An Interactive Colour Video Segmentation using Granular Reflex Fuzzy Neural Network

Abhijeet V. Nandedkar, *Member, IAENG*

***Abstract*— Granular data processing using neural networks is an upcoming approach in pattern recognition. This paper proposes an interactive colour video segmentation technique based on Granular Reflex Fuzzy Min-Max Neural Network (GrRFMN). Inspired from the structure of human nervous system, GrRFMN architecture is consisting of a reflex mechanism to handle overlaps amongst the classes. It is noted that most of the image and video segmentation techniques are pixel based. It means that segmentation is carried out on pixel-by-pixel basis. Instead, in this paper a novel data granule based approached for colour video segmentation is presented. The proposed technique processes granules of a video frames. This results into a fast segmentation process. The video segmentation discussed here is of supervised type. In the training phase, GrRFMN learns different classes in an image or a video frame with an interaction with user. A trained GrRFMN is then used to segment the subsequent video sequences. The main advantage of using GrRMN is its capability to learn on-line in a single pass through data and property to deal with granular data which is helpful to segment a video sequence in an interactive mode. Results of the proposed method on standard images and video sequences are also presented.**

Index Terms—Granular Computing, Image Segmentation, Video Segmentation, Supervised learning, Granular Neural Network.

I. Introduction

A CENTRAL and prerequisite step in colour image and video understanding is segmentation. In this paper, a novel interactive colour video segmentation approach using Granular Reflex Fuzzy Min-Max Neural Network (GrRFMN) [1] is proposed. It has been observed that most of the image and video segmentation techniques are pixel based [3-7]. A data granule represents a bunch or a group of pixels in the form of hyperbox [1]. In the proposed segmentation technique, instead of pixels, such granules of an image or a video frame are processed.

Many successful applications of computer vision to image or video manipulation are interactive by nature. However, parameters of such systems are often trained neglecting the user [11]. It is observed that the conventional image segmentation is carried out in an unsupervised mode i.e. pixels of an image are classified into different categories using some homogeneity criteria.

In the proposed interactive segmentation, user interaction is added to the segmentation process. With the help of user interaction a semantic object can be defined easily and some uncertain borders can be decided. User interaction happens usually at the initial stage of the segmentation.

Abhijeet V. Nandedkar is with the Department of Electronics and Telecommunication Engineering, S.G.G.S. Institute of Engineering & Technology, Vishnupuri, Nanded, Maharashtra, PIN 431606, India. (e-mail: avnandedkar@yahoo.com).

In general, colour image segmentation techniques can be classified as histogram thresholding based, neighborhood based, clustering based and neuro-fuzzy based. Histogram thresholding is one of the simple and widely used techniques for image segmentation. The underlying assumption of histogram thresholding is that objects in the scene give rise to explicit peaks in image histogram. Thus, the segmentation task is reduced to find thresholds dissecting the image histogram [3]. An adaptive multi-thresholding approach for CIS can be found in [4]. However, a major drawback in the histogram thresholding techniques is the lack of use of spatial relationship amongst the pixels [5]. The neighborhood based approach (e.g. region growing) generally uses the uniformity criteria to segment regions in the image. These methods are better than histogram thresholding since they consider spatial relationship amongst. However, problem with these methods is the selection of initial seed points and the order in which pixels and regions are examined [6]. Clustering based approaches generally use fuzzy logic to define membership of the pixels [7]. Regions are created by inspecting the membership values of pixels using partition method (e.g. Fuzzy C-Means (FCM) clustering algorithm) [7].

This work proposes a use of image granules for segmentation. Granulation of information is an inherent and omnipresent activity of human beings carried out with intent of better understanding of the problem [8]. In fact, information granulation supports conversion of clouds of numeric data into more tangible information granules. The concept of information granulation within the frame work of fuzzy set theory was formalized by Zadeh in his pioneering work [9]. He observed that humans mostly employ words in computing and reasoning; and information granulation is a part of human cognition and proposed Theory of Fuzzy Information Granulation (TFIG) and Fuzzy Logic as a tool for Computing with Words (CW) [9]. This paper is a small step towards implementation of this concept for segmentation of colour video sequences. The major advantage of this technique is that it incorporates user's knowledge for segmentation. In addition, since granules of image are used for segmentation, execution of algorithm is fast. The rest paper is organized as follows. Section 2 briefly discusses GrRFMN architecture and its learning. Section 3 elaborates the proposed method. Section 4 presents experimental results and conclusions.

II. GrRFMN Architecture Details

The segmentation task in the proposed method is carried out using GrRFMN (Figure 1). GrRFMN is a fuzzy hyperbox set based neural network. A hyperbox is a simple geometrical object which can be defined by stating its min and max points (e.g. is shown in Figure 1). GrRFMN learns

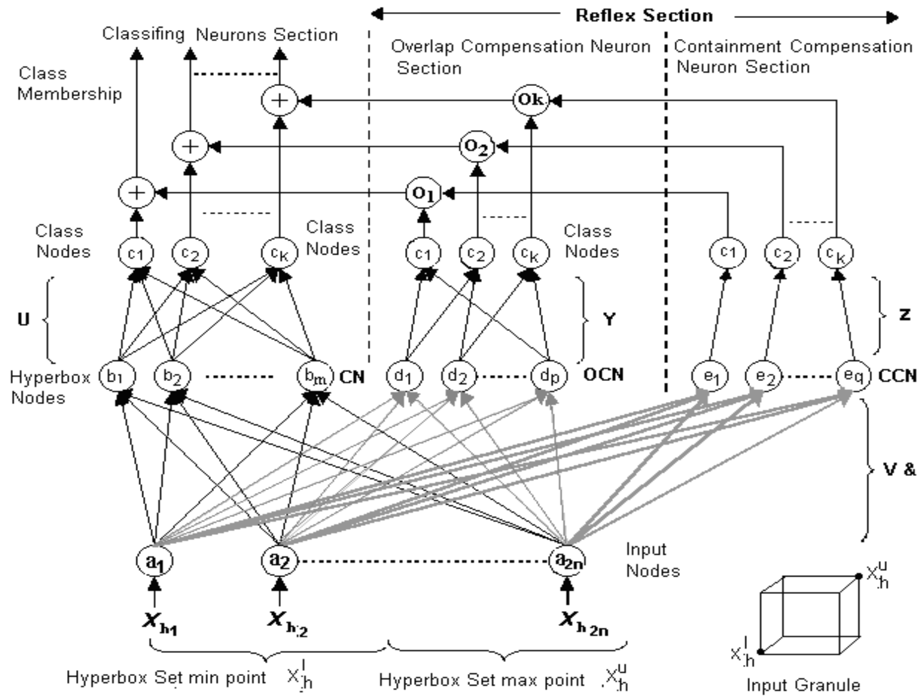


Figure 1. GrRFMN Architecture

different classes by aggregating hyperbox fuzzy sets [1]. It accepts input in the form of min-max oint of a granule i.e. in the form of hyperbox. Advantage of GrRFMN is that it is capable to handle data granules of different sizes efficiently, on-line in a single pass through data.

GrRFMN architecture is divided into two sections as Classifying Neuron section and Reflex section. The main task of classification of input data is carried out by Classifying Neuron section. It computes membership with the learned classes. Reflex section which based on reflex mechanism of human brain, adds compensation to the output if input data belongs to a class overlap region. Reflex section is further subdivided as overlap compensation and containment compensation sections. Neurons in these sections get activated only if the input sample or granule belongs to the overlap region. This action is very similar to the reflex action of human brain which takes over the control unconsciously in hazardous conditions.

Here it is assumed that all input features are scaled in the range [0-1]. An n-dimensional input granule is represented by, $X_h=[X_h^u, X_h^l]$ where X_h^l, $X_h^u=(x_{h1}, x_{h2,\ldots,}x_{hn})$ are min and max point vectors of the input granule respectively. A point data is a special case with $X_h^l=X_h^u$. Appending min and max point vectors, the input is connected to the nodes x_{h1}-x_{h2n}.

A. Classifying Neurons

The neurons b_1-b_m are classifying neurons. Outputs of classifying Neurons (CNs) belonging to a class are collected at a class node C_i in the output layer. The activation function of the **classifying neuron** b_j is given by Eq. 1-6,

$$\mu_{b_j}(X_h, V_j, W_j, \gamma) = \frac{1}{n}\sum_{i=1}^{n} A_{ji} + B_{ji} \tag{1}$$

where γ = Fuzzyness parameter, $y = 1/\gamma$

$$A_{ji} = \frac{(r_{ji} + l_{ji} + c_{ji})}{a_i - b_i + \varepsilon} \tag{2}$$

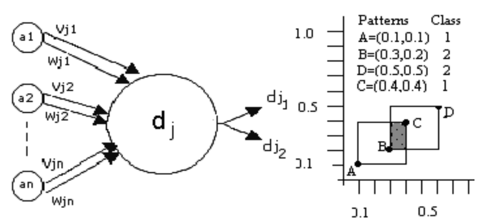
$$c_{ji} = \max(\min(b, w_{ji}) - \max(v_{ji}, a), 0) \tag{3}$$

$$l_{ji} = \frac{1}{2y}(\max((\min(v_{ji}, b_i) - \max(v_{ji} - y, a_i)), 0) \times (\max(a_i, v_{ji} - y) + \min(b_i, v_{ji}) - 2\times(v_{ji} - y))) \tag{4}$$

$$r_{ji} = \frac{1}{2y}(\max((\min(b_i, w_{ji} + y) - \max(w_{ji}, a_i)), 0) \times (-(\max(a_i, w_{ji}) + \min(b_i, w_{ji} + y) + 2\times(w_{ji} + y)))) \tag{5}$$

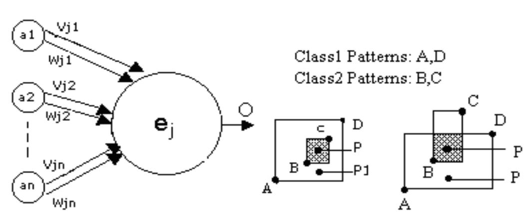
$$B_{ji} = \mathrm{U}(a_i - b_i) \times \max\left(\min\left(\frac{a_i - (v_i - y)}{y}, \frac{(w_i + y) - a_i}{y}, 1\right), 0\right) \tag{6}$$

$[a,b]$, $[v,w]$: min-max points of input and hyperbox fuzzy set (HBF set) respectively. $\varepsilon = 10^{-10}$ avoids division by zero error in case data is in point form, U: unit step function.



(a) (b)

Figure 2. Overlap Compensatory Neuron



(a) (b)

Figure 3. Containment Compensatory Neuron

ISBN: 978-988-19251-4-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

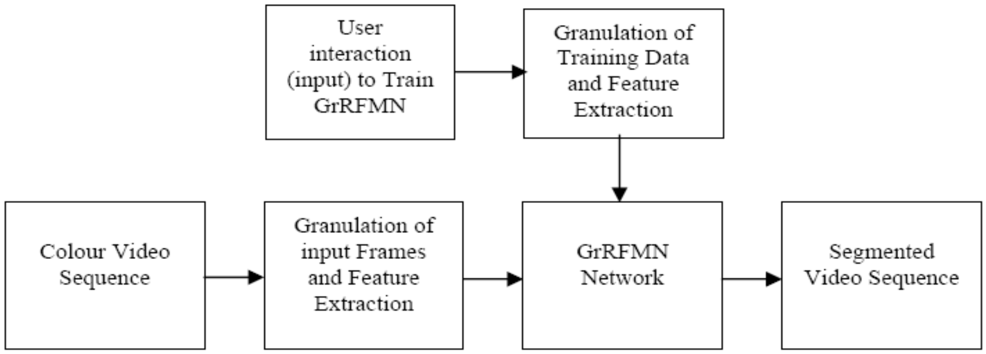


Figure 4. Proposed Colour Video Segmentation (CVS) technique.

B. Overlap and Containment Compensation Neurons

While training GRRFMN, situation depicted in Figure 2(*b*),3(*b*) where hyperboxes of different classes are overlapping, is bound to occur. Overlap compensation neurons (OCN) and containment compensation neurons (CCN) are trained to handle these situations. Nodes d_1-d_p and e_1-e_q represent OCNs and CCNs respectively. Outputs of these neurons are collected at the respective class nodes.

Figure 2(*a*) shows details of an OCN, which represents a hyperbox of size equal to the overlap region between two hyperboxes. Outputs of OCN are connected to the respective nodes of classes facing the overlap problem. OCN activation function is given by Eq.7,8.

$$\begin{aligned} d_{j_p} = {} & U(b_j(X_h,V,W,\gamma)-1))\times \\ & \left(-1+\max\left(b_{j1}(X_h^u,V_p,W_p),\, b_{j1}(X_h^l,V_p,W_p)\right)\right) \end{aligned} \tag{7}$$

where $$b_{j1}(X_h^k,V_p,W_p)=\left(\frac{1}{n}\sum_{i=1}^{n}\max\left(\frac{x_{hi}^k}{w_{p_{ji}}},\frac{v_{p_{ji}}}{x_{hi}^k}\right)\right) \tag{8}$$

p=1,2. d_{j1} and d_{j2} are outputs for Class1 and Class2. *V, W*: min-max point of OCN. V_1,W_1,V_2,W_2: min-max point of overlapping hyperboxes *U(x)* : is a unit step function. $b_j()$ is same as Eq.1.

A CCN is shown in Figure 3(a). This neuron represents a hyperbox of size equal to the overlap between two classes as shown in Figure 3(b). The activation function of CCN is:

$$O_{c_j} = -1\times U(b_j(\mathrm{X_h},\mathrm{V},\mathrm{W},\gamma)-1) \tag{9}$$

where Oc_j: output , *V,W*: min-max point of CCN, *U(x)* : unit step function, *bj*( )*:* same as Eq. 1.

The output of CCN is connected to the class that contains the hyperbox of other class. The number of output layer nodes in CL section is same as number of classes learned. The number of class nodes in reflex section depends on the nature of overlap network faces during the training process. Final membership calculation is given by,

$$\mu_i = \max_{j=1..m}(b_j u_{ji}) + \min(\min_{j=1..p}(d_j y_{ji}), \min_{j=1..q}(e_j z_{ji})) \tag{10}$$

where *u,y,z* are the connection matrices for the neurons in the three sections. *m,p,q* are number of neurons in respective sections.

Training of GrRFMN begins by presenting training granules sequentially. Network tries to accommodate the training samples in hyperboxes for the given class. During training if hyperboxes belonging to different class are found overlapping, respective compensation neuron is added to the network. Training of GrRFMN is on-line and single pass through the data. More details about the training algorithm for GrRFMN are given in [1].

III. Granular Video Segmentation

Colour Video Segmentation (CVS) is an important issue for understanding a scene in a video. Here the problem of CVS is tackled using granular computing. The proposed method utilizes capability of GrRFMN to acquire knowledge through granules of data. The main endeavor behind development of this technique is to demonstrate that instead of processing individual colour pixels (a 3D vector), a group of pixels (granules) can be processed very easily using GrRFMN. Obviously this reduces lot of computational cost required to process individual pixels. The proposed for CVS technique is shown in Figure 4.

This CVS method uses CIE (L-a-b) colour space. The reason is that CIE space can control colour and intensity information more independently than RGB colour space. This colour space is especially efficient in the measurement of small colour changes, as a result direct colour comparison can be performed based on geometric separation within this colour space [10].

The main steps in CVS implementation are a) Granulation of user interaction through samples of an image or video frame presented for training, b) Training of GrRFMN, c) Granulation of frames of test video sequence. To perform CVS, the system is supposed to be trained with a user intervention for the first video frame. Subsequently, proposed CVS system segments the frames based on the earlier acquired knowledge. The detailed scheme is explained as follows.

In the proposed CVS method GrRFMN is trained with labelled granules (seed granules) constructed from the seed images of different objects/parts in the first training video

ISBN: 978-988-19251-4-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

frame. These seed-images are sub-divided into grids of size (k x k). A granule for each grid is represented by a hyperbox which can be represented by simply stating its min and max vertices. Thus, a hyperbox is computed by finding min-max values of the pixels in that grid,

$$\text{i.e.} \quad V=[L_{min}, a_{min}, b_{min}] \qquad W=[L_{max}, a_{max}, b_{max}]) \tag{11}$$

Along with this information, a mean value of the grid for the three planes is also extracted and added to the min-max vector as,

$$\text{i.e. } V=[L_{min}, a_{min}, b_{min,} L_{mean}, a_{mean}, b_{mean}] \qquad W=[L_{max}, a_{max}, b_{max}, L_{mean}, a_{mean}, b_{mean}] \tag{12}$$

Such granules are then used to train GrRFMN. To segment a given frame, it is sub-divided into small grids of size (n x n) and granules are formed. These granules are then fed to GrRFMN for segmentation.

An example of colour image segmentation using proposed method is shown in Figure 5. Here four sub-images (Fig. 5(b)) belonging to two different classes (i.e. house and sky) are used to train GrRFMN. During training seed-images are divided into a grid (e.g. size 3x3, 5x5 and 10x10) and granules (hypeboxes) are formed. These granules are then used for training GrRFMN. Here, GrRFMN is trained with an expansion coefficient (θ) [1] equal to 0.2. In the test phase, the given image is granulated with various grids sizes (5x5, 10x10, and 15x15). The segmentation results are shown in Fig. 5. Note that for various training and test grid sizes, the performance of GrRFMN is almost consistent.

The following section discusses few more experimental results of colour image and video segmentation.

IV. EXPERIMENTAL RESULTS

Here aim is to test proposed system on some real images and video sequences. The results are as follows:

A. Colour Image segmentation

In this case, training granules size here is kept as 5x5 and test granule size is 3x3 and 5x5, Expansion coefficient of GrRFMN is (θ=0.2). Note that training GrRFMN training is done on samples of classes and are represented by a pseudo colours in segmented output image (Figure 6).

Note that in Figure 6 (d), the image consists of two classes 1) lady and 2) background. Observe that class “1” consists of three different colours. To segment the lady properly, the training sample is chosen such that it consists of both coloured body parts of lady. It may be observed that the two classes in the image i.e lady and background are segmented properly. This shows the capability of proposed system to group different colours in a class, if required. From results demonstrated in Figure 6, one can note that proposed is capable to classify image granules efficiently.

B. Colour Video Segmentation

The performance of the proposed CVS system on video sequence is tested in this section. In this experiment, with the help of user interaction, few image/frame samples of each class are selected. These samples are granulated using grid size (5x5). These granules are used to train GrRFMN. Expansion coefficient (θ) of GrRFMN was kept equal to 0.2. To segment subsequent frames, test granules with size (5x5) are formed. The results depicted in Figure 7 demonstrate the capability of proposed CVS system.

One may add more user interaction to improve resultant segmentation.

V. CONCLUSION

Granular computing is a powerful tool and is found suitable in colour video segmentation. It is observed that even with different training and test granule sizes the proposed method works efficiently. This allows a trained CVS system to work on required resolution without retraining. Since it is an interactive segmentation technique, one can assign different coloured regions which belong to one category easily. Hence, an improved segmentation result can be achieved.

ACKNOWLEDGEMENT

The author would like thank All India Council for Technical Education (AICTE), New Delhi for financially supporting the project under the scheme of Career Award to Young Teachers.

REFERENCES

[1] A. V. Nandedkar, P. K. Biswas, “ A Reflex Fuzzy Min Max Neural Network for Granular Data classification**,** IEEE Transaction on Neural Networks**,** vol. 20, no.7, pp. 1117-1134, 2009.

[2] Hanfeng Chen, Feihu Qi, and Su Zhang, “Supervised video object segmentation using a small number of interactions,” IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '03). Vol. 3, no. 3, pp. 365-368, 2003.

[3] Rosenfeld, L. S. Davis, “Iterative histogram modification,” IEEE Trans. Syst. Man Cybern. Vol. 8, pp. 300-302, 1978.

[4] Kanchan Deshmukh, A.V. Nandedkar, Y.V. Joshi and G.N. Shinde, “Multilevel Approach for Color Image Segmentation,” Proc. Indian Conference on Computer Vision and Graphics, ICVGIP 2004, pp. 338-342, 2004.

[5] M.K. Kundu, S.K. Pal, “Thresholding for edge detection using human psychovisual phenomena,” Pattern Recognition Letter, vol. 4, pp. 433-441, 1986.

[6] W. K. Pratt, Digital Image Processing. Second Edition. New York: Wiley-Inter science, 1991.

[7] Y.W. Lim, S.U. Lee, “On the color image segmentation algorithm based on the thresholding and the fuzzy c-means technique,” Pattern Recognition, vol. 23, no. 9, pp.935–952, 1990.

[8] W. Pedrycz, "Granular Computing: An Introduction," in Proc. Joint IFSA World Congress and 20th NAFIPS Int.Cnf., vol. 3, pp. 1349-1354, 2001.

[9] L. A. Zadeh, "Towards a Theory of Fuzzy Information Granulation and its Centrality in Human Reasoning and Fuzzy logic," Fuzzy Sets and Systems, vol. 90, pp. 111-127, 1997.

[10] H.D. Cheng, X.H. Jiang, Y. Sun, and J. Wang, "Colour image segmentation: Advances and propects," Pattern Recognition, 34(12), 2259-2281, 2001.

[11] Hannes Nickisch, Carsten Rother, Pushmeet Kohli, Christoph Rhemann, “Learning an Interactive Segmentation System,” Proc. of Indian Cnf. on Computer Vision and Graphics, Chennai, Dec. 2010.

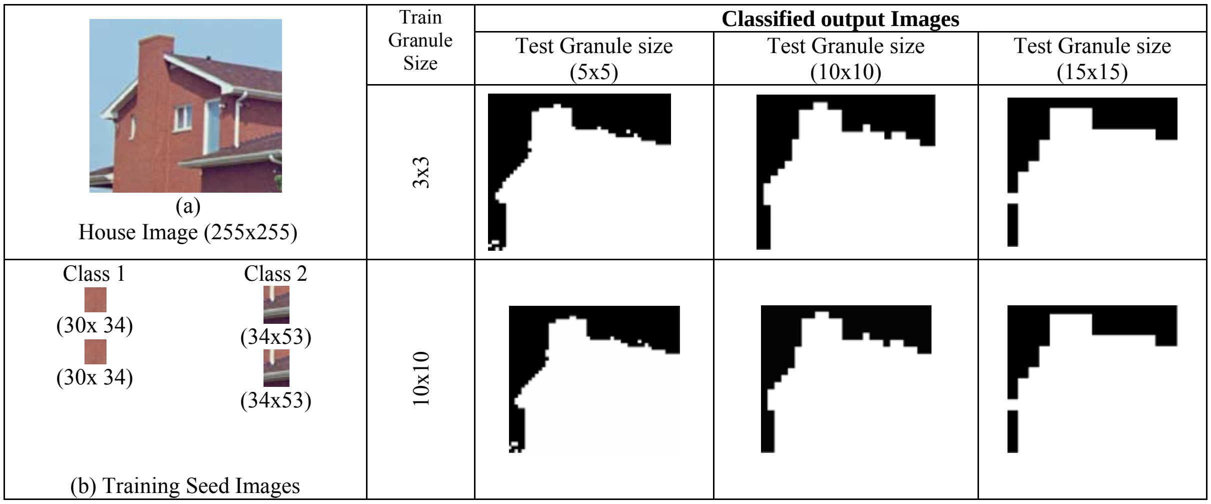


Figure 5. House Image (for output images Black: Class 1 and White: Class2).

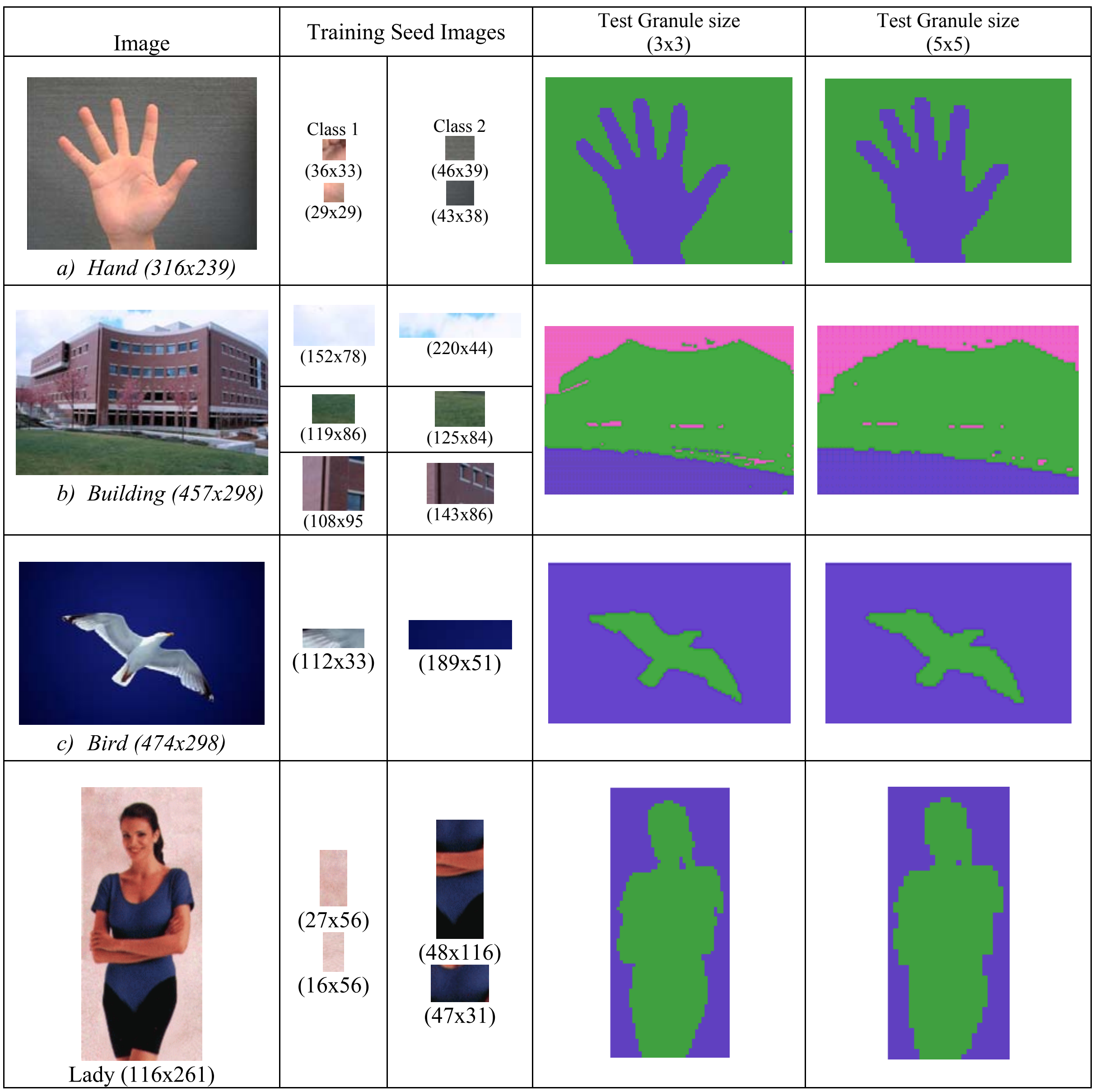


Figure 6. Segmentation results on different real images.

ISBN: 978-988-19251-4-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

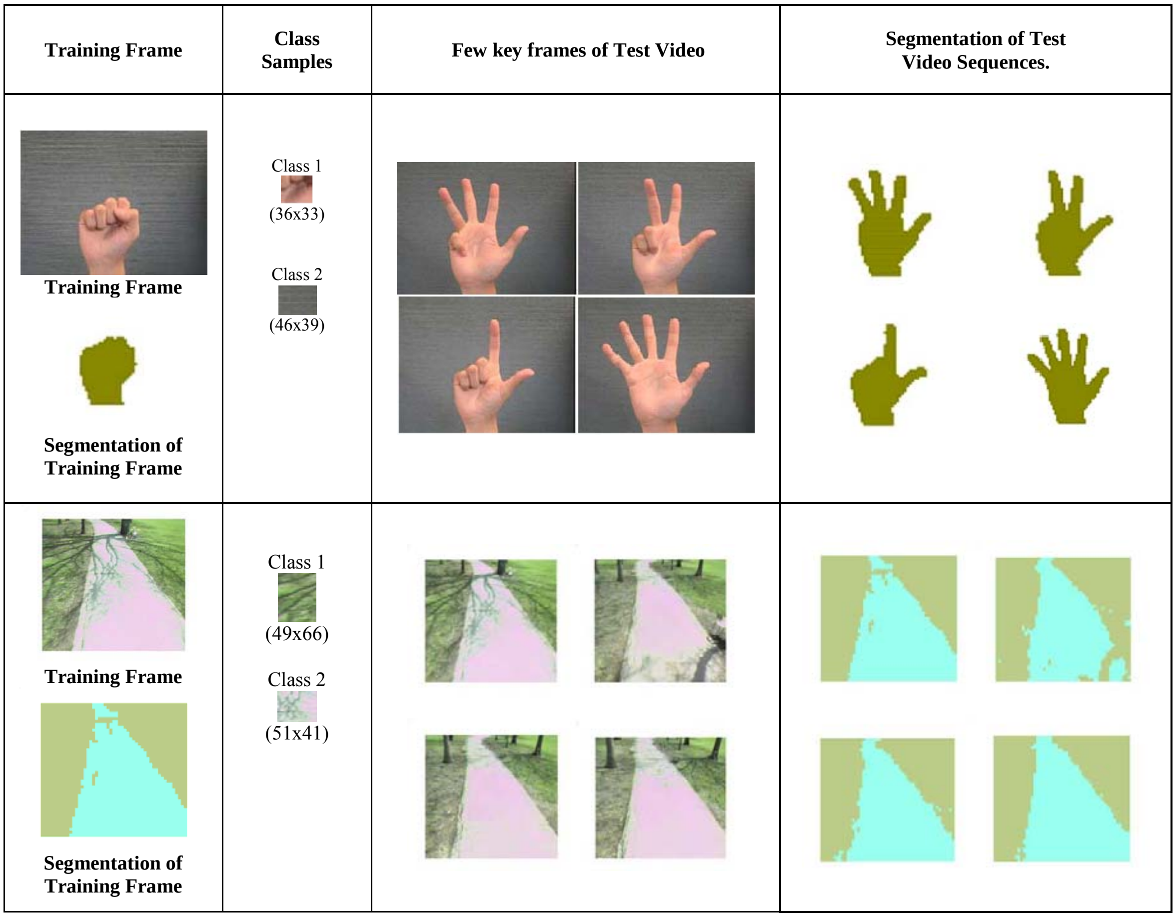


Figure 7. Result for Colour Video Sequences

ISBN: 978-988-19251-4-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)