknown data of what the trainee has already learnt by using data processed by experts.

In the case of a finite cardinal output values, we talk about a classification problem. Especially since, the main purpose of a classification problem is to assign a label to a given input.

In the same way, in the next section we state a supervised learning to a classification problem using the negative selection algorithm.

4. Problem Statement

An approach based on artificial immune system ought to describe two aspects:

1. The projection and the models advocation of immune elements in the real world problem.
2. The use of the appropriate immune algorithm or approach to solve the problem.

4.1 Immune Representation

Before describing the immune representation in the context of skin detection, we must depict the skin features space.

4.1.1 Skin Characteristics:

We consider two spaces to define the skin characteristics: color and texture.

1. Color Characteristics:

The color moment is a compact representation of the color features to characterize a color image. It was shown that most of the color distribution is captured by three moments' orders. The first order moment (μc) captures the average color, the second order (σc) captures the standard deviation, and third order moment captures the asymmetry color (θc). These three order moments (μc, σc, θc) are extracted from each of the three color planes (RGB), using the mathematical formula shown in the table 2.

2. Texture Characteristics:

Texture analysis is an active area of research in pattern recognition. A variety of techniques have been used to measure the similarity of texture. One approach proposes a co-occurrence matrix that uses a grayscale image. We can extract the texture features, such as: entropy, energy, contrast, and homogeneity from the co-occurrence matrix [16].

The gray level of co-occurrence matrix C (i, j) is specified by a displacement vector (where Dx, δy are the displacements in X and Y respectively), followed by counting all pixels pairs separated by the displacement vector, and having gray levels i and j. The matrix C (i, j) is normalized by dividing each element in it by the total number of pixels pairs [16]. To obtain the characteristics of the texture using the mathematical formula is shown in table 2.

4. 2 Immune Elements Modeling:

1. Antibodies:

Antibodies represent candidate solutions to the classification problem. They correspond to the set of skin characteristics. It is relived as the schematically representation shown in fig. 2.

2. Antigens:

In the artificial immune system, the antigen usually corresponds to the problem to be resolved and its constraints. In this context, the antigen ties on a pixels set to be classified on two classes: corresponding to the skin or non skin, selected randomly. It is presented as shown in fig.2.

3. The Shape space:

The shape-space is the antigen-antibody interaction. It allows a quantitative description of receptors identification to antigens [2].

The shape-space type, used to describe antigens and antibodies, will determine in part a measure to calculate their affinity. In this work, a real shape-space is considered. On account of both antigens and antibodies are real valued. They represent pixel area referring to texture and color features.

Table 2: Representation of color and texture features

	Features	Formula
Color moments	The first order moment (μc)	$\mu_c = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} p_{ij}^{c}$
	The second order moment (σc)	$\sigma_c = \left[\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(p_{ij}^{c}-\mu_c)^2\right]^{1/2}$
	The third order moment (θc)	$\theta_c = \left[\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(p_{ij}^{c}-\mu_c)^3\right]^{1/3}$
Texture Characteristics	Entropy	$Entropy = \sum_i\sum_j C(i,j)\log(C(i,j))$
	Energy	$Energy = \sum_i\sum_j C^2(i,j)$
	Contrast	$Contrast = \sum_i\sum_j (i-j)^2\, C(i,j)$
	Homogeneity	$Homogeneity = \sum_i\sum_j \frac{C(i,j)}{1+\lvert i-j\rvert}$

μc_R	μc_G	μc_B	σc_R	σc_G	σc_B	θc_R	θc_G	Θc_B	ENT	E	C	H

Fig.2 *Immune antibody and antigen representation from color and texture features. (μc_R, μc_G, μc_B) The red, green and blue color moment of the first order. (σc_R, σc_G, σc_B)The red, green and blue color moment of the second order. (θc_R, θc_G, θc_B)The red, green and blue color moment of the third order. ENT, E, C and H represent respectively the texture characteristics: entropy energy, contrast and homogeneity.*

4. The affinity antibody-antigen:

The affinity antibody-antigen describes the link force by which the antibody and the antigen are connected [2]. It can be estimated by measuring a distance. Here, the Euclidean distance is considered. It is given by the formula 2.

$$aff = \sqrt{\sum (Ag - Ab)^2} \quad (2)$$

5. The Proposed Approach

The principle of the algorithm, as defined previously, adjusted to the skin color classification is presented in the following steps:

1. Initialization phase.
2. Censoring phase which corresponds to the learning process.
3. Monitoring phase corresponding to the classification process.

Initialization:
- Features extraction: Select a set of pixels and their neighborhoods in the training set representing a skin area of a chosen image. This set denotes the self set.
- Parameters extraction: Select the threshold r which is a parameter that establishes the degree of matching on r positions.

Learning:
- Generate randomly a number N of candidate detectors;
- Affinity evaluation: Determine the Euclidean distance between the sensors and each candidate element of the training set (the self set).
- Selection:
If a candidate detector matches an element of self on r contiguous positions, then The detector is removed;
Otherwise the candidate is added to a relevant detectors set (possible Solutions);

Classification:
- Select a pixel and calculate appropriate measures on its neighborhood.
- Match the selected pixel with relevant detectors set.

If two strings match on r positions Then
the element fits to non-self set, is classified among the non skin class and it is removed from the population;
Otherwise the pixel fits to the self set and it is classified among the skin class;

6. Experimental Results and Discussion

The negative selection algorithm learning process is done by using a collection of 300 images composing the library of skin texture. Some examples of the library are presented in fig.3. The skin types typically include: whitish, brownish, yellowish and darkish skins. Indoor and outdoor images with varying lighting conditions and varied backgrounds are also included in the learning database.

Fig.3 Some samples of the skin window library.

For each image in the learning database, we have considered a set of pixels with a neighborhood window size of 16*16 pixels. The number N of detectors, which are generated randomly, is 500 detectors. The parameter r defines the number of the contiguous position by which skin strings and detectors are connected, it is chosen at the value 6.

At the end of learning process, we have hold 300 relevant detectors, defining non skin pixels. Therefore, 200 detectors have been removed from the initial population of self detectors.

In the following step, we have performed the classification process. To test our system we have used images of human beings, which were not used in the learning process. They were being the input of the negative selection algorithm, after the first step, in order to obtain the generalization results.

The used images set, is a collection of different web persons and human faces image make up the database. Some samples are presented in table 3. We have mainly considered images with different ethnic groups, individually or the main in the same image (image1 to 15), different backgrounds (image11 to image 15) and different lighting conditions (images 8, 10, 13,15).

The software tool developed in MATLAB 9 displays the image in a window with GUI support as shown in fig.4. User clicks on a pixel within a skin area which is then taken as a middle pixel for 16*16 window's pixels. The interface right side shows results.

Table 3: Images sample, with diverse backgrounds and different ethnic groups.

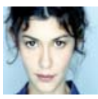


Image 1	Image 2	Image 3	Image 4	Image 5

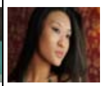
				
Image 6	Image 7	Image 8	Image 9	Image 10
				
Image 11	Image 12	Image 13	Image 14	Image 15

Fig. 4 Skin detection tool in action: Showing the screen immediately after loading image; Identified skin pixels are shown at the right side of the window.

This user interactive software tool is developed by authors on the similar lines as that in [8], It can be used to display skin pixels from an image. See fig.5.

Fig. 5 Another screen test applied on image 10. Identified skin pixels are shown on the same image.

We evaluate the performance of the proposed approach in terms of sensitivity, specificity, precision and overall accuracy. Sensitivity, true positive fraction, is the probability that a test is positive, that the given pixel is part of a skin class. Specificity, true negative fraction, is the probability that a test is negative, that the given pixel is not a part of the skin class. Precision is defined as the proportion of the true positives against all the positive results (both true positives and false positives). Overall accuracy is the probability that a test is correctly performed. The four indices are defined as in (3):

$$Sensitivity = \frac{TP}{TP+FN}; \qquad Specificity = \frac{TN}{TN+FP} \qquad (3)$$

$$Precision = \frac{TP}{TP+FP}; \quad Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

Where: TP, FP, TN, and FN refer respectively to: true positive, false positive, true negative and false negative. Those values are derived from a confusion matrix.
The true positive (TP) represents the number of skin pixels correctly classified as skin pixels while false positive (FP) represents the number of non-skin pixels incorrectly classified as skin pixels. The true negative (TN) represents the number of non skin pixels correctly classified as non skin pixels while false negative (FN) represents the number of skin pixels incorrectly classified as non skin pixels.

Some classification rates, by applying the negative selection algorithm to skin detection, assessed from table 3 are shown in Table 4.

Table 4: Qualitative Analysis of the proposed approach.

Images	Specificity	Sensitivity	Precision	Accuracy (%)
Image 1	0,92	0,98	0,89	94,02
Image 2	0,94	0,98	0,94	96,05
Image 3	0,95	0,95	0,93	95,66
Image 4	0,87	0,93	0,86	91,10
Image 5	0,96	0,88	0,88	94,11
Image 6	0,98	0,76	0,97	89,03
Image 7	0,90	0,88	0,65	90,15
Image 8	0,97	0,80	0,91	93,22
Image 9	0,84	0,96	0,89	91,01
Image 10	0,86	0,97	0,80	90,02
Image 11	0,75	0,69	0,53	73,12
Image 12	0,93	0,93	0,85	93,22
Image 13	0,76	0,84	0,54	78,23
Image 14	0,90	0,87	0,71	89,06
Image 15	0,91	0,88	0,83	90,02
		Overall Accuracy	89,67 %	

Some conclusions can be extracted from the experiments:

- For the considered test images the skin detection accuracy dependent greatly on the content of the image. Those with simple backgrounds, image 1 to 10, give better results compared to those with detailed background, image 11 to 15. They have high accuracy and a better precision.

- We can sort the considered images samples set on three categories:
  The first category is a set of persons' face with simple background and different illumination conditions (image 1 to 5). The second is a set of images containing persons groups also with different illumination conditions but with detailed background (image 6 to 10). The last one considers person's images in real scenes, so more detailed backgrounds are treated here (image 11 to 15).

- In the first category, the test images yield very interesting results with an overall accuracy varying between 91.10 % and 96.05 % , and a sensitivity of 0,88 to 0,98 (See pink cells in table 4). This is due to some misclassified pixel in persons' necks or in chestnut hair regions in whitish persons, the case in image 4.

- For the test images in the second category, we obtain an accuracy varying between 89, 03 % and 93,22% and a sensitivity of 0,76 to 0,97 (See mallow cells in table 4). The accuracy reduction is mainly due to the color clothing, like in image 6, or the shadow in image 10,where those parts are classified as skin regions

- In the last category, the test images gives an accuracy rate varying between 73,12% and 93,22%, a sensitivity of 0,69 to 0,93 (See blue cells in table 4). Skin textures are mainly smooth surfaces. If this same kind of texture can be found on other surface of the analyzed pictures, the negative selection algorithm possibly classify this texture as skin, this is the case in the image 10 (results in fig. 5) and image 13, in both wood is classified as skin. Also, the brightness in image 15 is identified as skin region.

- By considering the results given above, we can conclude that negative selection algorithm to skin detection and classification gives very promising results on different ethnic groups, different illumination condition more than different backgrounds, with an overall accuracy of 85% in the generalized case.

7. Conclusion and Perspectives

In this paper, we have proposed a skin detection and classification approach based on the negative selection algorithm from artificial immune systems. The proposed approach is based on a supervised learning coming which aims to classify pixel representations as skin or non skin.

The approach mainly operates on colors images; the processing is done on two steps: learning and classification. Besides the use of colors features, we also make use of texture information, which makes it robust against noisy pixels and so increase the overall accuracy. Moreover, we have not used any strict threshold, as some approaches dedicated for this purpose, which makes it applicable in a variety of imaging conditions. We are pretty confident about its performance in different circumstances such: different ethnic groups, over different backgrounds and illuminations.
In future work, we will further investigate the potential influence of other parameters and we will use alternative indices for measuring the affinity measure, the representations space and the image colorimetric space.

References

[1] S.Mitra, Tinkuacharya, Data Mining, Multimedia, Soft Computing and Bioinformatics, Published by John Wiley & Sons, Inc., Hoboken, New Jersey. 2003.
[2] L. R. De Castro and J.Timmis, Artificial Immune Systems: A New Computational Intelligence Paradigm, Springer-Verlag New York, Inc., Secaucus, NJ, 2002
[3] D. Dasgupta, Artificial Immune Systems and their applications, Springer-Verlag (Ed.), 1999.
[4] E. Hart and J.Timmis , Application areas of AIS: The past, the present and the future, Journal of Applied Soft Computing, v.8 n.1, p.191-201, January, 2008
[5] J.Timmis, Artificial immune systems---today and tomorrow, Natural Computing: an international journal, v.6 n.1, p.1-18, March 2007
[6] S. Forrest and Al, Self-Non self Discrimination in a Computer, Proceedings of the IEEE Symposium on Security and Privacy, p.202, May 16-18, 1994
[7] P. Kakumanu, S. Makrogiannis, N. Bourbakis, A survey of skin-color modeling and detection method, Journal of The Pattern Recognition Society 40 1106 – 1122, 2007.
[8] K. Bhoyar and O. Kakde, Skin Color Detection Model Using Neural Networks and its Performance Evaluation, Journal of Computer Science 6 (9): 963-968, ISSN 1549-3636. 2010
[9] A. Conci, E. Nunes J.J. Pantrigo, A. S´anchez, Comparing Color And Texture-Based Algorithms For Human Skin Detection, The 10th International Conference on Enterprise Information Systems June (13-16) in Barcelona, Spain. Proceedings Databases and Information Systems Integration – volume 5 – Human – Computer Interaction pp. 168-173 - ISBN: 978-989-8111-40-1, 2008.
[10] M. Abdullah-Al-Wadud, Mohammad Shoyaib, and Oksam Chae, A Skin Detection Approach Based on Color Distance Map, EURASIP Journal on Advances in Signal Processing Volume 2008, Article ID 814283, 10 pages doi:10.1155/2008/814283, 2008.
[11] Z .Ji and D.Dasgupta, Revisiting Negative Selection Algorithms, Evolutionary Computation, Massachusetts Institute of Technology v.15 n.2, p.223-251, 2007
[12] K. Igawa , H. Ohashi, A negative selection algorithm for classification and reduction of the noise effect, Journal of Applied Soft Computing, Elsevier Ed, 431–438, 9 - 2009.
[13] Thair Nu Phyu, Survey of Classification Techniques in Data Mining, Proceedings of the International Multi Conference of Engineers and Computer Scientists 2009 Vol I, IMECS 2009, March 18 - 20, Hong Kong, 2009.
[14] S. B. Kotsiantis, Supervised Machine Learning: A Review of Classification Techniques, Informatica 31 pp 249-268, 2007.
[15] M.Kamber, J.Han, Data Mining:Concepts and Techniques, Second Edition University of Illinois at Urbana-Champaign ISSN 13: 978-1-55860-901-3 ISBN 10: 1-55860-901-6, by Elsevier, 2006.
[16] R. M. Haralick, K. Shanmugam, and I. Dinstein, Textural Features for Image Classification, IEEE Transactions on Systems, Man, and Cybernetics, 3, 610-621, 1973.

Esma BENDIAB: Assistant professor at Mentouri University of Constantine, at Computer science department. In 2003, she received the M.S. degree in informatics from Mentouri University of Constantine, Algeria. Now, she is working toward the Ph.D.degree and she is a member of modelisation and implementation of complex system laboratory MISC at the same university. Her main research interests are classification, intelligent algorithms, artificial immune system and complex systems.

Dr. Mohamed Kheireddine KHOLLADI: is a Doctor of state in data processing (graduate doctor of the INSA of Lyon). He is currently a head of computer science department of the faculty of Science of the engineer at Mentouri University of Constantine and president of association science and development of the university. He is a head of modelisation and implementation of complex system laboratory (MISC). He is member of the network Cassini in France and member of network ACIT. His main research interests are geographical information systems computer graphics, knowledge bases and the new technology of information and communication, etc.