

# A Survey Paper on Methods of Detecting Human under Partial Occlusion

Sayali.S.Baxi, S.V.Dabhade, Priyanka.D.Varma

Department Of Computer Science, Smt Kashibai Navale College Of Engineering  
Pune University

## Abstract

Identifying humans under partial occlusion is a challenging problem in unconstrained scene understanding. Detecting humans from still images is an extensively studied problem in computer vision. We address this problem by studying the semantic context between human face and other body parts using Markov logic networks. By learning a set of probabilistic first-order logic rules that capture interactions between body parts under varying degrees of occlusion, and the relationship they share with the neighboring spatial windows. We can obtain a graphical model representation of these instances to facilitate inference. In this method, parts of human objects are detected individually using template matching. Another approach is to detect human objects in an image by combining the use of active contour models 'snakes', which detect human objects in an image, with a 2 layer feed forward back propagation neural network, to categorize the detected shape as human, or not. It was found that combining the neural network's output values with its confidence value provided a means of classifying unseen shapes into 'human' and 'non-human'.

## 1. Detecting Partial Occlusion of Humans by Using Snakes and Neural Networks.

This paper presents a technique for detecting human shapes in images, and for determining whether or not those human shapes are being partially occluded. An active contour [1], or 'snake', is used to detect and track objects in an image or sequence of images. When the snake has relaxed onto an object, that is, when its energy has been minimised, the contour's vector of (x, y) coordinates is re-represented using a novel encoding algorithm called the axis crossover representation. The resulting scale- and location-invariant vector can be used as an input pattern for neural networks. A feedforward neural network has been found to classify 90% of unseen human shapes correctly, when trained with both human and non-human crossover vectors. Furthermore, when the network is presented with unseen, partially occluded, human shapes, a measure of occlusion can be obtained by analysing the neural network's output values. This method is used to indicate a level of

occlusion for a given pose and, as such, could be used as a 'higher level cognitive process to control snakes.

### 1.1. Detecting Human Shapes with Snakes.

A snake is an energy minimising spline whose energy function can be tailored to detect specific features in images, allowing for the detection of particular classes of object. Since their original design [1], several variants have been developed for specific scenarios or for computational efficiency [4, 5, 6]. Despite these improvements, snakes have no knowledge of what they are detecting or tracking, and thus cannot categorise their own shape. This makes the snake less robust in complex visual environments, where it is often unknown what objects may enter and divert the snake's attention away from the target outline and onto other edges or noise. Without a mechanism for identifying what is being tracked, snakes have limited appeal to artificial intelligence applications.

A snake is initialised around the target human shape in an image by the user (Figure 1). The snake is then iteratively mapped onto the human outline in that frame by repeatedly minimising its energy function, resulting in a human shaped contour. Once relaxed, the snake can be moved into the next frame of the movie and mapped onto the human's new position using its relaxed position from the previous frame as a starting point.

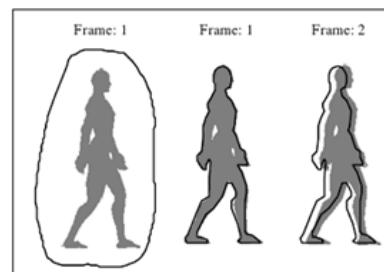
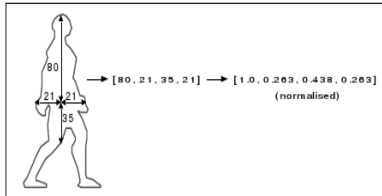


Figure 1: A snake relaxing on a human shape.

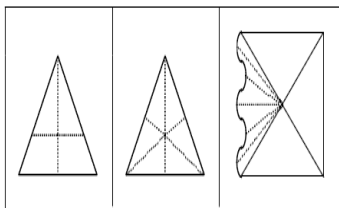
## 1.2 The Axis Crossover Representation.

Snakes are stored as a vector of (x,y) coordinates, from which a spline is constructed. This native representation is neither scale- nor location-invariant, so that similarly shaped contours may not have similar vectors. A representation of the snakes is presented which, in addition to being both scale- and location-invariant, can be customised so that it encapsulates salient features of the object class being detected.



**Figure 2: Encoding a human contour**

We must obtain an axis crossover representation of a snake, first centre of snake is calculated, by representation of a snake, the centre of the snake is calculated, using the mean of the snake's control point (x,y) coordinates. Axes are projected from the snake's centre point to its edges at specified angles. For example an equi-spaced 4-axis representation would grow axes at  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$  and  $270^\circ$  within the snake (Figure 2)..All axes are equi-spaced for simplicity, though axes could in theory be projected at irregularly-spaced angles. In tasks where only certain parts of an object need to be inspected by the neural network, for example production line assembly, projecting axes at particular angles during encoding may generate a more compact representation for the object class being detected.(Figure 3).



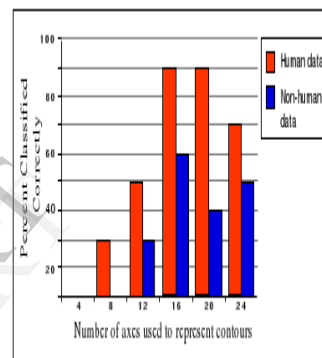
**Figure 3: The axis crossover representation**

After projection of axes at particular angles by using encoding we can generate more compact representation for object class which is being detected. The distance between snakes centre and edge from origin. It is stored in vector. The resulting vector, whose length equals the number of axes being used in the representation, is then normalised. This normalised vector can then be used as a training or test pattern to the neural network.

## 1.3. Neural Network Experiments

### 1.3.1 Categorisation.

It was necessary to determine whether or not a neural network could distinguish one group of crossover vectors (humans) from other groups of crossover vectors (non-humans). The vectors were evaluated with simple hidden layer back propagation networks, as the task, at this stage at least, was a categorisation of the vectors. The axis crossover representation allows for different numbers of axes to be used in the contour representation. Double output unit networks were used so that their categorisation confidence values, based upon the two output units' values, could later be analysed. The networks were trained with a range of different hidden layers, to allow the network with the optimal generalisation skills to be identified. The training set contained 150 human and 150 non-human shapes.



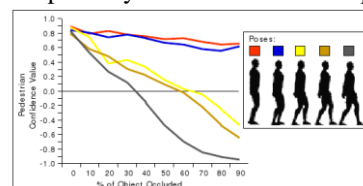
**Figure 4. Results from experiments using double output unit neural networks**

### 1.3.2 Confidence values.

Since the network has two output units, a confidence value of its classification can be obtained by differencing the two values.

### 1.3.3 Partial Occlusion.

Having identified that the axis crossover representation was a suitable means of encoding snakes for neural networks, it was interesting to test how robustly the networks behaved when presented with partially occluded human shapes.



**Figure 5: Confidence values for partially occluded human shapes.**

## 2. Detecting humans under partial occlusion using Markov Logic Networks.

There are 2 basic categories for human detection: holistic & window-based and part-based.

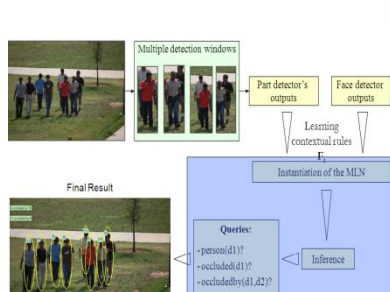


Figure 6. An overview of approach.

Logic (by itself) cannot accommodate for the probabilistic nature of the real world, and hence a more formal approach that accommodates the uncertainties of the visual scene is needed. Further by focusing on different aspects of human in isolation (within a single window), the existing works do not account for the information conveyed by the surrounding scene. Since human vision perceives the real world by associating a set of contextual constraints prevalent in nature.

### 2.1 Background on Markov Logic Networks.

Markov Logic Networks are one type of the unrolled graphical models developed in SRL to combine logical and probabilistic reasoning. In MLN, every logic formula  $F_i$  is associated with a nonnegative real-valued weight  $w_i$ .

Every instantiation of  $F_i$  is given the same weight. An undirected network, called a Markov Network, is constructed such that,

- Each of its nodes correspond to a ground atom  $x_k$ .
- If a subset of ground atoms  $x\{i\} = \{x_k\}$  are related to each other by a formula  $F_i$ , then a clique  $C_i$  over these variables is added to the network.  $C_i$

is associated with a weight  $w_i$  and a feature  $f_i$  defined as follows

$$f_i(x\{i\}) = 1, \text{ if } F_i(x\{i\}) \text{ is true,} \\ = 0 \text{ otherwise} \quad (1)$$

### 2.2 Approach for detecting partial occlusion by markov logic networks.

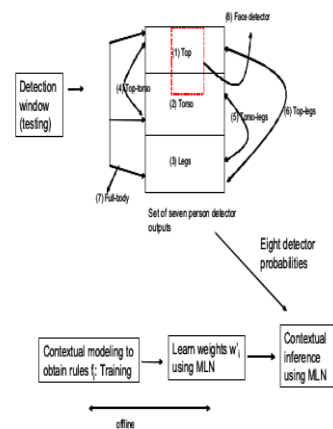


Figure 7. Partial least squares (PLS)-based dimensionality reduction

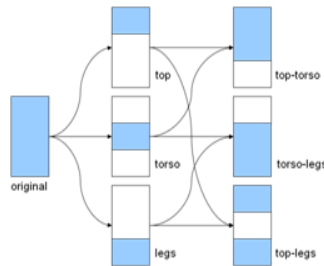
Given a set of  $N$  detectors corresponding to different body parts, and a set of detection windows in a close spatial neighbourhood, we propose the following:

- Learn the relation shared between the  $N$  detectors under varying degree of occlusion - Intra-window context
- Learn the relation of a window with the surrounding windows under visual uncertainties - Inter-window context
- Formulate the contextual information with a set of logic rules, and perform probabilistic inference within the framework of Markov logic networks.

### 2.3 Design of detectors.

- Exploit the use of more representative features to provide richer set of descriptors to improve detection results – edges, textures, and color.
- Consequences of the feature augmentation:
  - extremely high dimensional feature space (>170,000)
- number of samples in the training dataset is smaller than the dimensions.
- These characteristics prevent the use of classical machine learning such as SVM, but make an ideal setting for Partial Least Squares (PLS)\*.

**2.4 Part-based detectors.** To handle human detection under occlusion, our original detector is split into parts, then MLN is used to integrate their outputs.



**Figure 8: Partial based detectors**

**2.5 Detector – Partial Least Squares (PLS).**

PLS is a wide class of methods for modeling relations between sets of observations by means of latent variables. Although originally proposed as a regression technique, PLS can be also used as a *class aware* dimensionality reduction tool. By setting the dependent variable to a set of discrete values (class ids), we use PLS for dimensionality reduction followed by classification using a classifier in low dimensional space. The extracted feature vector is projected onto a set of latent vectors (estimated using PLS), then a classifier is used in the resulting low dimensional sub-space. PLS models relations between predictors variables in matrix  $\mathbf{X}$  ( $n \times p$ ) and response variables in vector  $\mathbf{y}$  ( $n \times 1$ ), where  $n$  denotes number of samples,  $p$  the number of features.

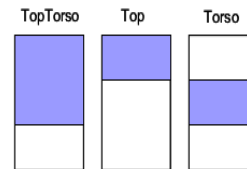
$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E}$$

$$\mathbf{y} = \mathbf{Uq}^T + \mathbf{f}$$

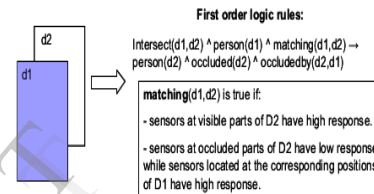
$\mathbf{T}$ ,  $\mathbf{U}$  are ( $n \times h$ ) matrices of  $h$  extracted latent vectors.  $\mathbf{P}$  ( $p \times h$ ) and  $\mathbf{q}$  ( $1 \times h$ ) represent the matrices loadings and  $\mathbf{E}$  ( $n \times p$ ) and  $\mathbf{f}$  ( $n \times 1$ ) are the residuals of  $\mathbf{X}$  and  $\mathbf{Y}$ , respectively. PLS method NIPALS (nonlinear iterative partial least squares) finds the set of weight vectors  $\mathbf{W}_{(p \times h)} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_h\}$  such that

**2.6. Context: Consistency between detector outputs.**

Each detector acts in a specific region of the body. One can look at the output of sensors acting in the same spatial location to check for consistency – similar responses are expected.



First order logic rules:  
 topTorso(d1) ^ top(d1) ^ torso(D1) → person(d1)  
 Detection window d1 topTorso(d1) ^ !top(d1) ^ !torso(d1) → !person(d1)



**Figure 9. Examples of Contextual modeling.**

**2.7 Inference using MLN\* - The basic idea.**

A logical knowledge base (KB) is a set of **hard constraints** ( $F_i$ ) on the set of possible worlds. Let's make them **soft constraints**: When a world violates a formula, It becomes less probable, not impossible. Give each formula a **weight** ( $w_i$ ) (Higher weight  $\Rightarrow$  Stronger constraint)

- Instantiation MLN is **template** for ground Markov nets
- Probability of a world  $x$ :
- Learning of weights, and inference performed using the open-source Alchemy system.

## Advantages of Using Markov Logic Networks.

- A contextual analysis is in charge of assessing levels of intra-object or inter-object relationship, ultimately integrated into a Markov logic network.
- 
- A formal model can be built using logic networks and we can perform inference on it.
- Also partial occlusion problems can be solved using probabilistic logic networks.

## Advantages of using Snake based approach.

- Snake is an energy minimizing, deformable spline influenced by constraint and image forces that pull it towards object contours. Snakes are greatly used in applications like object tracking, shape recognition, segmentation, edge detection, stereo matching.
- Snakes are autonomous and self-adapting in their search for a minimal energy state.
- They can be easily manipulated using external image forces.
- They can be made sensitive to image scale by incorporating Gaussian smoothing in the image energy function.
- They can be used to track dynamic objects in temporal as well as the spatial dimensions.

## Conclusion.

Importance of context, and the use of probabilistic interpretation of first-order logic to perform robust inference under visual uncertainties. Also by using snake based active contour approach, feedforward neural network can be trained to categorise them as either 'human' or 'non-human'. The axis crossover representation forms a scale- and location-invariant representation of the shape. The representation can be customised in terms of the number of axes used, allowing more detailed representations to be encoded

## Future Scope.

More robust network removing noise can be built using Markov Logic Networks.

## References.

- [1] M. Kass, A. Witkin and D. Terzopoulos, "Snakes, Active Contour Models. International", Journal of Computer Vision, (1988) , pp 321-331.
- [2] J. J. Little and J. E. Boyd. , "Recognizing people by their gait: The shape of motion. Vider", Journal of Computer Vision Research Vol 1 No 2, (MIT Press 1988) ,pp 2-32.
- [3] A. M. Baumberg and D. C. Hogg., "An efficient method for contour tracking using active shape models", University of Leed School of Computer Studies Research Report Series, Report 94.11,(1994).
- [4] D. J. Williams and M. Shah, "A fast algorithm for active contours and curvature estimation", CVGIP - Image Understanding 55 ,(1992), pp 14-26
- [5]. T. F. Cootes and C. J. Taylor., "Active shape models - Smart snakes". British Machine Vision Conference, Sept 1992,pp 276-285
- [6]. R. Curwen and A. Blake., "Dynamic contours: Real-time active splines",. In A. Blake and A .Yuille (Eds). Active Vision, (MIT Press 1992),pp 39-58.
- [7]. K. Tabb and S. George. , "Snakes and their influence on visual processing",. University of Hertfordshire Department of Computer Science Technical Report No, (Feb 1998)., pp 309
- [8]. W. Schwartz, A. Kembhavi, D. Harwood, and L. Davis., "Human detection using partial least squares analysis", In Proceeding of the International Conference on Computer Vision, 2009,pp 24-31.
- [9]. W. R. Schwartz, R. Gopalan, R. Chellappa, and L. S. Davis., "Robust human detection under occlusion by integrating face and person detectors", In International Conference on Biometrics , 2009 ,pp 970-979.
- [10] O. Tuzel, F. Porikli, and P. Meer, "Human detection via classification on riemannian manifolds. In IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp 1 -8.