# *Biologically-inspired robust motion segmentation using mutual information*

Article

Accepted Version

# Biologically-Inspired Robust Motion Segmentation using Mutual Information

Anna-Louise Ellis[a,], James Ferryman[a],

[a]*Computational Vision Group, School of Systems Engineering, University of Reading, Whiteknights, Reading, Berkshire RG6 6AY, UK*

**Abstract**

This paper presents a neuroscience inspired information theoretic approach to motion segmentation. Robust motion segmentation represents a fundamental first stage in many surveillance tasks. As an alternative to widely adopted individual segmentation approaches, which are challenged in different ways by imagery exhibiting a wide range of environmental variation and irrelevant motion, this paper presents a new biologically-inspired approach which computes the multivariate mutual information between multiple complementary motion segmentation outputs. Performance evaluation across a range of datasets and against competing segmentation methods demonstrates robust performance.

*Keywords:*
biologically-inspired vision, background modelling, segmentation, surveillance, performance evaluation

## 1. Introduction

The ability to extract objects of interest from video sequences, using detected motion, remains an active area of research within the computer vision

*Corresponding author; email: james@computer.org

community. The capacity to provide real-time segmentations - silhouettes and bounding boxes - of objects (especially pedestrian) assists in the tracking and reasoning of the behaviour. Surveillance scenes often contain change that may be inaccurately detected as object motion such as changes in lighting, periodic motion, moving shadows and reflections. In addition the quality of surveillance footage is often poor, and at a low resolution resulting in noisy motion and ghosts. An example of these challenges is shown in Figure 1The extraction of objects of interest is frequently tackled by removing all irrelevant pixels in each frame. This is referred to as motion segmentation. To date no segmentation algorithm is robust under all these conditions.

In this paper, we propose a new formulation of pixel-based foreground segmentation which is motivated by recent results in biological vision which exploit the mutual information between multiple segmentation channels. The paper is divided as follows. Firstly, Section 2 details the biological motivation and mapping to a combination of parametric background modelling approaches. This is followed in Section 3 by approaches to fusing the outputs of multiple segmentation algorithms and introduces the multivariate mutual information forumulation adopted in this work. In Section 4 the datasets, evaluation methodology and the results of experiments are presented before concluding in Section 5 with conclusions and recommendations for future research.

## 2. Biologically-Inspired Segmentation

The ability of primates to recognise objects of interest, regardless of illumination and background, drives much of the biologically inspired computa-


Figure 1: PETS 2009 dataset original frame annotated with automated visual surveillance challenges.

tional vision systems. A new biologically inspired vision system is introduced in this section that models current vision research which has not previously been examined by the computational vision community.

In Section 2.1 the model of primate vision conventionally accepted by the computer vision community is presented. Section 2.2 provides descriptions of state of the art biologically inspired computational vision systems that refer to this model. Section 2.3 progresses on to accounts of current published neuro-biological, physiological and psychological vision research and highlights descriptions of retinal functions, inputs to the ventral and dorsal streams, and ventral and dorsal stream behaviour that have not been considered in modelling primate visual systems in the computer vision community. Based on this, a new model of understanding is presented and the behaviours

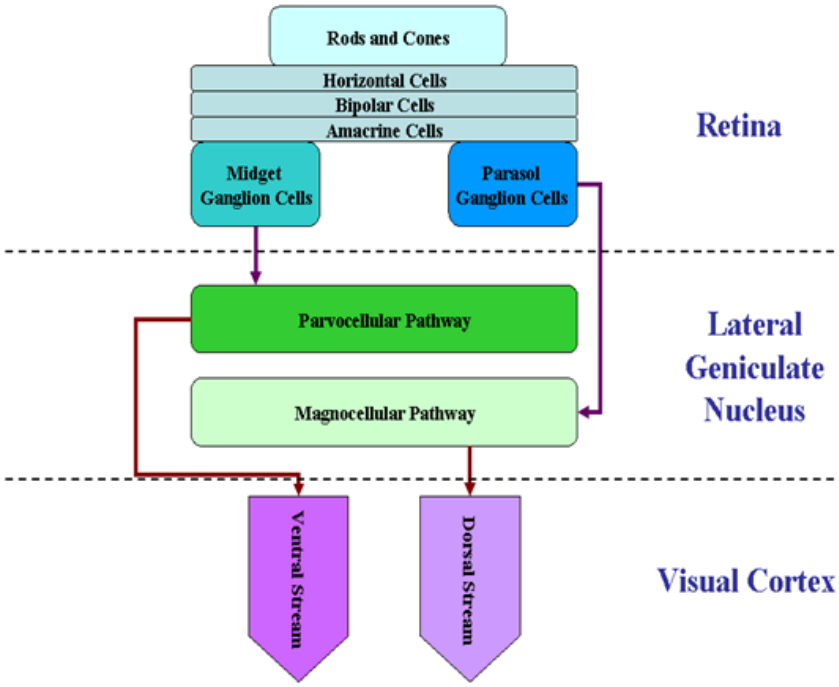


Figure 2: Model of traditional computational vision process

of these retinal functions are summarised.

### *2.1. Conventional Model of Primate Vision*

It is widely acknowledged that the rods and cones (photoreceptors) of the primate retina detect light and cells of the inner retina providing the initial stages of the visual processing. The retinal ganglion cells convey this information, via pathways in the lateral geniculate nucleus, to the ventral and dorsal streams in visual cortex. Figure 2 represents a model of these traditionally accepted components, frequently referred to in biologically inspired computational vision systems.

Within the retina, shown in Figure 2 as the blue area, the photoreceptor rod cells respond to achromatic brightness and the photoreceptor cone cells respond to short (blue), medium (green) and long (red) chromatic wavelengths. These nerve impulses are passed on to the network of horizontal, amacrine and bipolar cells, which provide cumulative information to retinal ganglion cells, shown in Figure 2 as the midget and parasol ganglion cells. The midget ganglion cells have been associated with providing chromatic information and parasol ganglion cells with luminance and contrast.

The lateral geniculate nucleus (LGN), illustrated as the green area in Figure 2, receives the assembled information from the ganglion cells, in the form of pathways. The parvocellular pathway is conventionally understood to receive information from the midget ganglion cells, and as such provides a means to direct colour information to the visual cortex. It is customary to describe the magnocellular pathway as a swiftly responsive structure, presenting the visual cortex with luminance and contrast information.

Finally, the visual cortex (VC), emphasised as the purple area in Figure 2,

includes two different streams: the ventral stream, associated with form, and the dorsal stream associated with motion.

### *2.2. Existing Bio-Inspired Computational Models*

(Mota et al., 2006) state that because bio-inspired vision models based on a vertebrates visual system are limited and require high computational cost, real-time applications are seldom addressed. As flies are capable of exploiting optical flow, which modelled by calculating the local image motion with Reichardt motion detectors (and referred to as Elementary Motion Detectors), they use this as inspiration and employ EMD as the first extraction primitive to characterise motion in a scene. Sequences are initially pre-processed by extracting edges within each frame using a Sobel edge extraction procedure. The Reichardt motion detector is then used to extract sideways moving features. Noise is removed from the resulting saliency map with a neural structure that allows the emergence of rigid bodies (independent moving objects in the scene) using "velocity channels". The technique is limited to greyscale images and suffers from being unable to identify to objects moving in parallel at the same speed. The system proposed by (Serre et al., 2007) follows on from their own theory of a feed forward path of object recognition that accounts for the first 100-200 milliseconds of processing in the ventral stream of primate visual cortex. It is based on Hubel and Wiesels findings in 1962 of a cats visual cortex (Hubel and Wiesel, 1985). Unlike the conventionally accepted chromatic input to the primate ventral stream, the approach takes a grey scale input and uses a set of scale and position-tolerant feature detectors, to simulate the properties of V1 and V4 (Figure 2 shows V1 and V4 within the ventral stream). A major limitation of the system

for real-time application is the processing speed which is limited by some of its modules that typically take tens of seconds, depending on the size of the input image. The authors have yet to address whether the recognition results obtained can be extended to the analysis of video. (Huang et al., 2011) offer an improvement on the system proposed by (Serre et al., 2007) focusing on improving the biological Standard Model Feature (SMF) for scene classification in a video surveillance environment. They develop a new energy computation component to improve SMF in occlusion and disorder cases as basic SMF models can only handle shift and invariance. An energy function is used in order that patches for saliency are not chosen randomly. An earlier analysis of energy density is used to conduct a local energy measurement after the initial basic feature extraction stage. Again the technique is limited to greyscale images. Using accounts of the primate visual cortex (Bayerl et al., 2007) have developed a neurodynamical computational vision model of motion segregation in the dorsal stream, as described in (Mishkin et al., 1983). The model includes two modules, corresponding to the primate visual cortex (highlighted as the purple area in Figure 2): V1 represents a motion hypothesis on the same scale of resolution on which it was detected, and V5 uses a coarser spatial resolution, where the accuracy of both location and velocity is reduced by a factor of five in accordance with physiological findings of Albright and Destmone in 1987 (Albright et al., 1987). The authors conclude that it is a step towards producing a biologically inspired model which may be capable of real-time computation. (Thriault et al., 2013) use a principle referred to as Slow Features Analysis (SFA) which bears foundations in neuroscience. SFA extract slowly varying features from a quickly varying input

signal. These features have been shown by (Thriault et al., 2013) to reveal sensible motion components correlated with specific semantic classes such as complex flame motion, waterfalls and fountains. As perceptions vary on a slower timescale compared to input signals from the environment, the SFA model learns to generate a slower, more invariant output signal. Temporal variations created by motion are minimised to in order to learn the stable representations of objects in motion. Motion features are defined by threading together short temporal sequences of SFA outputs. The motion features can be interpreted as spatio-temporal atoms describing the stable motion components inside a small space time window. Again this model relies on grey scale video as an input. The authors state that employing it for motion segmentation is a direction for future work. In (Yuen et al., 2009) features of objects are extracted "in a way similar to that of the ventral stream processing", referring to Diddays two visual stream model (Didday et al., 1975) published in 1975 and Mishkins slightly earlier publication than previously mentioned, with Ungerleider, in 1982 (Ungerleider et al., 1982). They use an RGB image input and proceed with a cortex-like centre surround operation in the spatiotemporal domain, by sub-sampling the image data into various spatial scales resulting in a set of images with horizontal and vertical scale reductions. Sets of features are extracted from the spatiotemporal stream and manipulated across various scales to detect those which locally stand out from their surround, similar to that of an edge detector. The authors state that due to the lack of a full understanding about the object recognition process in the visual cortex, the recognition mechanism that was implemented was a statistical classifier (SVM). In contrast Benoit et al. (Benoit et al.,

2010) recognise that consideration must be taken of the processing of the retinal signals that occur in primate vision, in order to assist further processing of that input, in a primate biologically inspired manner, in the visual cortex. They base their retinal architecture on Meads silicon model (Mead et al., 1988) albeit improved in terms of spatial and temporal properties. Their system contains two processing modules, one based on the retina for motion information extraction and the second representing a model of the V1 cortex area providing motion event detection. Their focus on the retinal processing includes passing information to their parvocellular channel model and magnocellular channel model from the midget ganglion cells model and parasol ganglion cells model respectively. These are shown in Figure 2 in green. This transformed information then is presented to their V1 model of the visual cortex. The system concentrates on using grey level image processing as the authors state the cell actions at the retinal level are unknown and further investigation is required to produce a better model.

### *2.3. Current Primate Vision Research*

Current neurobiology, visual neuroscience, physiology and psychology research provide descriptions of the input to the ventral and dorsal streams that have not been considered in computational vision systems modelling primate visual systems. Ganglion cell types other than midget and parasol cells also project to the LGN (Nieuwenhys et al., 2008; Dacey et al., 2000; Chatterjee and Callaway, 2003). (Dacey et al., 2000) provides a detailed description of these cell types, referred to as bistratified ganglion cells. They project their information to a further pathway in the lateral geniculate nucleus which is referred to as the koniocellular pathway (Nieuwenhys et al., 2008; Dacey

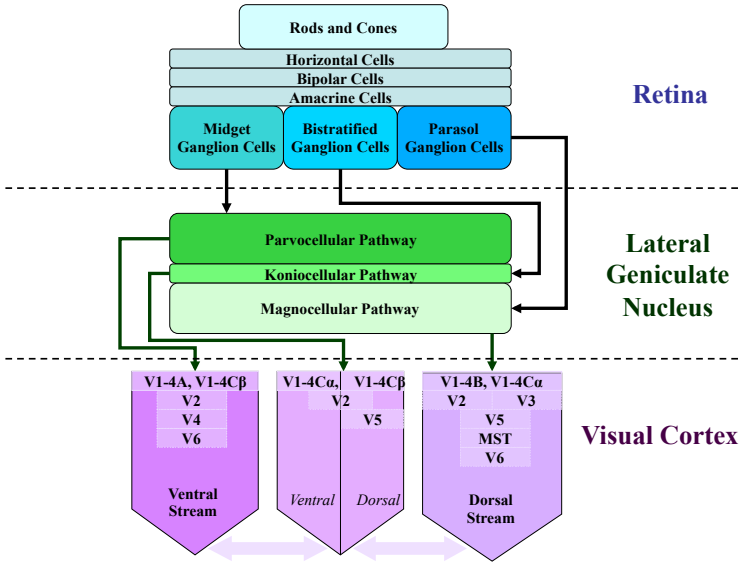


Figure 3: Model of recognised primate vision processes

et al., 2000; Chatterjee and Callaway, 2003; Hendry, 2000; Morand et al., 2000; Briggs and Usrey, 2011). A new illustration representing these recognised processes, including the bistratified ganglion cells and the koniocellular pathway is shown in Figure 3.

The retinal ganglion cells function in a distinct manner. The received wavelength signals can be used in the course of perceiving form or motion, independent of their role in the subjective experience of colour. Contradistinctively to the traditional accepted processes, the networked routing provides the midget cells with some contrast information (Kentridge et al., 2002), alongside the bistratified and parasol cells and therefore contrast information is present within both the ventral and dorsal streams. In addition prominent computation has been found to occur in the retina: the detection

of object motion while rejecting background motion (resulting from subtle eye movements) (Baccus et al., 2008) through specific interactions of amacrine and bipolar cells and presented to the ganglion cells. The koniocellular layer has been found to project to both the ventral and dorsal streams (Hendry, 2000). Finally recent primate vision research suggests there is communication between the dorsal and ventral streams, contrary to the traditionally accepted definitions used by the computer vision community of independent luminance motion information and colour object information occurring in the dorsal and ventral streams respectively. (McKeefry et al., 2010) ascertain that both luminance and chromatically defined motion is analysed in the dorsal stream and (Farivar et al., 2009) provide evidence that the dorsal stream participates in object recognition and some dorsal-ventral integration may be considered. Furthermore the study by (Zanon et al., 2010) states that the continuous interchange of information between the two streams is necessary and provides evidence that interaction is present in order to produce adaptive behaviour, for example, in order to elaborate the position in space and the shape of a 3D object. In effect the individual streams of information are weaved back together.

### *2.3.1. Ganglion Cells and the Lateral Geniculate Nucleus Pathways*

The current understanding of the individual behaviours of the three types of ganglion cells is described in detail in a vast array of vision research literature. These components in turn project this information to their respective lateral geniculate nucleus (LGN) streams, and these three streams have been ascertained by the neuroscience vision research community to have distinct behaviours and output. In this section brief descriptions of these components

and their respective LGN streams and behaviours are presented.

Parasol retinal ganglion cells receive many inputs and are responsively fast. They react to achromatic information and low contrast stimuli from the rods, and medium and long wavelength cones. They are unable to transmit information about wavelength independent of intensity and as such are not very sensitive to changes in colour. These cells are more sensitive to light since they are three times larger in diameter to the midget retinal ganglion cells. This information is relayed to the magnocellular pathway which is a fast system which contributes to the perception of luminance and motion derived from both achromatic and chromatic wavelengths, though it is unable to transmit any chromatic wavelength signals (Nieuwenhys et al., 2008; Kentridge et al., 2002; Dacey et al., 2000; Chatterjee and Callaway, 2003; Briggs and Usrey, 2011).

Midget retinal ganglion cells are involved in colour encoding. They react to chromatic information from the rods, and medium and long wavelength cones (green and red cones respectively) in the retina. They have low sensitivity because of their small receptive fields, but because of that they are densely packed and their resolution ability is higher. They respond weakly to changes in contrast unless that change is great. However, though these cells are found predominantly in the fovea of the retina, those located in the periphery show a non-opponent luminance response, indistinguishable from the parasol cells. The red/green colour opponent information and achromatic contrast detection information, provided by the synergy of the medium and long wavelength cones in the fovea, and those of the periphery able to distinguish brightness only, are relayed through the slow parvocellular pathway.

This pathway transmits information about long and medium wavelengths and fine detail. Motion perception information is presented but is far weaker than that of the magnocellular pathway and is dependent on the available chromatic contrast (Nieuwenhys et al., 2008; Kentridge et al., 2002; Dacey et al., 2000; Chatterjee and Callaway, 2003; Briggs and Usrey, 2011).

Bistratified retinal ganglion cells are involved in colour perception. They receive inputs from all rods and cone types but respond to rods and small wavelength cones (blue cones) 23 only. They have the lowest resolution ability, their density is extremely low and they have very large receptive fields. They have moderate to low spatial resolution and react to moderate changes in contrast. This information is projected to the koniocellular pathway which contributes to colour perception dependant on the small wavelength cone output and contributes to motion perception (Nieuwenhys et al., 2008; Kentridge et al., 2002; Dacey et al., 2000; Chatterjee and Callaway, 2003; Morand et al., 2000; Briggs and Usrey, 2011). Table 1 summarises the functions of the Magnocellular, Parvocellular and Koniocellular streams in the Lateral Geniculate Nucleus.

| | Magnocellular | Parvocellular | Koniocellular |
|---|---|---|---|
| Ganglion Cell | Parasol | Midget | Bistratified |
| Colour | No | Yes (R, G cones) | Yes (B cones) |
| Sensitivity to Contrast | High | Low | Moderate |
| Spatial Resolution | Low | High | Low |
| Temporal Resolution | Fast | Slow | Slow |

Table 1: Magnocellular, Parvocellular and Koniocellular Functions

### *2.4. Modelling the Lateral Geniculate Nucleus Pathways*

Recent research in (Zanon et al., 2010; Briggs and Usrey, 2011) have shown that the output of the magnocellular, koniocellular and parvocellular pathways provide mutual information to both ventral and dorsal streams, in order to supply the visual cortex with robust data about objects of interest and their location. Modelling this behaviour a form of multivariate mutual information is employed to enable the quantification of the amount of mutual information provided by the foreground segmentations of the modelling approaches described in this section. Background models may be seen to be analogous with the retinal suppression of global image motion as described by (Baccus et al., 2008). Using RGB colour space video sequences as input, the function of each of the parvocellular, magnocellular and koniocellular streams may each be modelled in a similar statistical manner. This section provides details of how these streams may be mapped to computational vision pixel-based background models.

#### *2.4.1. Parvocellular*

A background statistical model, which approximates behaviour of the parvocellular stream function (Kentridge et al., 2002), is able to distinguish between the brightness and its chromaticity of any one pixel, over time. This relates most closely to the method of (Horprasert et al., 1999). It is able to separate its wavelength (colour) information to include pixels with changes in luminance and contrast within its background model. The remaining pixels, with changes in colour and a limited amount of motion information. Figure 4 represents a graphical representation of the brightness distortion and chromaticity distortion in three dimensional RGB colour space. $E_i$ is the

initial (background) colour value for pixel $i$, and $I_i$ is the current colour value of the image. The line OE from the origin to $E_i$ represents the chromaticity line. Brightness distortion is a scalar value $\alpha$ and scales the point along OE where the orthogonal line from $I_i$ intersects OE. Chromaticity distortion $CD_i$ is the orthogonal distance between the observed colour and the line OE. The values for $\alpha$ and $CD$ are calculated for each of $N$ background frames

$$\alpha_i = \frac{\left(\frac{I_R(i)\mu_R(i)}{\sigma_R^2(i)} + \frac{I_G(i)\mu_G(i)}{\sigma_G^2(i)} + \frac{I_B(i)\mu_B(i)}{\sigma_B^2(i)}\right)}{\left(\left[\frac{\mu_R(i)}{\sigma_R(i)}\right]^2 + \left[\frac{\mu_G(i)}{\sigma_G(i)}\right]^2 + \left[\frac{\mu_B(i)}{\sigma_B(i)}\right]^2\right)}$$

where $\sigma_R(i)$, $\sigma_G(i)$ and $\sigma_B(i)$ are the standard deviation and $\mu_R(i)$, $\mu_G(i)$ and $\mu_B(i)$ are the means of the $i^{th}$ pixel's red green and blue values computed over $N$ background frames

$$CD_i = \sqrt{\left(\frac{I_R(i) - \alpha_i\mu_(i)}{\sigma_R(i)}\right)^2 + \left(\frac{I_G(i) - \alpha_i\mu_(i)}{\sigma_G(i)}\right)^2 + \left(\frac{I_B(i) - \alpha_i\mu_(i)}{\sigma_B(i)}\right)^2}$$

and then normalised to find a single threshold for all pixels

$$a_i = \sqrt{\frac{\sum_{i=0}^{N}(\alpha_i - 1)^2}{N}}$$

$$\widehat{\alpha}_i = \frac{\alpha_i - 1}{a_i}$$

$$b_i = \sqrt{\frac{\sum_{i=0}^{N}(CD_i)^2}{N}}$$

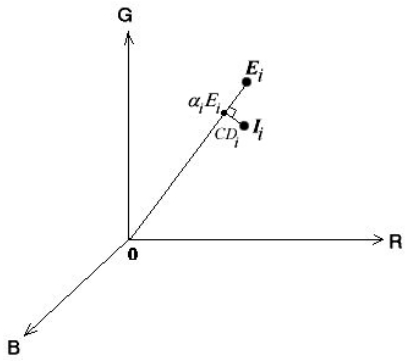


Figure 4: Graphical representation of the brightness distortion and chromaticity distortion in 3D RGB colour space.

$$\widehat{CD_i} = \frac{CD_i}{b_i}$$

The method constructs histograms of the normalised $\widehat{\alpha}$ and $\widehat{CD}$ values and takes a detection rate as input to automatically select thresholds. For segmentation, incoming pixels are used to calculate $\widehat{\alpha_i}$ and $\widehat{CD_i}$ values which are compared to those of the background model. The pixel classification for the $i$th pixel as defined by (Horprasert et al., 1999) is:

1. Original background if both $\widehat{\alpha_i}$ and $\widehat{CD_i}$ are within a threshold of those in the background model
2. Shadows or shaded background if the chromaticity $\widehat{CD_i}$ is within the threshold, but the brightness $\widehat{\alpha_i}$ is below
3. Highlighted background if the chromaticity $\widehat{CD_i}$ is within the threshold, but the brightness $\widehat{\alpha_i}$ is above

Figure 5: PETS 2009 dataset frame - BC algorithm approximating Parvocellular behaviour.

4. Moving foreground object if the chromaticity $\widehat{CD_i}$ is outside of the threshold

The resulting motion segmentation (Figure 5) from the original frame (Figure 1) show the model is able distinguish subtle differences in colour due to its motion sensitivity, but because of its motion sensitivity (due to both the temporal resolution and contrast sensitivity) parts of fluttering tape in the wind appear as foreground. Both the illumination and motion sensitivity provide the foreground segmentation with shadows.

### *2.4.2. Magnocellular*

A statistical model that presents foreground segmentation approximating behaviour of the magnocellular stream function is one that is able to provide high contrast information but does not distinguish between colour and its intensity. It must be sensitive to changes in luminance and motion (Kentridge et al., 2002). This most closely relates to the mixture model approach

of Stauffer and Grimson (Stauffer et al., 1999). Gaussian mixture models (GMM)s are able to model each component distribution as a soft classification; that is they are able to produce a distribution without specifying exactly what each cluster must represent. Yet as a whole, the mixture model covers the entire set of features (colour, brightness, intensity and luminance) that the data represents. The clusters formed represent more than one feature of information, and in this way the model becomes sensitive to contrast and motion. The resulting motion segmentations show that the model is able distinguish subtle differences in colour due to its motion sensitivity. Both the illumination and motion sensitivity provide the foreground segmentation with shadows. The recent history of a pixel is modelled by a mixture of K Gaussians (K usually varies from 3 - 5). The mixture is weighted by the frequency with which each of the Gaussians explains the background. The probability of observing a foreground pixel $x$ is:

$$P(x) = \sum_{j=1}^{K} w_j N(x, \mu_j, \Sigma_j) \tag{1}$$

where $w$ is the weight of the $Kth$ Gaussian distribution, $\mu$ is the mean, $\Sigma$ is the covariance matrix and $N$ is a multivariate Gaussian density function.

The resulting motion segmentation (Figure 6) from the original frame (Figure 1) show the model is able distinguish subtle differences in colour due to its motion sensitivity, but because of its motion sensitivity (due to both the temporal resolution and contrast sensitivity) parts of fluttering tape in the wind appear as foreground. Both the illumination and motion sensitivity provide the foreground segmentation with shadows.

Figure 6: PETS 2009 dataset frame - GMM algorithm approximating Magnocellular behaviour.

### *2.4.3. Koniocellular*

Similar to that of the Gaussian Mixture Model, the Colour Mean and Variance (CMV) algorithm, described in (Wren et al., 1997) captures the brightness, motion and colour information but only for a single colour channel. In this way the algorithm is able to provide foreground segmentation, similar to the behaviour of the koniocellular pathway (Kentridge et al., 2002). Encapsulating features in distinct distributions, using one independent channel value, removes the ability to capture some of the colour contrast information in the model, enabling any subtle changes to appear as foreground. The changes in the objective luminance of a pixel provide additional necessary motion information, but it is not as precise a measure as perceived brightness change and as such the motion sensitivity is coarser. The resulting motion segmentations show the model is able distinguish between some subtle differences in colour, however is of lower resolution and provides low

resolution shadow information from its motion sensitivity. CMV builds a statistical background model to represent an independent Gaussian distribution for each normalised colour channel (R,G,B) and a Gaussian distribution of the luminance (A) of each normalised pixel colour:

$$n(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{-(x-\mu)^2/2\sigma^2} \tag{2}$$

where x is the value of a single channel R, G, or B, or luminance (A), $\mu$ is the mean and $\sigma$ is the standard deviation of that channel. A pixel is classified as foreground if it is found to be more than 3 standard deviations of the R, G, B or A distributions.

The resulting motion segmentation (Figure 7) from the original frame (Figure 1) show the model is able to distinguish between some subtle differences in colour, but is of lower resolution (shown by the merging of moving objects in close proximity in Figure 7 and provides low resolution shadow information from its motion sensitivity.

## 3. Combining Algorithms

A number of approaches have been adopted in the literature for combining or fusing the outputs of multiple motion segmentation algorithms. (Martin et al., 2006) exploit optimal algorithm selection and key parameters tuning. A library of segmentation algorithms are fine tuned against predetermined ground truth images. The features extracted, alongside the optimal algorithm parameters, are saved as a case. They are ranked by a number of criteria. For each image a new case is created composed of a vector of image features, the chosen algorithm, and its optimised parameters. A multilayer

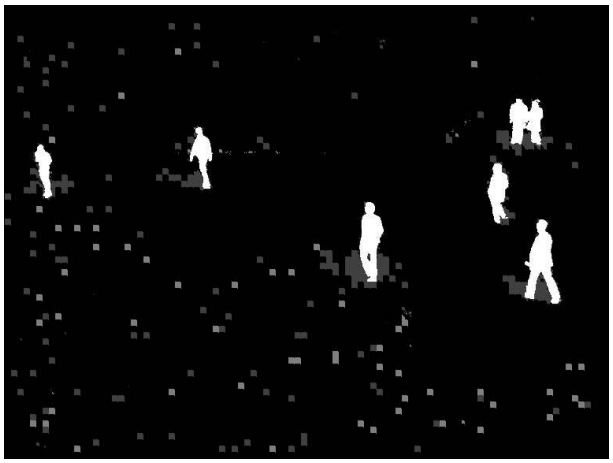

Figure 7: PETS 2009 dataset frame - CMV algorithm approximating Koniocellular behaviour.

perceptron (MLP) neural network is trained with this stored knowledge for algorithm selection. As the technique relies on predetermined ground truth this rules out generality. A Support Vector Machine (SVM), used by (Avidan et al., 2004), views the feature information as two sets of vectors in an n-dimensional space. It constructs a separate hyper-plane in that space which maximizes the margin between the two data sets. (Farmer et al., 2006) employ Expectation Maximisation (EM) as a fusion engine. Principal Component Analysis (PCA) is first applied to perform dimensionality reduction to improve the performance of EM and reduce the computational load. It is claimed that the approach applied to fusion of three popular optical flow algorithms (where the U and V component images are treated as image planes and EM applied to them) reduces the percentage of missing target pixels by 33%, although only one outdoor driving sequence has been used for evaluation. Boosting is an alternative. In (Zhou et al., 2004) each base classifier

must be trained, sequentially, using feature points that are weighted. The weight of a feature point is increased if a previous classifier misclassifies it. Once all of the classifiers are trained, their decisions can be combined through a weighted majority vote method or others. Popular boosting methods Adaboost and LogitBoost both have structural space, a cost function, and a selection algorithm. The AdaBoost algorithm minimises an upper bound of the target misclassification error, and LogitBoost minimises a negative binomial log-likelihood, as cost functions. Serre, Wolf, Bileschi, Riensenhuber and Poggio model a neurobiological design of a primate cortex (Serre et al., 2007). It is designed using hierarchical alternating layers of simple units and complex units. Simple units (16 Gabor filters for each layer) combine their inputs with a (bell shaped) tuning function to increase selectivity. Complex units pool their inputs (from the output of the previous Simple unit layer) through a MAX function. The image (grey scale only) is propagated through the hierarchical architecture. Standard Model Features (SMFs) are extracted from the complex units and classified using SVM or boosting (Gentle boosting providing the best performance). It was discovered that because there are variations in the amount of clutter and in the 2D transformations, it is beneficial to allow the classifier to choose the optimal features extracted from either the high or low level SMFs at a point in time, to improve the performance. A major limitation of the system in the use of real world applications remains its processing speed which is typically tens of seconds per image. (Jodoin and Mignotte, 2005) fusion of motion segmentation approach is based on a K-nearest-neighbour-based fusion procedure that mixes spatial and temporal data taken from two input label fields. The first one

is a spatial segmentation of a frame at time t which contains regions of uniform brightness while the second label field is an estimated version of the motion partition. The two segmentation maps are estimated separately with an unsupervised Markovian segmentation routine. The fusion occurs with an iterative optimization algorithm called Iterative Conditional Mode whose maximum local energy for each site, at each iteration, is obtained with a K-nearest neighbour algorithm.

Mazeed, Nixon and Gunn (Al-Mazeed et al., 2004), whose work is closest to the work described in ths paper, employ Bayes. Two background models are produced using a Mixture of Gaussians algorithm and a brightness and chromaticity algorithm referred to as Statistical Background Disturbance Technique (SBD). When the classifiers agree (pixel is foreground or background) a decision is set accordingly. When classifiers disagree, conditional probability for the chosen class by each class is calculated. The product of each class of conditional probabilities provide the parameters for the final decision

$$arg \max_{i \in \{1,2\}} p(x|w_{CLSF_i})P(w_{CLSF_i}) \tag{3}$$

where $w$ is a class of either a background (BG) or a foreground (FG) for the classifier $CLSF_i$. The maximum conditional probability for each classifier is used with the classifer's confidence measure $P(w_{CLSF_i})$ to find the decision for the algorithm. The main limitation of the approach is that it limited to combination of two classifiers and that the priors are calculated using an exhaustive search method based on the training data to obtain the optical values giving minimum classification errors.

While Bayesian inference, as well as other methods details above, have been exploited for classification in motion segmentation, application of mutual information to fuse multiple motion segmentation outputs has not been studied. The approach taken here in selecting mutual information as a method to combine multiple classifiers (the output from the LGN pathways) is threefold: Firstly, in the same way the recognised behaviours of the LGN pathways influenced the modelling of such, the identified interactions between these channels of visual information that occur in the visual cortex influenced the choice of mathematical approach we use to model such findings. Recent neurophysiological and vision research highlight that the output of all three LGN pathways is shared within the visual cortex (McKeefry et al., 2010; Farivar et al., 2009; Zanon et al., 2010; Briggs and Usrey, 2011). Indeed (Clery et al., 2013) state that when considering the encoding of visual information in the brain, the statistical independence between luminance and chromatic edges in natural scenes vary depending on the dataset of natural images used and "mutual information" may be found. These findings rule out choosing methods of combining classifiers where the classifiers are competing and a single classifier is found to be the "expert" at each instance for example Behaviour Knowledge Space (Raudys et al., 2003) and those such as the majority vote and K-nearest neighbour algorithm. As the information theory principle of mutual information measures the amount of information one random variable contains about another it is seemingly a sensible mapping to choose to model the neurophysiological and vision findings. Secondly, consideration is taken regarding the data used from a statistical view point. Multiple classifiers that produce probabilities as an output may be combined

using the product or average of the probabilities or the "Naïve Bayes" rule however these combiners require that the individual classifiers use mutually independent subsets of features (Kuncheva, 2001). This is not the case with the output from the LGN pathways as each pathway produces an interpretation of identical data that each is presented with. Mutual information may also be described as a technique that measures the mutual dependency of one random variable with another and it is certainly the case with the LGN outputs that there will be some commonality. In addition mutual information classifiers have been found to provide an objective solution (Hu, 2012). Finally, as the LGN pathways are modelled using real-time computational vision techniques, it is pertinent to choose a combining method such as mutual information which, unlike techniques such as boosting, requires no additional training on the data presented and may provide a fused result "on-the-fly".

### *3.1. Mutual Information*

In information theory the entropy of a discrete random variable $X$ is the measure of the amount of uncertainty associated with the value of $X$. Shannon entropy, denoted by $H$, of a discrete random variable $X$, includes a probability measure. If $p$ represents a probability mass function of $X$ then Shannon entropy can be described in terms of a discrete set of probabilities

$$H(X) \quad = \quad -\sum_{i=1} p(x_i) \log p(x_i) \tag{4}$$

Mutual information $I$ measures the amount of information that can be obtained about one random variable by observing another. Mutual information can be expressed as

$$
\begin{aligned}
I(X;Y) &= H(X) - H(X|Y) \\
&= H(Y) - H(Y|X) \\
&= H(X,Y) - H(X|Y) - H(Y|X) \\
&= H(X) + H(Y) - H(X,Y) \qquad (5)
\end{aligned}
$$

where $H(X)$ and $H(Y)$ are the marginal entropies, $H(X|Y)$ and $H(Y|X)$ are the conditional entropies, and $H(Y|X)$ is a measure of what $Y$ does not say about $X$. $I(X;Y)$ is non-negative. Mutual information is a well established technique for medical image registration of several modalities (Pluim et al., 2003; Cheah, 2012) due to its insensitivity to changes in lighting conditons ability to address a wide range of non-linear image transformations. It has also been shown to be well suited to registration of images of the same modality (Pluim et al., 2003).

Trivariate mutual information is described in various ways by authors of research literature with reference to both the definition and in the use of notation. Figure 8 provides examples of the assorted ways that (Pluim, 2003) discovered it had been defined and used in his survey of multivariate mutual information in terms of entropies. The darker shaded areas represent the mutual information in each case. (Pluim, 2003) asserts that a property of the definition of Figure 8a. is that it is not necessarily nonnegative. In Figure 8b. the deeper shaded middle section denotes that this area is counted twice.

Figure 9 provides examples of how the notation varies between authors. The diagrams labelled Figure 9a., Figure 9b. and Figure 9c. depict a bi-

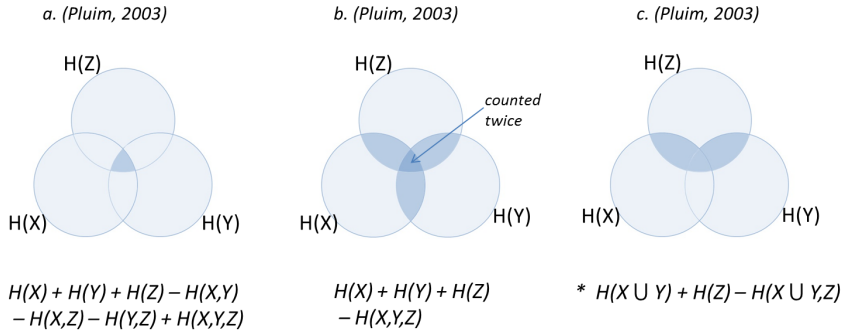


Figure 8: Different definitions of trivariate mutual information in terms of Shannon entropies. Each circles denote the entropy of an image. *Definition from (Pluim, 2003) text.

variate and two trivariate examples respectively and the notation to describe them given by (Studholme, 1996). He uses a '**;**' to separate the arguments for mutual information, while a '**,**' denotes a union of two variables. The notation used by (Pluim, 2003) differs in that to describe the same examples in the diagrams labelled Figure 9d., Figure 9e. and Figure 9f. '**,**' is used as the separator between the arguments and is not a union. Further to the differences found in literature in the notation, (MacKay, 2003) states that the term $I(X;Y;Z)$ is illegal. For clarity in this work the notation used throughout is that of (MacKay, 2003) which is consistent with (Studholme, 1996) and later authors (Escolano et al., 2009).

In this work the variables $X$, $Y$ and $Z$ are the probability in each LGN stream (parvocellular, magnocellular, and koniocellular) that a pixel is foreground. Here mutual information is used as a measure of the information or interaction between any two or all three LGN streams. To this end, CMI (Combined Mutual Informations) is defined as a linear combination of

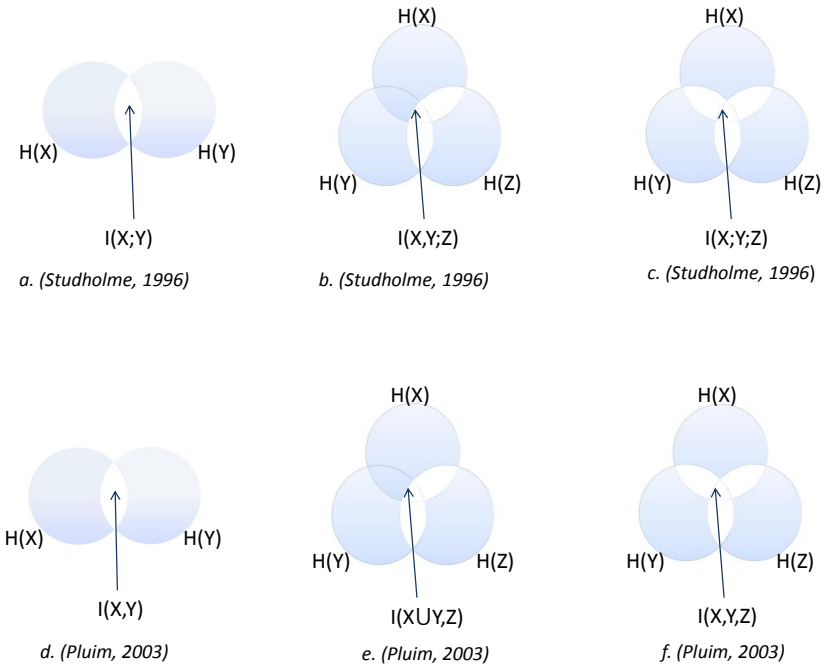


Figure 9: Differing notations describing the same mutual information examples

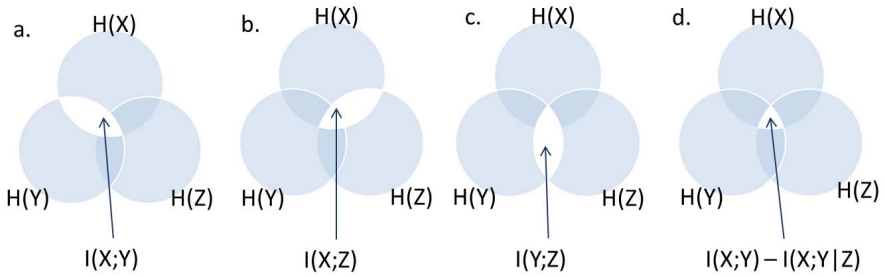


Figure 10: Bivariate and trivariate mutual information in terms of Shannon entropies

trivariate mutual information for all three LGN streams and bivariate mutual information for each pair of LGN streams such that none of the constituent entropies are counted twice. To avoid the use of any terms which could be considered illegal, the only trivariate mutual information used here will be of the form $I(X;Y|Z)$ which is the mutual information between $X$ and ($Y$ given $Z$) and is considered a legal term (MacKay, 2003).

Bivariate mutual informations are $I(X;Y)$, $I(X;Z)$ and $I(Y;Z)$ (Figure 10 a., b. and c. respectively) and are expressed in terms of Shannon entropies as

$$\begin{aligned} I(X;Y) &= H(X) + H(Y) - H(X,Y) \\ I(X;Z) &= H(X) + H(Z) - H(X,Z) \\ I(Y;Z) &= H(Y) + H(Z) - H(Y,Z) \end{aligned} \tag{6}$$

Trivariate mutual informations are $I(X;Y|Z)$, $I(X;Z|Y)$ and $I(Y;Z|X)$. In

terms of Shannon entropies $I(X;Y|Z)$ is defined as

$$I(X;Y|Z) = -H(Z) + H(X,Z) + H(Y,Z) - H(X,Y,Z) \tag{7}$$

The quantity $I(X;Y) - I(X;Y/Z)$ is shown in Figure 10d. and may also be defined as

$$\begin{aligned} I(X;Y) - I(X;Y/Z) &= I(X;Z) - I(X;Z|Y) \\ &= I(Y;Z) - I(Y;Z|X) \end{aligned} \tag{8}$$

Therefore a consistent quantity $CMI$, with no overlapping entropies may be defined as

$$\begin{aligned} CMI &= I(X;Y) + I(X;Z) + I(Y;Z) \\ &- 2[I(X;Y) - I(X;Y|Z)] \end{aligned} \tag{9}$$

$CMI$ can thus be expanded to give

$$\begin{aligned} CMI &= I(X;Y) + I(X;Z) + I(Y;Z) - 2[I(X;Y)] \\ &+ 2[I(X;Y|Z)] \\ &= -I(X;Y) + I(X;Z) + I(Y;Z) \\ &+ 2[I(X;Y|Z)] \end{aligned} \tag{10}$$

which can be expressed in terms of Shannon entropies as

$$\begin{aligned} CMI &= -H(X) - H(Y) + H(X,Y) \\ &+ H(X) + H(Z) - H(X,Z) \\ &+ H(Y) + H(Z) - H(Y,Z) \\ &+ 2[H(X,Z) + H(Y,Z) - H(X,Y,Z) - H(Z)] \end{aligned} \tag{11}$$

and can be simplified as

$$CMI = H(X,Y) + H(X,Z) + H(Y,Z) - 2H(X,Y,Z) \tag{12}$$

Since

$$H(X) = -\sum_{i=1} p(x_i) \log p(x_i) \tag{13}$$

$CMI$ may be rewritten as

$$\begin{aligned} CMI &= -\sum_{x,y} p(x,y) \log p(x,y) - \sum_{z,y} p(y,z) \log p(y,z) \\ &- \sum_{x,z} p(x,z) \log p(x,z) + 2 \sum_{x,y,z} p(x,y,z) \log p(x,y,z) \end{aligned} \tag{14}$$

and yields an expected value over all possible instances of $X$,$Y$ and $Z$.

The quantities given below, that are summed to find CMI, exist at all

points $x, y, z$.

$$\begin{aligned} p(x,y)\log p(x,y) \\ p(x,z)\log p(x,z) \\ p(y,z)\log p(y,z) \\ p(x,y,z)\log p(x,y,z) \end{aligned} \tag{15}$$

The two variable quantities are each defined on a 2D grid and the three variable quantity is defined on the 3D space $(x, y, z)$. Hence $p(x, y, z) \log p(x, y, z)$ may have a different value at all points $(x, y, z)$ where as $p(x, y) \log p(x, y)$ is only defined on the $x, y$ grid and values at any point $(x, y)$ are the same for all $z$. It is therefore possible to define a quantity $pVC$ at each point based on the point wise constituents of CMI.

$$\begin{aligned} pVC &= -p(x,y)\log p(x,y) - p(y,z)\log p(y,z) - p(x,z)\log p(x,z) \\ &\quad +2p(x,y,z)\log p(x,y,z) \end{aligned} \tag{16}$$

This provides a nonnegative result and is referred to as the Visual Cortex (VC) model in the following text.

The approximated probability mass functions produced by respectively the GMM, Brightness and Chromaticity, and Colour, Mean and Variance algorithms provide the mutual information required to produce silhouettes of objects of interest. For Brightness and Chromaticity, the probability that a pixel is foreground (FP) may be computed as (see Section 2.4.1 for notation)

$$FP = \frac{p(1 - p(\widehat{CD}_i))p(\widehat{\alpha}_i)}{p(\widehat{\alpha}_i)} \tag{17}$$

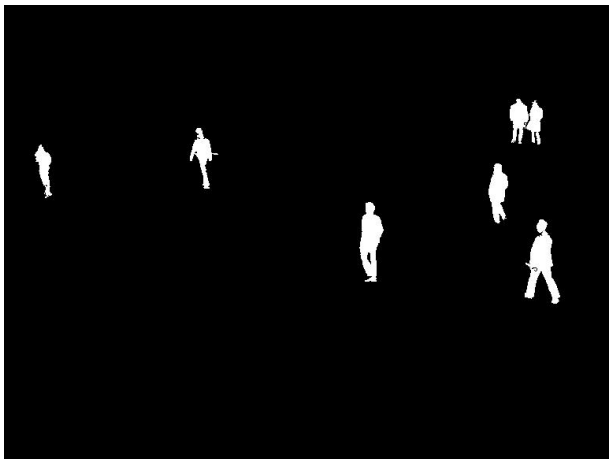

Figure 11: PETS 2009 dataset frame - resulting segmentation using the VC model

For Colour, Mean and Variance, the probability that a pixel is foreground (FP) may be computed as follows:

$$FP = p(R_i \cup G_i \cup B_i \cup A_i) \tag{18}$$

The probability for the Gaussian Mixture Model may be computed as given in equation 1.

Figure 7 represents the classification by the VC model of foreground pixels (white) from the original frame in Figure 1

## 4. Experimental Results

### *4.1. Ground Truth*

#### *4.1.1. Sihouettes*

The binary silhouettes of both the MuHAVi and PAMELA datas were hand labelled for all frames. For MuHAVi, Manually Annotated Silhouette Data (MAS) consists of annotated footage of 5 action classes. They include

two different actors and two separate camera views. In this case the annotation consists of white silhouettes of the actors performing their actions on a black background.

### *4.1.2. Objects*

Each of the PETS2009 seven independent 2D camera views (views 1,3,4,5,6,7,8) and CAVIAR"Walk" and "Walk 2" sequences were ground truthed frame by frame using the Video Performance Evaluation Resource (ViPER-GT) ground truth tool (Mariano et al., 2002). The ground truth consists of bounding boxes that are created around the objects and the coordinate positions of these boxes within the scene are given in a ground truth XML file.

## *4.2. Background Learning*

Each of the three motion segmentation methods used to model the LGN pathways require an initial "learning" phase, where the algorithms produce a statistical interpretation of the initial scene. Visual surveillance scenes are frequently dynamic in nature and whilst lengthy "background learning" sequences may produce a better motion segmentation from each of the algorithms this is mostly not practical due to rapidly changing scenes. To capture a scene or "background" where there is little of interest happening it is prudent to use as short a number of frames as is possible when initialising each of the motion segmentation algorithms. With this in mind for all datasets and sequences the following initialisations to the algorithms were given. The BC algorithm was set to a "background run length" of 100 frames, the initial $a_i$ and $b_i$ calculations used 50 frames and the initial histograms were created with just 10 frames. The GMM in this case was set to three gaussians, had

a "background run length" of 100 frames and calculated Expected Maximisation (EM) from just 20 frames. The CMV algorithm initialised with 10 background frames. For all algorithms a weight of 0.0001 was set for the learning rate.

### *4.3. Datasets*

Four different datasets are used to test the performance of the proposed Visual Cortex model, the publicly available MuHAVi (Singh et al., 2010), CAVIAR, PETS2009 (Ferryman and Ellis, 2009), and the datasets produced for the Background Models Comparison (BMC) challenge (Vacavant et al., 2012).

The first dataset, MuHAVi (Singh et al., 2010), introduces the challenge of real night-time street lighting, street paving (reflective) and real high street surveillance camera footage (with glare and lare prominet shadows) to the motion segmentation algorithms. There is also some camouflage of individuals present, where the clothing and the background are similar in colour.

CAVIAR Walk 1 and Walk 2 indoor datasets include sunlight shining through large glass panels and producing variable lighting within an indoor scene, alongside intermittent and unpredictable shadows of the panel frames on the floor. Reflections appear intermittently on additional glass panels that reside inside the building, and sunlight reflects from these panels. Shadows are present when individuals walk through the scene and some camouflage is present with the clothing of certain individuals and the background.

The third dataset, (Ferryman and Ellis, 2009), comprises multi-sensor sequences containing crowd scenarios with increasing scene complexity. Dataset S2, used in this evaluation, addresses people detection and tracking. Spe-

cific challenges include occluding ,moving objects encompassing whole scenes; moving vegetation; vehicles; motion behind translucent windows; reflective surfaces; objects appearing both very large and close to the camera and small and in the far distance; lack of natural lighting to entire footage.

Finally, the BMC dataset consists of both synthetic and real world videos. The synthetic videos present a variety of cloudy, sunny, foggy and windy scenes with and without acquisition noise. The real world videos contain challenges such as outdoor scenes, lengthy videos, varying ground types, presence of vegetation, casted shadows and the presence of continuous flow of objects.

### *4.4. Evaluation Metrics*

Performance evaluation was based on Precision and F1 Score Metrics andthe framework by (Kasturi et al., 2009), a well established protocol for performance evaluation of object detection and tracking in video sequences. These metrics are formally used by the Video Analysis and Content Extraction (VACE) programme and the CLassification of Events, Activities, and Relationships (CLEAR) consortium.(Vacavant et al., 2012) provides details for the F-score and SSIM metric used for the Background Model Challenge dataset.

*Notation.*

- $G_i^t$ denotes $i^{th}$ ground-truth object in frame $t$; $G_i$ denotes the $i^{th}$ ground-truth object at the sequence level; $N_{frames}$ is the number of frames in the sequence

Figure 12: Datasets used. Top row: Four views from MuHAVi which contains sequences with realistic street scenes. Second row: Four example frames from CAVIAR Walk 1 (left two images) and Walk 2 (right two images) sequences. Third row: Four example frames from Background Model Challenge dataset which contains both synthetic and real videos. Fourth row:Four views from the PETS2009 dataset which contains a range of crowd-based scenarios.

- $D_i^t$ denotes the $i^{th}$ detected object in frame $t$; $D_i$ denotes the $i$th detected object at the sequence level

- $N_G^t$ and $N_D^t$ denote the number of ground-truth objects and the number of detected objects in frame $t$, respectively; $N_G$ and $N_D$ denote the number of unique ground-truth objects and the number of unique detected objects in the given sequence, respectively

- $N^i_{frames}$ refers to the number of frames where either ground-truth object ($G_i$ ) or the detected object ($D_i$ ) existed in the sequence

- $N_{mapped}$ refers to sequence level detected object and ground truth pairs, $N^t_{mapped}$ refers to frame $t$ mapped ground truth and detected object pairs

- $m_t$ represents the missed detection count, ($fp_t$ ) is the false positive count, $c_m$ and $c_f$ represent respectively the cost functions for missed detects and false positives, and $c_s = log_{10}ID - SWITCHES_t$

*4.4.1. Precision and F1 Score*

Pixel based metrics are computed from pixel counts that may be classified as true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). FP and FN refer to those that are misclassified as pixels belonging to the objects of interest (FP) or the background (FN) while TP and TN account for accurately classified pixels.

The precision of a silhouette is an important factor for the reasoning of behaviour using pose and gait techniques, and is found by:

$$Precision = 100 - \left[\left(\frac{FN + FP}{TP + FN}\right) \times 100\right] \tag{19}$$

The F1 score is a popular metric for evaluation of segmentation and represents a measure of the accuracy of an algorithm and is found by:

$$F1Score = \frac{2TP}{((TP + FN) + (TP + FP))} \tag{20}$$

*4.4.2. Sequence Frame Detection Accuracy (SFDA)*

SFDA uses the number of objects detected, the number of missed detections, the number of falsely identified objects, and the calculation of the spatial alignment between the algorithm's output for detected objects and that of the ground truthed objects. It is derived from a Frame Detection Accuracy (FDA) measure. The FDA is calculated using a ratio of the spatial intersection and union of an output object and mapped ground truthed objects

$$OverLapRatio = \sum_{i=1}^{N^t_{mapped}} \frac{|G^t_i \cap D^t_i|}{|G^t_i \cup D^t_i|} \tag{21}$$

$$FDA(t) = \frac{OverlapRatio}{\left[\frac{N^t_G + N^t_D}{2}\right]} \tag{22}$$

$$SFDA = \frac{\sum_{t=1}^{N_{frames}} FDA(t)}{\sum_{t=1}^{N_{frames}} \exists\,(N^t_G \vee N^t_D)} \tag{23}$$

For this study although the annotation of the ground truth was challenging, an overlap threshold of 100 percent for the intersection over union scores, was used.

For both detection and tracking metrics in the following descriptions the accuracy metrics provide a measure of the correctness of the detections or tracks. The precision metrics provide the measure of, in the instance where there has been a correct detection or track, how close to the ground truth that detection or track may be.

#### *4.4.3. Multiple Object Detection Accuracy (MODA)*

MODA is an accuracy measure that uses the number of missed detections and the number of falsely identified objects. Cost functions to allow weighting to either of these errors are included, however for the sake of both PETS 2009 evaluations they were equally set to 1.

$$MODA = 1 - \frac{c_m(m_t) + c_f(f_{p_t})}{N_G^t} \tag{24}$$

#### *4.4.4. Multiple Object Detection Precision (MODP)*

MODP gives the precision of the detection in a given frame. Again, with this metric, an overlap ratio is calculated as previously defined in (1), and, in addition to a count of the number of mapped objects, the MODP is defined as:

$$MODP(t) = \frac{OverLapRatio}{N_{mapped}^t} \tag{25}$$

### *4.5. Results*

#### *4.5.1. MuHAVi*

The three individual segmentation algorithms and Visual Cortext algorithm were evaluated on the MuHAVi dataset against ground truth using the Precision and F1 Metrics. Comparisons are then made frame by frame between the algorithms resulting silhouette and the ground truth. True positive, false positive, true negative and false negative pixels are counted for each frame. Figure 13 shows the robust nature of the Visual Cortex model, respectively for F1 score (14) and Precision (13), using the mutual information of the three LGN pathways, in comparison to their independent performances.

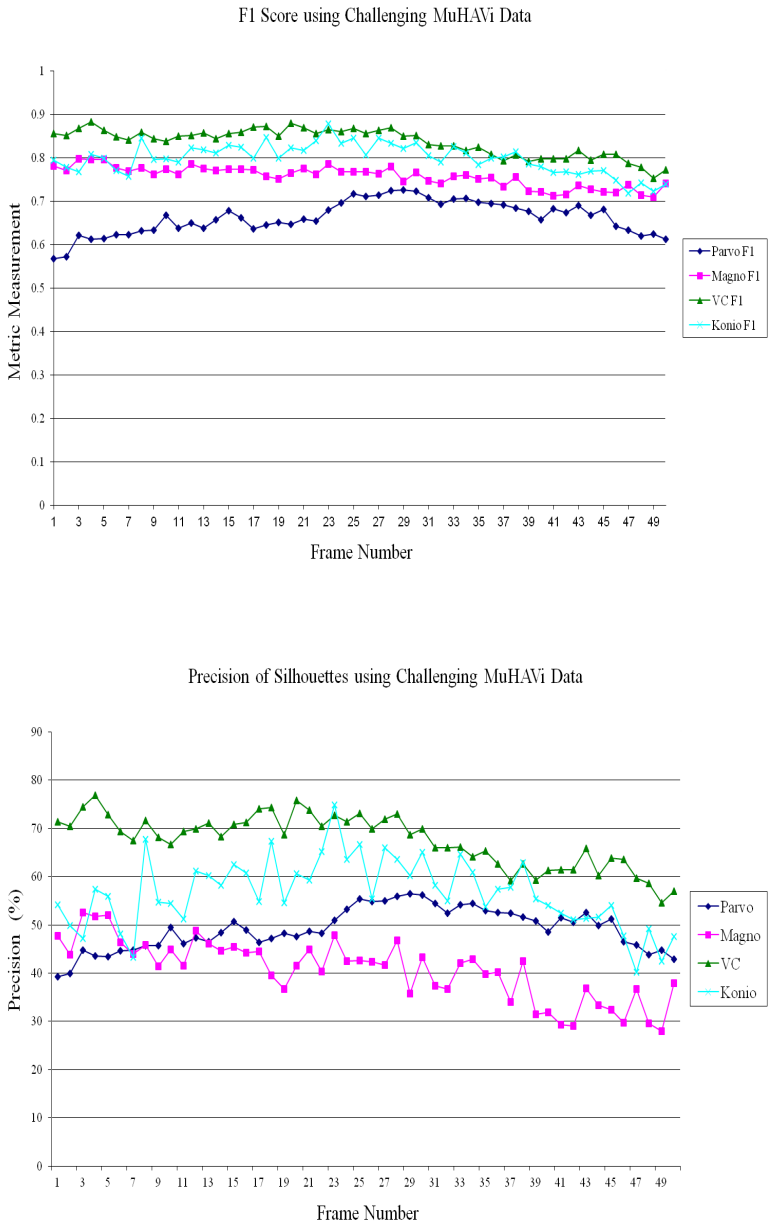


Figure 13: Accuracy (top) and precision (bottom) of the silhouettes produced by the independent LGN pathways versus the mutual information of the VC model on the challenging MuHAVi dataset.

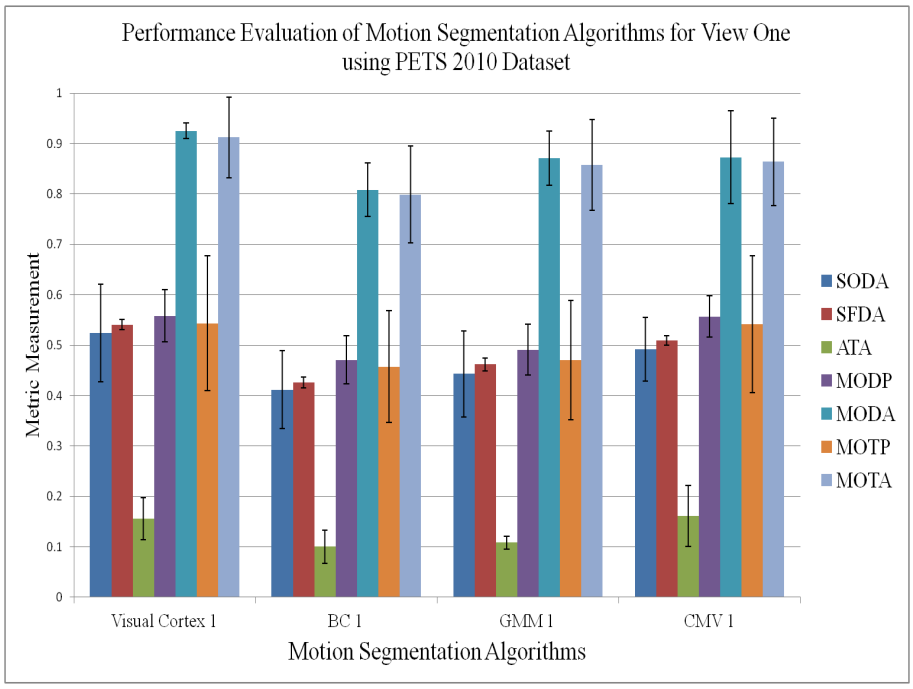


Figure 14: Performance of Visual Cortex and individual motion segmentation algorithms for view one of PETS2009 dataset.

### *4.5.2. PETS2009*

The next set of evaluations show comparisons of the performance of individual motion segmentation algorithms against the Visual Cortex model for the PETS2009 dataset. Figure 14 represents the evaluation results for sequence S2.L1, at time sequence 12.34, for the first camera view. and illustrates the superior performance of the Visual Cortex model, in comparison to the established motion segmentation algorithms, for the detection of objects within the surveillance scene. Every object detection metric, SODA, SFDA, MODA and MODP evaluates the Visual Cortex model (VC) as the best in

performance for its criteria, with the detection precision (MODP) metric proving the performance of the CMV algorithm as equal to that of the Visual Cortex model. Referring to the MOTA tracking metric, further analysis of Figure 14 demonstrates the increase in performance in tracking accuracy using the Visual Cortex model as the motion segmentation algorithm base for the tracker.

Next, to assess robustness in real world scenarios the Kanade-Lucas-Tomasi (KLT) tracking algorithm (Tomasi and Kanade, 1991) was used with individual sets of motion segmentation silhouette results using the PETS2009 dataset to produce tracking results, and in turn 2D bounding box coordinate positions and unique identifiers for each object for view one of the PETS2009 dataset. The performance evaluation results of the PETS 2009 and PETS2010 workshops (Ellis et al., 2010) were used to enable the comparisons. The SODA, SFDA, MODA and MODP metrics are relevant to the evaluation of the motion segmentation algorithms of the workshop's participating authors systems in addition to that of the Visual Cortex model. A summary of their motion segmentation/object detection techniques follow in order that comparisons may be drawn:

(Arsic et al., 2009) employ a multi-layer homography, which is capable of creating a three dimensional representation of the scene. Homography frameworks rely on the fusion of previously segmented foreground regions visible from multiple views. In the case of (Arsic et al., 2009) system, these foreground segmentations are produced by finding the median of pixel values and composing a reference image for simple background subtraction. Brightness invariance is achieved by normalised cross covariance when compared

with the reference image and contrast invariance is achieved using normalised cross-correlation. A graph cut optimisation algorithm is then optionally carried out to fill in small holes in foreground silhouettes.

(Breitenstein et al., 2009) presents a HOG object detector producing the input for the observation model of a particle filter, which includes not only the objects detected, but their confidence density of that detection (rep-resented as a colour heat map). Each object has its own particle filter initialised which includes its position and velocity. Bounding boxes are created by a boosted ensemble of weak classifiers employing colour histograms.

(Yang et al., 2009) utilises dynamic appearance models, using single Gaussians for foreground descriptions, and a Gaussian background model.

(Alahi et al., 2009) creates degraded foreground silhouettes from some binary silhouette image and its approximation, using rectangular and ellipse shapes. These then help form the input to a Multi-Silhouette Dictionary which is made up of atoms modelling the presence of individuals at give locations on an occupancy grid. The atoms are generated using homographies mapping points in a three dimensional scene to their two dimensional coordinates in the planar view.

(Bolme et al., 2009) approaches the challenge with the object detection filtering method Average of Synthetic Exact Filters which considers the entire output of the filter un-der a full convolution operation. He also uses a Viola and Jones cascade classifier with both visual and motion features used for detection. The third detector he uses is based on the deformable parts model system.

(Ge et al., 2009) regard people in a crowd scene as a realisation of a

Marked Point Process. Each person is associated with a random mark that specifies their location and size within the frame. A binary foreground mask is obtained by an adaptive background subtraction method and is subjected to further morphological processing. This then becomes the input to the detector.

(Conte et al., 2010) utilise an adaptive background image difference algorithm to detect moving objects. In order to make the system robust in realistic environments this has been extended to included processes that handle illumination, camouflage detection, noise filtering, shadow filtering and reflection removal.

(Berclaz et al., 2009) employ an object detector that produces a probabilistic occupancy grid, using a set of prob-abilities of the presence of objects, at a discrete set of locations, at each time step. These objects are represented as cylinders that project to rectangles in the frame sequences.

Figure 15 shows that the Visual Cortex model outperforms the evaluation of the individual algorithms with respect to the accuracy of both the detection of the objects and the tracking, using view one of the PETS 2009 datasets and the SODA, SFDA, MODP and MODA metrics.

It should be noted that the accuracy of the tracking algorithm used improves with the accuracy of the segmentation. The precision of any single detected object in this case refers to the precision of the location of its bounding box enclosing the object, that the tracker has produced, and not the precision of the silhouettes previously measured. Note that the standard error of mean (SEM) error bars have been added to the performance evaluation results charts. These quantify how precisely the true mean is known, taking

into account both the standard deviation and the sample size. Looking at whether the error bars overlap, therefore enables comparison of the difference between the mean with the precision of those means. It is very important to note that if two SEM error bars do overlap, and the sample sizes are equal the difference is not statistically significant, however if two SEM error bars do not overlap no conclusions may be made about statistical significance.

It is clear that for this sequence, the systems described by (Breitenstein et al., 2009) performed strongly at multiple object detection and tracking, with (Yang et al., 2009) outperforming all others. However the Visual Cortex model provides a strong performance in object detection and outperforms Breitenstein's system for detection accuracy (MODA) using the Visual Cortex model motion segmentation algorithm alone. Most detection and tracking systems employ further processing filters after any initial segmentation to improve the motion segmentation quality. This is not the case with the Visual Cortex model. The tracking accuracy (MOTA) gained from using the Visual Cortex model is second only to the system produced by Yang. As both Breitenstein and Yang did not provide results for views 5,6, and 8 no further comparisons or analysis of robustness using these systems may be drawn. (Ge et al., 2009), (Berclaz et al., 2009) and (Conte et al., 2010) detection accuracy measures (MODA) also suggested a good performance for these particular areas, as do (Berclaz et al., 2009), (Conte et al., 2010), and AlahiOlasso (Alahi et al., 2009) for tracking accuracy (MOTA).

### *4.5.3. CAVIAR*

Two "Walk" sequences from CAVIAR were evaluated against using the SODA, SFDA, MODP and MODA metrics. The Visual Cortex model again

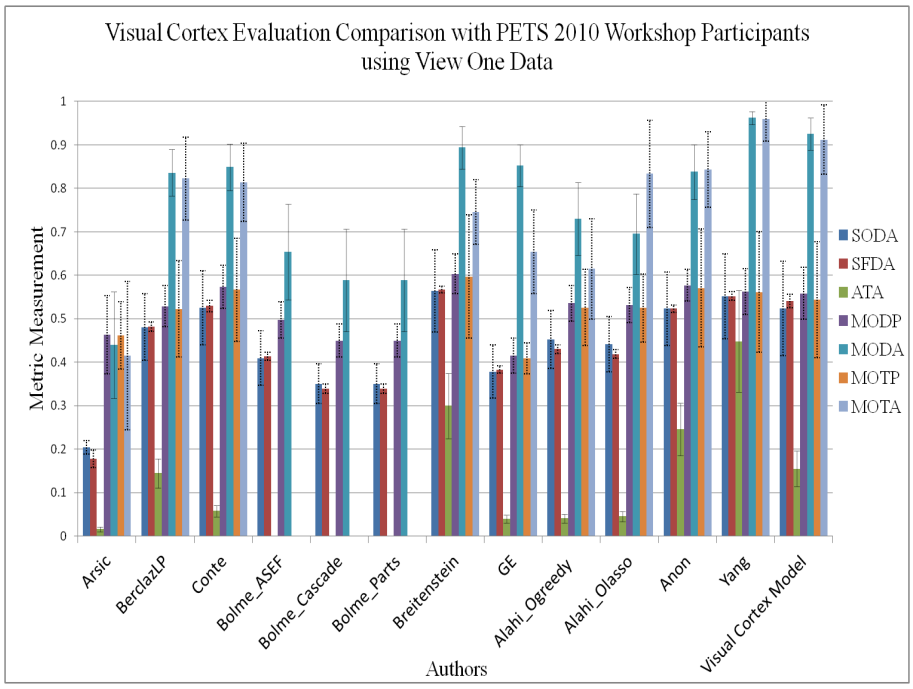


Figure 15: Performance of participating authors' systems, using CLEAR and VACE metrics for view one of PETS2009 dataset, mean SEM, N=109.

outperforms all three motion segmentation algorithms for each metric category despite the datasets being of a completely different nature to MuHAVi and PETS2009.

### *4.5.4. BMC dataset*

Finally, the synthetic and real datasets provided for this BMC special issue were evaluated and are shown in Figure 17. You can see from these that the VC model generally performs more robustly to the variety of sequences than published algorithms BC, GMM and CMV, in both synthetic and real world scenarios. The results for the synthetic videos show improvement on

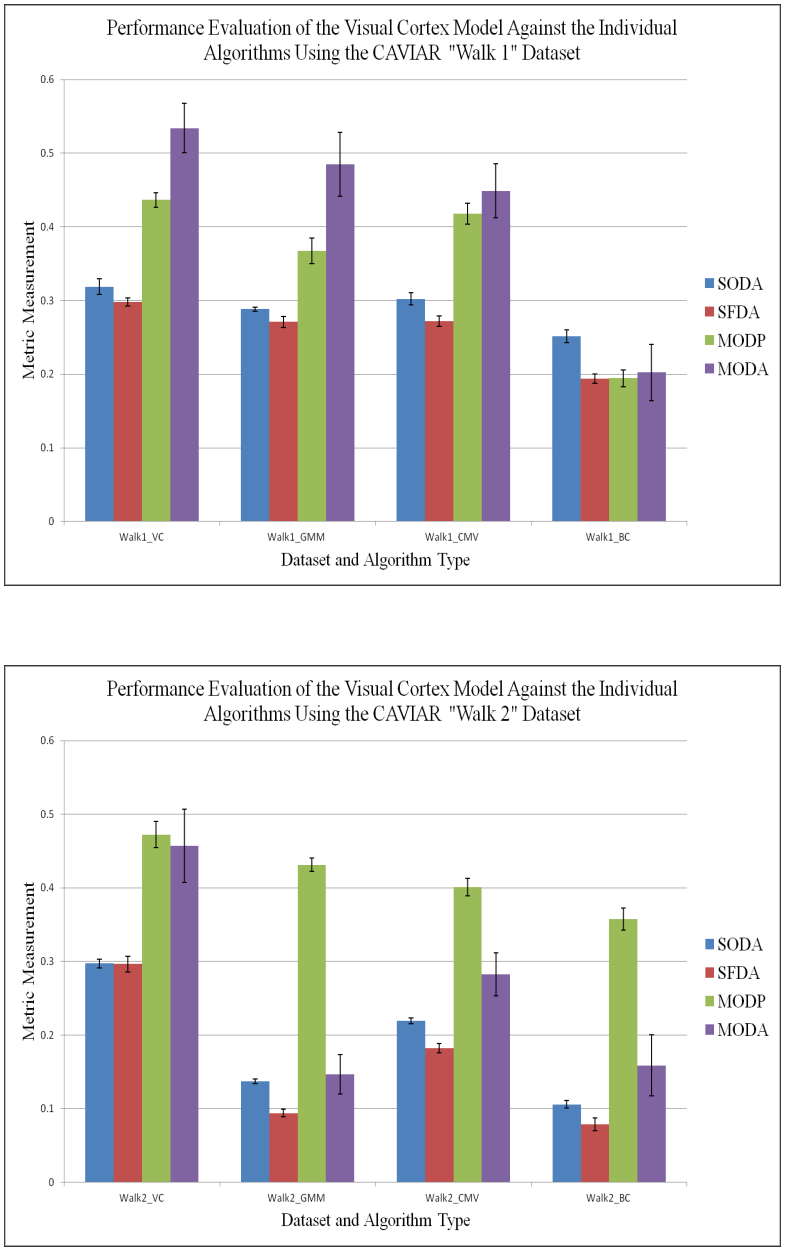


Figure 16: Comparing the Performance Evaluation of the Visual Cortex model with established motion segmentation algorithms using the CAVIAR (top) "Walk 1" and (bottom) "Walk 2" dataset, mean SEM, N=610.

the CMV, BC and GMM algorithms by employing the VC model, using both the F-Score and the SSIM metric as a measure, for all cases of videos tested. The individual algorithms however do not include any form of additional object recognition processing (and this is outside the scope of the biological model presented) that would distinguish between the cars travelling on the road and moving ground-truthed objects in the car park within the real world Video 1 scenario. In addition the VC model attempts to create a better silhouette of both the cars on the road and the ground-truthed cars in the car park than ones presented by the individual CMV, GMM and BC algorithms and as such is penalised by the pixel-based F Score metric for doing so. This is also the case for Video 8 where there is an additional flow of traffic to that which has been ground-truthed. It should be noted that pixel based metrics such as the F score can be heavily biased towards the larger moving objects within a frame when a video sequence contains more than one object and/or perspective plays a part. This bias is inherent in the results. The SSIM metric measures, for each real video sequence, highlight the visual structural (silhouettes) improvement gain made using the VC model, as opposed to the individual CMV, BC, and GMM algorithms.

The performance evaluation results of the Background Models Challenge workshop (Vacavant et al., 2012) participating authors' systems are shown in Figure 18. The VC model represents the results of motion segmentation only and does not include any additional processing techniques that may be added to assist in the elicitation of objects from the background. The VC model shows a noticeable comparison to all participating authors' background model systems with regard to the SSIM metric. The F-score metric highlights the

difficulty in producing a robust background model system for all scenarios, where generally the performances of each individual system appears to vary depending on the scenario it is presented with. A summary of the workshop's participating authors' techniques follow:

(Yoshinga et al., 2013) use illumination invariant local features and describe their distribution by Gaussian Mixture Models. The local feature has the ability to tolerate the effects of illumination changes, and the GMM can learn the variety of motion changes. Radial distances control the local feature and the localized regions focused by each pixel.

For (Shah et al., 2013) A Gaussian mixture model is used as a background basis and a new match function is used by computing separate variances for colour and intensity channels. For every foreground blob SURF features are matched and irrelevant features are removed using RANSAC sampling. The weight of winning Gaussian is increased a little for foreground blobs detected as paused objects. Automatic parameter adaptation is achieved using a fixed length sliding window to keep the most recent N frames in order to capture continuing statistical changes.

(Glazer et al., 2013) use one-class SVM classifiers to model the distribution of the background. Three levels of resolution are used: block, region and frame. Images are divided in to equal-sized blocks of pixels and the one-class SVMs are independently trained on each block to model its background distribution. Inter block relationships are used to refine the classification results at region level and at frame level an adaptive background method is used to re-initialise the model with regions considered to be part of the background.

(Tavakoli et al., 2013) introduce a method of estimating motion saliency

based on temporal cues obtained using frame de-correlation. Temporal salience maps are computed, presenting the amount of motion in a frame. Salient motion is assumed steady and the focus is on the detection of firm movements. Principal components analysis is applied for reconstruction whilst suppressing background clutter and noise.

(Guyon et al., 2013) use Robust Principal Components Analysis (RPCA) to separate moving objects from the background. The background sequence is then modelled by a low rank subspace, using a low-rank matrix factorization with iteratively reweighted least squares that can gradually change over time. The moving foreground objects constitute the correlated sparse outliers.

## 5. Conclusions and Future Work

This paper has presented a novel neuroscience inspired information theoretic approach to motion segmentation. In applying current neurological and physiological research in primate vision, a system has been created to improve the robustness of a multidimensional motion segmentation system. The major result found in this investigation is in using the current understanding of the primate visual system as inspiration and guidance for choosing both feature sets (the LGN pathways), and the means of fusing them (the Visual Cortex model), considerably improves the appearance of the obtained silhouettes, without the need for subjective parameter adjustments, or the use of arbitrary thresholds. This presents an advantage over established multidimensional models which frequently rely on decisions, based on some weighting, whether a feature set provides the correct segmentation. These techniques are burdened with adjusting parameters, which do not necessarily

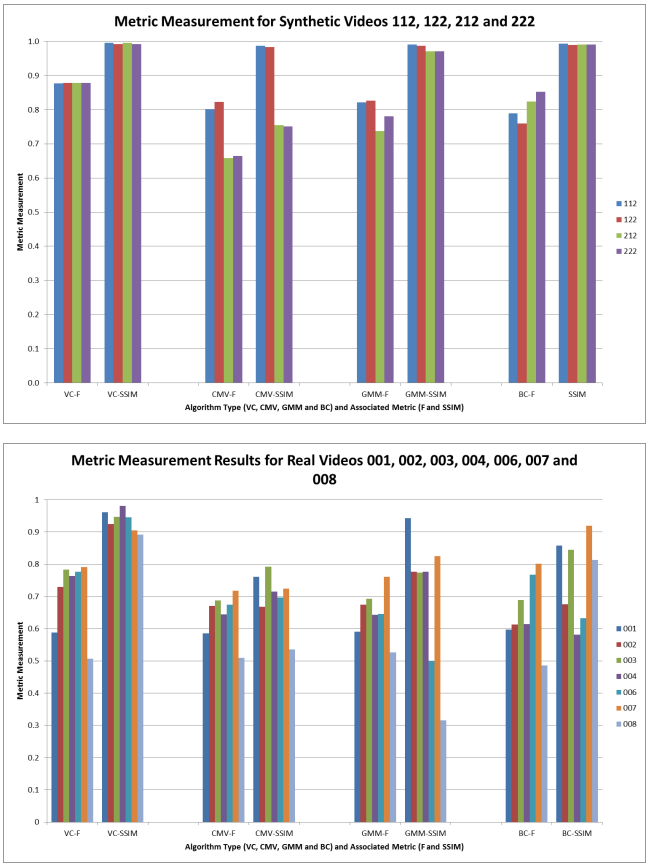


Figure 17: Comparing the performance of the Visual Cortex model with established motion segmentation algorithms using the BMC (top) synthetic and (bottom) real videos.

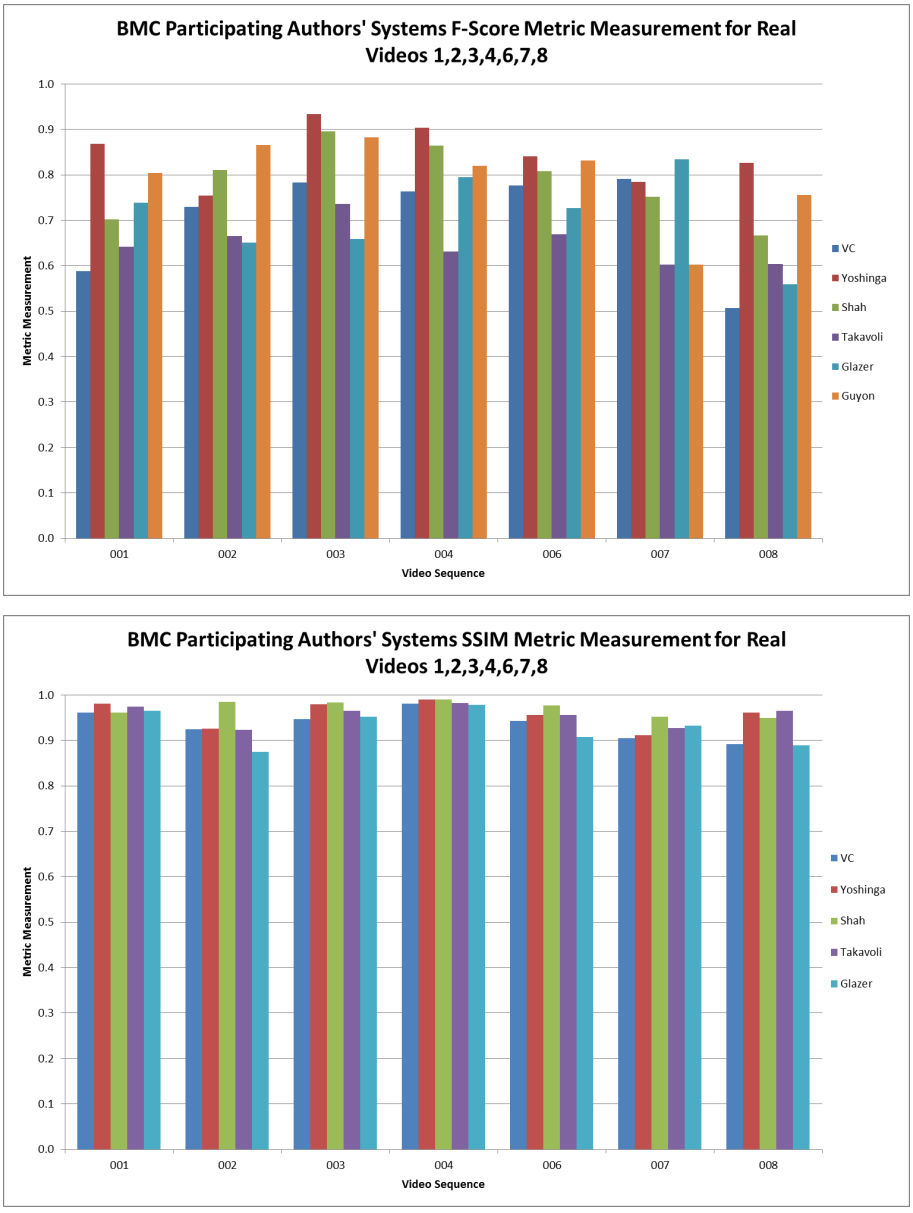


Figure 18: Comparing the performance of the Visual Cortex model with those of the participating authors' systems in the BMC challenge with the real videos dataset and F-Score(top) and SSIM (bottom) metrics.

provide the correct decision for all cases. This work has presented the performance evaluation of the biologically inspired motion segmentation system in challenging and diverse scenarios using a variety of evaluation metrics. In addition the evaluation results of state of the art automated visual surveillance systems have been presented to enable comparisons to be drawn. It shows that biologically inspired automated visual surveillance detection systems may be considered comparable to the current state of the art surveillance systems in detection and tracking. Existing real-time computational vision techniques have been exploited in the production of feature sets similar to that which the primate retina produces with a view towards real-time biologically inspired visual surveillance systems. The "reasoning" made within the visual cortex model employs a technique already well-established in the registration of medical images. It is envisaged that refining the LGN pathway approximations to closer representations of the biological system may result in robust performance beyond that of the current model. Further research into biologically guided object detection may provide a further processing model with a view to presenting robust object detection in addition to motion segmentation.

## Acknowledgements

This work was supported by the EC project ARENA Grant Agreement No. 261658. Any opinions expressed in this paper do not necessarily reect the views of the European Community. The Community is not liable for any use that may be made of the information contained herein.

The authors would like to thank M. J. Lally, School of Mathematical and

Physical Sciences, University of Reading, UK

A. Alahi, L. Jacques, Y. Boursier and P. Vandergheynst, Sparsity-Driven People Localization Algorithm: Evaluation in Crowded Scenes Environments, Proceedings of the Twelfth IEEE International Workshop on Evaluation of Tracking and Surveillance (PETS-Winter), 2009, DOI: 10.1109/PETSWINTER.2009.5399487

T. D. Albright, R. Desimone, Local Precision of Visuotopic Organization in the Middle Temporal Area (MT) of the Macaque, Experimental Brain Research, vol. 65(3), pp. 582-592, 1987.

A. Al-Mazeed, M. Nixon and S. Gunn, Classifiers Combination for Improved Motion Segmentation, Proceedings of International Conference on Image Analysis and Recognition, vol. 3212, pp. 363-371, 2004, ISBN: 3-540-23240-0

D. Arsic, A. Lyutskanov, G. Rigoll and B. Kwolek, Multi Camera Person Tracking Applying a Graph-Cuts Based Foreground Segmentation in a Homography Framework, Proceedings of Twelfth IEEE International Workshop Performance Evaluation of Tracking and Surveillance (PETSWinter), 2009, DOI: 10.1109/PETS-WINTER.2009.5399723

S. Avidan, Support Vector Tracking, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26(8), pp. 1064-1072, 2004, DOI: 10.1109/TPAMI.2004.53

S. A. Baccus, B. P. lveczky, M. Manu and M. Meister, A Retinal Circuit

That Computes Object Motion, The Journal of Neuroscience, vol. 28, pp. 6807-6817, 2008, DOI: 10.1523/JNEUROSCI.4206-07.2008.

P. Bayerl and H. Neumann, A Fast Biologically Inspired Algorithm for Recurrent Motion Estimation, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29(2), pp. 246-260, 2007, DOI: 10.1109/TPAMI.2007.24

A. Benoit, A. Caplier, B. Durette and J. Herault, Using Human Visual System modeling for Bio-Inspired Low Level Image Processing, Computer Vision and Image Understanding, vol. 114, pp. 758-773, 2010.

J. Berclaz, F. Fleuret and P. Fua, Multiple Object Tracking using Flow Linear Programming, Proceedings of the Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETSWinter), 2009, DOI: 10.1109/PETS-WINTER.2009.5399488

D. Bolme, Y.M. Lui, B. Draper and J. Beveridge, Simple Real-Time Human Detection using a Single Correlation Filter, Proceedings Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETSWinter), 2009, DOI:10.1109/PETS-WINTER.2009.5399555

M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier and L. van Gool, L., Markovian Tracking-by-Detection from a Single, Uncalibrated Camera, Proceedings of the Eleventh IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, pp. 7178, 2009.

F. Briggs and W. M. Usrey, Corticogeniculate Feedback and Visual Pro-

cessing in the Primate, The Journal of Physiology, vol. 589(1), pp.33-40, 2011.

S. Chatterjee and E. M. Callaway, Parallel Colour-Opponent Pathways to Primary Visual Cortex, Nature, vol. 426, pp. 668-671, 2003.

T. C. Cheah, Medical Image Registration by Maximizing Mutual Information Based on Combination of Intensity and Gradient Information, Proceedings International Conference on Biomedical Engineering, pp. 368-372, 2012.

S. Cheng, L. Xingzhi and S. M. Bhandarkar, A Multiscale, Parametric Background Model for Stationary Foreground Object Detection, IEEE Workshop on Motion and Video Computing, pp. 18, 2007.

D. Conte, P. Foggia, G. Percannella and M. Vento, Performance Evaluation of a People Tracking System on the PETS Video Database, Proceedings of the Thirteenth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, pp. 119-126, 2010, DOI: 10.1109/AVSS.2010.87

D. Dacey, Parallel Pathways for Spectral Coding in Primate Retina, Annual Review of Neuroscience, vol. 23, pp743-775, 2000, DOI: 10.1146/annurev.neuro.23.1.743.

R. L. Didday and M. A. Arbib, Eye Movements and Visual Perception: A Two Visual Stream Model, International Journal of Man-Machine Studies, vol. 7, pp. 499508, 1975.

R. P. W. Duin, The Combining Classifier: To Train or Not to Train?, Proceedings of the Sixteenth International Conference on Pattern Recognition, pp. 765-770, 2002.

A. Elgammal, D. Harwood D, L. Davis, Non-Parametric Model for Background Subtraction, Proceedings of the Sixth European Conference on Computer Vision, Part II, pp. 751767, 2000.

A. Ellis and J. Ferryman, PETS2010 and PETS2009 Evaluation of Results Using Individual Ground Truthed Single Views, Proceedings of the Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance, pp.135-142, 2010, DOI: 10.1109/AVSS.2010.89

F. Escolano, P. Suau and B. Bonev, Information Theory in Computer Vision and Pattern Recognition, Springer, 2009, ISBN: 978-1-84882-296-2, DOI: 10.1007/978-1-84882-297-9

R. Farivar, O. Blanke and A. Chaudhuri, DorsalVentral Integration in the Recognition of Motion-Defined Unfamiliar Faces, The Journal of Neuroscience, vol. 29(16), pp. 5336 5342, 2009.

M. E. Farmer, X. Lu, H. Chen and A. K. Jain, Robust Motion-Based Image Segmentation using Fusion, Proceedings of International Conference on Image Processing, vol. 5, pp. 3375-3378, 2004, DOI: 10.1109/ICIP.2004.1421838

J. Ferryman and A. Ellis, A., PETS2009: Dataset and Challenge, Proceedings of the Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, pp. 1-6, 2009, DOI:10.1109/PETS-WINTER.2009.5399556

W. Ge and R. Collins, Evaluation of Sampling-Based Pedestrian Detection for Crowd Counting, Proceedings of the Twelfth IEEE International Work-

shop on Performance Evaluation of Tracking and Surveillance (PETSWinter), 2009, DOI: 10.1109/PETS-WINTER.2009.5399553

A. Glazer, M. Lindenbaum and S. Markovitch, One-Class Background Model, Proceedings of the 11th International Conference on Computer Vision, vol. I, pp. 301-307, 2013, DOI: 10.1007/978-3-642-37410-4_26

C. Guyon, T. Bouwmans and E. Zahzah, Foreground Detection Via Robust Low Rank Matrix Decomposition Including Spatio-Temporal Constraint, Proceedings of the 11th International Conference on Computer Vision, vol. I, pp.315-320, DOI: 10.1007/978-3-642-37410-4_28

K. Huang, D. Tao, Y. Yuan, X. Li and T. Tan, Biologically Inspired Features for Scene Classification in Video Surveillance, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 41(1), pp. 307-313, 2011, DOI: 10.1109/TSMCB.2009.2037923

S. H. C. Hendry, The Koniocellular Pathway in Primate Vision, Annual Review of Neuroscience, vol. 23(1), pp 127, 2000, DOI: 10.1146/annurev.neuro.23.1.127.

T. Horprasert, D. Harwood, D. and L. S. Davis, A Statistical Approach for Real-Time Robust Background Subtraction and Shadow Detection, Proceedings of the Seventh IEEE ICCV Frame-rate Workshop, pp. 1-19, 1999.

B-G. Hu, What are the Differences between Bayesian Classifiers and Mutual-Information Classifiers?, CoRR, vol. abs/1105.0051v2, 2011.

D. H. Hubel and T. N. Wiesel, Receptive Fields and Functional Architecture

of Monkey Striate Cortex, Journal of Physiology, vol. 196, pp.117-151, 1985.

P. Jodoin and M. Mignotte, Motion Segmentation Using a K-Nearest-Neighbor-Based Fusion Procedure of Spatial and Temporal Label Cues, Proceedings of the Second international conference on Image Analysis and Recognition, pp778-788, 2005, DOI:10.1007/11559573-95

P. KaewTraKulPong and R. Bowden, An Improved Adaptive Background Mixture Model for Real-Time Tracking with Shadow Detection, Proceedings of the Second European Workshop on Advanced Video Based Surveillance Systems, pp. 149-158, 2001.

K. Kasturi, D Goldgof, P. Soundararajan, V. Manohar, J. Garofolo, R. Bowers, M. Boonstra, V. Korzhova and J. Zhang, Framework for Performance Evaluation of Face, Text, and Vehicle Detection and Tracking in Video: Data, Metrics, and Protocol, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.31(2), pp. 319 336, 2009.

R. Kentridge, C. Heywood, and J. Davidoff, Color Perception, The Handbook of Brain Theory and Neural Networks, Second Edition, Part III: Articles, pp. 230 233, 2002, ISBN-10: 0-262-01197-2, ISBN-13:978-0-262-01197-6

D. J. C. MacKay, Information Theory, Inference and Learning Algorithms, First Edition, Cambridge Press, 2003, ISBN-10: 0521642981, ISBN-13: 978-0521642989

V.Y. Mariano, J. Min, J.H. Park, R. Kasturi, D. Mihalcik, D. Doermann and T. Drayer, Performance Evaluation of Object Detection Algorithms,

Proceedings of International Conference on Pattern Recognition, pp. 965-969, 2002.

V. Martin, M. Thonnat and N. Maillot, A Learning Approach for Adaptive Image Segmentation, Proceedings of the Fourth IEEE International Conference on Computer Vision Systems, pp. 40  40, 2006.

D. J. McKeefry, M. P. Burton and A. B. Morland, The Contribution of Human Cortical Area V3A to the Perception of Chromatic Motion: A Transcranial Magnetic Stimulation Study, European Journal of Neuroscience, vol. 31, pp.575584, 2010, DOI:10.1111/j.1460-9568.2010.07095.x

C. A. Mead and M. A. Mahhowald, A Silicon Model of Early Visual Processing, Neural Networks, vol. 1, pp.91-97, 1988.

M. Mishkin, L. G. Ungerleider, K. A. Macko, Object Vision and Spatial Vision: Two Central Pathways, Trends in Neuroscience, vol. 6, pp 414-417, 1983.

S. Morand, G. Thut, R. Grave de Peralta, S. Clarke, A. Khateb, T. Landis and C.M. Michel, Electrophysiological Evidence for Fast Visual Processing through the Human Koniocellular Pathway When Stimuli Move, Cerebral Cortex, vol.10(8), pp. 817-825, 2000.

S. Mota, E. Ros, J. Díaz, R. Agis and F. de Toro, Bio-inspired Motion-Based Object Segmentation, Proceedings of the Third International Conference on Image Analysis and Recognition, pp196-205, 2006, DOI:10.1007/11867586_19.

R. Nieuwenhuys, J. Voogd and C. van Huijzen, The Human Central Nervous System, Fourth Edition, Springer, 2008, ISBN: 978-3-540-346864-5

J. P. W. Pluim, J. B. A. Maintz and M. A. Viergever, Mutual-Information Based Registration of Medical Images: A Survey, In IEEE Transactions Medical Imaging, vol. 22(8), pp. 986-1004, DOI:10.1109/TMI.2003.815867, 2003.

T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber and T. Poggio, Robust Object Recognition with Cortex-Like Mechanisms, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29(3), pp. 411-426, 2007.

M. Shah, J. Deng and B. Woodford, Illumination Invariant Background Model using Mixture of Gaussians and SURF Features, Proceedings of the 11th International Conference on Computer Vision , vol. I, pp. 308-314, 2013, DOI: 10.1007/978-3-642-37410-4_27

A. Shimada, D. Arita and R. Taniguchi, Dynamic Control of Adaptive Mixture-of-Gaussians Background Model, IEEE International Conference on Video and Signal Based Surveillance, pp. 5, 2006.

S. Singh, S. Velastin and R. Hossein, Muhavi: A Multicamera Human Action Video Dataset for the Evaluaton of Action Recognition Methods, Proceedings Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 48-55, 2010, DOI:10.1109/AVSS.2010.63, 2010

C. Stauffer and W. Grimson, Adaptive Background Mixture Models for Real-Time Tracking, Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 23-25, 1999.

C. Studholme, D. L. G. Hill, and D. J. Hawkes, Incorporating connected region labelling into automated image registration using mutual information, Mathematical Methods in Biomedical Image Analysis, A. A. Amini, F. L Bookstein, and D. C. Wilson, Eds. 1996, pp. 2331, IEEE Computer Society Press, Los Alamitos, CA.

H. R. Tavakoli, E. Rahtu and J. Heikkilä, Temporal Saliency for Fast Motion Detection, Proceedings of the 11th International Conference on Computer Vision , vol. I, pp. 321-326, 2013, 10.1007/978-3-642-37410-4_29

C. Thriault, N. Thome, M. Cord, Dynamic Scene Classification: Learning Motion Descriptors with Slow Features Analysis, Proceedings of IEEE ComputerVision and Pattern Recognition, 2013, To be issued.

C. Tomasi and T. Kanade, Detection and Tracking of Point Features, Carnegie Mellon University Technical Report CMU-CS-91-132, 1991.

L. C. Ungerleider and M. Mishkin, Two Cortical Visual Systems, Analysis of Visual Behavior, pp. 549586, Cambridge, MIT Press, 1982.

A. Vacavant, T. Chateau, A. Wilhelm, and L. Lequivre, A Benchmark Dataset for Outdoor Foreground/Background Extraction, Proceedings of the 11th International Conference on Computer Vision, vol. I, pp. 291-300, 2012, DOI=10.1007/978-3-642-37410-4_25

C. R. Wren, A. Azarbayejani, T. Darrell and A. Pentland, Pfinder: Real-Time Tracking of the Human Body, International Conference on Automatic Face and Gesture Recognition, pp. 51-56, 1997, DOI: 10.1109/AFGR.1996.557243

J. Yang, Z. Shi, P. Vela and J. Teizer, J., Probabilistic Multiple People Tracking through Complex Situations, Proceedings of the Eleventh IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, pp. 7986, 2009.

S. Yoshinaga, A. Shimada, H. Nagahara and R.Taniguchi, Background Model Based on Statistical Local Difference Pattern, Proceedings of the 11th International Conference on Computer Vision, vol. I, pp. 327-332, 2013, DOI: 10.1007/978-3-642-37410-4_30

P. Yuen, A. Tsitiridis, K. Hong, T. Chen, F. Kam, J. Jackman, D. James and M. Richardson, A Cortex Like Neuromorphic Target Recognition and Tracking in Cluttered Background, Third International Conference on Imaging for Crime Detection and Prevention, IET Seminar Digests, vol. 2, pp.27, 2009, DOI:10.1049/ic.2009.0255

S. K. Zhou, B. Georgescu, D. Comaniciu, S. Jie, BoostMotion: Boosting a Discriminative Similarity Function for Motion Estimation, Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 1761-1768, 2006.

M. Zanon, P. Busan, F. Monti, G. Pizzolato and P. Battaglini, Cortical Connections Between Dorsal and Ventral Visual Streams in Humans: Evidence By, Brain Topography, vol. 22(4), pp. 307-317, 2010.