

The Dynamic Image Segmentation for Sign Language Training Simulator

Oles Hodych¹, Kostiantyn Hushchyn¹, Iouri Nikolski², Volodymyr Pasichnyk², Yuri Shcherbyna¹

¹Ivan Franko Lviv National University, 1 Universytetska St, Lviv, Ukraine

²Nationa University "Lvivska Politechnica", street, Lviv, Ukraine

oles.hodych@gmail.com, kostya.gushtin@gmail.com, y_nikol@yahoo.com,
vpas@astra.lviv.ua, shcherbyna@franko.lviv.ua

Abstract. *The paper discusses the problem of noise reduction for the improvement of the hand recognition in a sequence of video frames. The presented results were obtained as a part of a larger project, which has an objective to build a training simulator for Ukrainian Sign Language. The proposed solution for image segmentation is build around the data mining techniques, which are based on clustering algorithms using Self-Organizing Maps. A particular emphasis in this research is made on the image preparation for Self-Organizing Map training, in order to provide the most optimal similarity measurement between different image segments.*

Keywords

Image segmentation, clustering, Self-Organizing Maps

1 Introduction

The presented in this paper results were obtained as part of the research for creating an adaptive sign language training simulator, which was already discussed in the number of articles [DN07] [DNP08] as well as demonstrated at CeBIT 2006, 2007 and 2008. The main motivation behind this research is to provide an affordable solution for people to use the end product for self training of the Ukrainian sign language. The current solution for recognizing sign language gestures is based on dactyl matching [DN07].

Unfortunately, this approach is limited to images (frames in video sequences) with uniform background. In order to overcome this limitation an image segmentation based solution has been developed for background removal utilising Self-Organizing Maps (SOM). One of the key SOM features is preservation of the topological order in the input data during the learning process. Some relatively recent research of the human brain revealed that the response signals are obtained in the same topological order on the cortex in which they were received at the sensory organs such as eyes transmitting the colour information [Ko01]. A large part of the conducted research is dedicated to image data preparation and colour space analysis to fully utilise the topology preservation feature of SOM when performing image segmentation. Image data preparation is the key component in the developed image segmentation methodology.

One possible way to approach image segmentation is to treat image segmentation as a clustering task, where objects for grouping are image pixels, and the actual groups (or clusters) are image segments. Hence the term *image clustering*, which refers to means for high-level description of the image content. SOM and its variations were the subject of our research for the past several years [Ho06] [Ho07], and therefore it was a natural choice for the task of image clustering. Image segmentation is a popular research subject and there is a large number of related research as well as readily available commercial and open source tools.

The use of SOM for image segmentation is a relatively popular research subject, and there is a number of SOM-based image segmentation algorithms with their virtues and weaknesses. In [MC96] authors proposed the use of the two-stage process based on one-dimensional SOM in combination with K-means algorithm for segmenting images with reduced colour information. Authors of [CTT96] developed a unique data preparation scheme where both colour and texture were

used, which yielded a success rate of 61.3% comparing to 53.6% when using only colour. In [BCL01] authors proposed a SOM-based algorithm for skin detection. Authors of [JCZ03] used both colour and spatial information. Thus, utilising five dimensional vectors (X, Y, R, G, B) for representing image data used in SOM training. In addition, a merging algorithm was introduced for clustered blocks to be combined into a specified number of regions with some semantic means. Paper [WLH00] discusses the use of an adaptive SOM colour segmentation algorithm, which supported pruning and merging techniques to automate search for an appropriate number of clusters. The presented results yield an excellent performance, however most of the processed images either have small size, thus automatically reducing the complexity, or have a close to uniform background.

2 Dynamic image segmentation

The discussion of the theoretical and algorithmic foundation of the conducted research is presented in two sections. The first sections, which is an essential part of our research, addresses data preparation. Specifically, the development of the most adequate method for transforming images from video sequences into the vector space suitable for SOM consumption (i.e. training and interpretation). This phase directly effects the quality of the image segmentation. For the purpose of clarity the following discussion assumes processing of a single image, which does not affect applicability of the proposed methodology to processing of video sequences.

2.1 Data preparation

The first stage of data preparation is to choose a vector space for representing each pixel on the image. The processing speed is of high importance for fulfilling the requirement of real-time processing. The training of SOM is the most time consuming operation. Therefore, the second stage of data preparation addresses the reduction of the number of data samples used for training.

Colour is the brain's reaction to specific visual stimulus. Therefore, in order to train SOM for it to reflect the topological order of the image perceived by a human eye, it is necessary to choose the colour space, which closely models the way sensors obtain the visual information. The eye's retina samples colours using only three broad bands, which roughly correspond to red, green and blue light (Abrian Ford, Alan Roberts). These signals are combined by the brain providing several different colour sensations, which are defined by the CIE (Commission Internationale de l'Eclairage (French))[Hu01]: Brightness, Hue and Colourfulness. The CIE commission defined a system, which classifies colour according to the human visual system, forming the tri-chromatic theory describing the way red, green and blue lights can match any visible colour based on the eye's use of three colour sensors.

The colour space is the method, which defines how colour can be specified, created and visualised. There are more than one colour space, some of which are more suitable for certain applications than others. Some colour spaces are perceptually linear, which means that an n -unit change in stimulus results in the same change in perception no matter where in the space it is applied (Abrian Ford, Alan Roberts). The feature of linear perception allows the colour space to closely model the human visual system. Unfortunately, the most popular colour spaces currently used in image formats are perceptually nonlinear. For example, BMP and PNG utilise RGB¹ colour space, JPEG utilises YCbCr, which is a transformation from RGB, HSL² is another popular space, which is also based on RGB.

The CIE based colour spaces, such as CIELuv and CIELab, are nearly perceptually linear (Abrian Ford, Alan Roberts), and thus are more suitable for the use with SOM. The CIEXYZ space devises an independent absolute colour space, where each visible colour has nonnegative coordinates X, Y and Z [Ho03]. The CIELab is a nonlinear transformation of XYZ onto coordinates L^*, a^*, b^* [Ho03]. The image format used in this research is uncompressed 24-bit BMP (8 bit per channel), which utilises the RGB colour space. In order to convert vectors $(r, g, b) \in RGB$ into $(L^*, a^*, b^*) \in CIELab$ it is necessary to follow an intermediate transformation via the CIE XYZ colour space (Gernot Hoffmann). Transforming each pixel of the original RGB image produces a transformed image in CILab space used for further processing. It is important to note that when using SOM it is common to utilise Euclidean metric³ for calculation of distances during the learning process [Ko01], which is also used in CIELab space for calculating distance between pixels [Ho03].

In order to reduce the dataset used for SOM training, it was decided to split the image into segments $n \times n$ pixels. Then for each such segment find two the most diverged pixels and add them to the training dataset. Finding the two most diverged pixels is done in terms of the distance applicable to the colour space used for image representation. Below is an

¹Uncompressed BMP files, and many other bitmap file formats, utilise a color depth of 1, 4, 8, 16, 24, or 32 bits for storing image pixels.

²Alternative names include HSI, HSV, HCI, HVC, TSD etc. (Abrian Ford, Alan Roberts)

³The selection of the distance formula depends on the properties of the input space, and the use of Euclidean metric is not mandatory.

example of image A of 4×4 pixels in size represented in the CIE Lab space, and split into four segments 2×2 pixels each.

$$A = \left(\begin{array}{cc|cc} (L_1^1, a_1^1, b_1^1)^T & (L_2^1, a_2^1, b_2^1)^T & (L_3^1, a_3^1, b_3^1)^T & (L_4^1, a_4^1, b_4^1)^T \\ (L_1^2, a_1^2, b_1^2)^T & (L_2^2, a_2^2, b_2^2)^T & (L_3^2, a_3^2, b_3^2)^T & (L_4^2, a_4^2, b_4^2)^T \\ \hline (L_1^3, a_1^3, b_1^3)^T & (L_2^3, a_2^3, b_2^3)^T & (L_3^3, a_3^3, b_3^3)^T & (L_4^3, a_4^3, b_4^3)^T \\ (L_1^4, a_1^4, b_1^4)^T & (L_2^4, a_2^4, b_2^4)^T & (L_3^4, a_3^4, b_3^4)^T & (L_4^4, a_4^4, b_4^4)^T \end{array} \right)$$

Algorithm 1 summarizes the above approach. It is important to note that an excessive reduction could cause omission of significant pixels resulting in poor training. At this stage it is difficult to state what rule can be used to deduce the optimal segment size, and the segmentation used in the presented results, was obtained empirically. However, even applying segmentation 2×2 pixels to an image of 800×600 pixels in size, reduces the training set from 460000 down to 240000 elements, which in turn enables the use of a smaller lattice and reduces the processing time required for SOM training.

Let n denote the size of segments used for image splitting, the value of which is assigned based on the image size. T – the training set, which is populated with data by the algorithm. Let's also denote j th pixel in segment S_i as $S_i(j)$. Further in the text both terms *pixel* and *vector* are used interchangeably.

Algorithm 1 Training dataset composition

Initialization. Split image into segments of $n \times n$ pixels; $N > 0$ – number of segments; $T \leftarrow \emptyset$; $i \leftarrow 1$.

1. Find two the most diverged pixels $p' \in S_i$ and $p'' \in S_i$ using Euclidian distance.
 - 1.1 $max = -\infty, j \leftarrow 1$
 - 1.2 $k \leftarrow j + 1$
 - 1.3 Calculate distance between pixels $S_i(j)$ and $S_i(k)$: $dist \leftarrow \|S_i(j) - S_i(k)\|$
 - 1.4 If $dist > max$ then $p' \leftarrow S_i(j), p'' \leftarrow S_i(k)$ and $max \leftarrow dist$
 - 1.5 If $k < n \times n$ then $k \leftarrow k + 1$ and return to step 1.3
 - 1.6 If $j < n \times n - 1$ then $j \leftarrow j + 1$ and return to step 1.2
 2. Add $p' \in S_i$ and $p'' \in S_i$ to the training set: $T \leftarrow T \cup \{p', p''\}$
 3. Move to the next segment $i \leftarrow i + 1$. If $i \leq N$ then return to step 1, otherwise stop.
-

2.2 Interpretation of Clusters

The guidelines from [Ko01] and [Ho06] were followed to conduct the self-organization process. The use of 2-dimensional lattice with hexagonal neighborhood relation proved to be the most efficient in our research producing more adequate segmentation results comparing to other evaluated configurations. Once the SOM structure and parameters for self-organization process are selected, the SOM is trained on the training set T , which is composed for the image to be segmented. The trained SOM is then used for the actual image segmentation. The topology preservation feature of SOM is fundamental to the proposed segmentation approach. The basic underlying principles of which are:

- Image pixels represented by topologically close neurons should belong to the same cluster and therefore segment.
- The marker used for segment representation is irrelevant as long as each segment is associated with a different one.

These two principles suggest that the neuron position in the lattice can be used for assigning a marker to a segment represented by any particular neuron instead of the neurons' weight vectors. This way weight vectors are used purely as references from 2D lattice space into 3D colour space, and neural locations represent the image colour distribution. As the result of a series of conducted experiments the following formulae for calculating an RGB colour marker for each neuron have produced good results: $R_j \leftarrow x_j + y_j \times \lambda$; $G_j \leftarrow x_j + y_j \times \lambda$; $B_j \leftarrow x_j + y_j \times \lambda$. Values x_j and y_j are coordinates of neuron $j = \overline{1, M}$, where M is the number of neurons in SOM. Constant λ should be less or equal to the diagonal of the SOM lattice. Applying the same formula for R, G, and B components produces a set of gray scale colours. However,

each neuron has its own colour, and one of the currently not fully resolved issues is how to group neurons based on the assigned colours into segments. There are several approaches, which are being currently developed to provide automatic selection of neurons pertaining to the same cluster [Ho07]. However, the presented in this paper results were obtained by applying a threshold to the segmented with SOM image, which requires human interaction in specifying the threshold value. Algorithm 2 summarizes the proposed approach.

Algorithm 2 Image segmentation

Initialization. $p_j = (R_j, G_j, B_j)$ – pixel j ; $j = \overline{1, K}$; $K > 0$ – total number of pixels; $j \leftarrow 1$; $i^*(p_j) = (R_{i^*}, G_{i^*}, B_{i^*})$ – a weight vector of the best matching unit (BMU – winning neuron) for input vector p_j ; (x_{i^*}, y_{i^*}) – coordinates of neuron i^* ; choose appropriate values for λ .

1. Find $BMU(p_j)$ for vector p_j in the trained SOM utilizing the distance used for training (Euclidian for CIELab).
 2. Calculate marker for pixel p_j : $R_j \leftarrow x_{i^*} + y_{i^*} \times \lambda$, $G_j \leftarrow R_j$, $B_j \leftarrow R_j$.
 3. Move to the next image pixel: $j \leftarrow j + 1$;
 4. If $j \leq K$ return to step 1, otherwise stop.
-

3 Experimental results

This section demonstrates some results of the developed image segmentation methodology. A specific interest is the case of training SOM on one of the frames in the video sequence only and applying it for segmenting subsequent frames. The recorded video captured an open palm closing and opening again during a period of several seconds. The recording was done using an ordinary PC web camera capable of 30FPS throughput with frame size of 800×600 pixels. The background of the captured scene is nonuniform, which increases the complexity of image segmentation. Fig. 1 depicts original and segmented images, which correspond to frames number 25 through to 50 of the recorded video. The training of SOM and determining of the appropriate threshold value was performed only on the first image (i.e. frame 25).

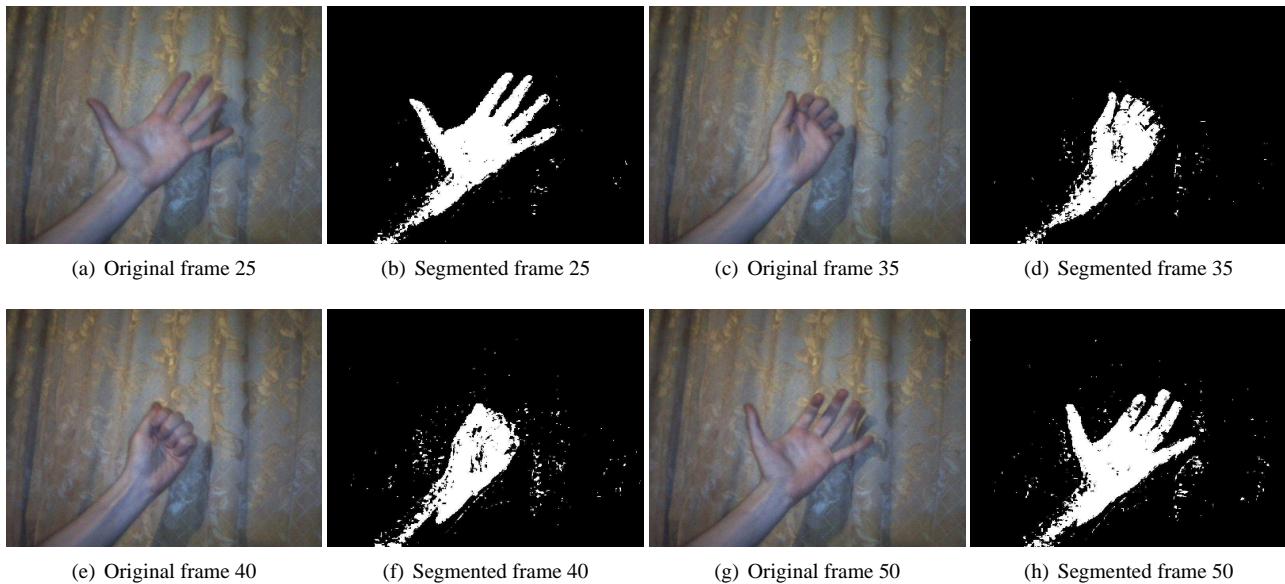


Figure 1. Original and segmented video frames 25, 35, 40, and 50

The key aspect of the presented in this section results is the use of the SOM trained on only one frame. This initial frame as well as all subsequent ones have been successfully segmented with clear separation of the human palm from the nonuniform background. Although, some elements of the background were recognized as part of the same segment and caused minor undesired artifacts scattered around the palm. The use of only one frame for SOM training provides a provision for much faster dynamic image segmentation needed for video, avoiding SOM retraining for every frame.

4 Conclusion and future work

The main purpose of developing an image segmentation algorithm in our case is to simplify the task of image analysis for dactyl matching. However, developed algorithms yielded good results not only for human hand recognition, and potentially can be used for other applications. The main disadvantage of the developed approach is the need for human interaction when specifying the threshold values in the final step of image segmentation. However, this aspect is being currently addressed utilising the results obtain in [Ho07], which allows automatic clustering of the trained SOM. Another important subject of the future research direction is increasing the quality of segmentation by applying hierarchical clustering. The basic idea behind this approach is to start SOM training on images with reduced information, following additional training based on the same image with increased information. There are many image smoothing methods, which provide a way of controlling the amount of image details (i.e. information), which may impact the quality of information reduction and thus segmentation results. The proposed in this research method for composing training dataset is also an example of the way to reduce image details.

References

- [DN07] Давидов М.В.; Нікольський Ю.В.: Автоматична ідентифікація елементів жестової мови за методом еталону. Вісник Нац. ун-ту "Львівська політехніка". Сер.: Інформаційні системи та мережі, №589, С.174-198, 2007.
- [DNP08] Давидов М.В.; Нікольський Ю.В.; Пасічник В.В.: Вибір ефективного методу опрацювання зображень на основі еталону для ідентифікації елементів жестової мови. Вісник Харківського національного університету радіоелектроніки. Сер. "АСУ і прилади автоматики", №139, С.59-68, 2008.
- [Ho06] Годич, О. et. al.: Дослідження ефективності алгоритмів навчання мереж Кохонена. Управляючі системи и машины, №2, 2006, С.63-80
- [MC96] Jander Moreira, Luciano Da Fontoura Costa: Neural-based color image segmentation and classification using self-organizing maps. 1996,
http : //mirror.impa.br/sibgrapi96/trabs/pdf/a19.pdf
- [CTT96] Neill W. Campbell, Barry T. Thomas, Tom Troscianko: Neural Networks for the Segmentation of Outdoor Images. International Conference on Engineering Applications of Neural Networks, pp.343–346, 1996.
- [BCL01] David Brown, Ian Craw, Julian Lewthwaite: A SOM Based Approach to Skin Detection with Application in Real Time Systems. University of Aberdeen, 2001,
http : //www.bmva.ac.uk/bmvc/2001/papers/33/accepted_33.pdf
- [JCZ03] Y. Jiang, K.-J. Chen, Z.-H. Zhou: SOM Based Image Segmentation. Lecture Notes in Artificial Intelligence 2639, pp.640-643, Springer, 2003.
- [WLH00] Ying Wu, Qiong Liu, Thomas S. Huang: An Adaptive Self-Organizing Color Segmentation Algorithm with Application to Robust Real-time Human Hand Localization. In Proc. Asian Conf. on Computer Vision, Taiwan, 2000.
- [Ko01] T. Kohonen: Self-Organizing Maps. 3rd edition, Springer, 2001
- [Hu01] R.W.G. Hunt: Measuring Colour, 3rd edition, Fountain Pr Ltd, 2001
- [Ho03] Gernot Hoffmann: CIELab Color Space. 2003,
http : //www.fho – emden.de/ hoffmann/cielab03022003.pdf
- [Ho07] Oles Hodych et. al.: High-dimensional data structure analysis using Self-Organising Maps. CAD Systems in Microelectronics, 2007, CADSM apos;07. 9th International Conference, 19-24 Feb. 2007 Page(s):218-221.