Mutual information for enhanced feature selection in visual tracking

Victor Stamatescu[a], Sebastien Wong[b], David Kearney[a], Ivan Lee[a] and Anthony Milton[a]

[a]University of South Australia, Mawson Lakes, South Australia, Australia;
[b]Defence Science & Technology Organisation, Edinburgh, South Australia, Australia

ABSTRACT

In this paper we investigate the problem of fusing a set of features for a discriminative visual tracking algorithm, where good features are those that best discriminate an object from the local background. Using a principled Mutual Information approach, we introduce a novel online feature selection algorithm that preserves discriminative features while reducing redundant information. Applying this algorithm to a discriminative visual tracking system, we experimentally demonstrate improved tracking performance on standard data sets.

Keywords: visual tracking, feature selection, *infomax space*, *mRMR*

1. INTRODUCTION

Visual tracking (or video tracking) is the process of detecting objects and tracing their positions over time in an image sequence. Discriminant tracking[1] is an approach to visual tracking that treats detection as a two class classification problem between the target object and its local background. Some challenging problems in visual tracking, which can lead to a loss of track, include target object occlusions and model drift[2] due to background clutter. An online feature selection scheme that limits the amount of assumed knowledge about each object can help to mitigate these problems. The adaptive approach for discriminant visual tracking presented by Collins et al.[1] proposed that the best features are those that best discriminate between the two classes. Using features extracted from combinations of RGB pixel values, they tested this hypothesis by implementing an online feature selection algorithm, coupling it with a mean shift tracking system.[3] Features were selected based on their two-class *Variance Ratio* (*VR*) scores, which were calculated from the log-likelihood ratio of the class-conditional feature response distributions.

An outstanding problem in selecting features using this heuristic approach is that some of the features can be highly correlated. This means that despite being discriminative, some level of redundancy is present among the features, which can also contain information about the class labels. If this is not properly accounted for in selecting the features then it can degrade overall performance in classification tasks[4–6] or in discriminant tracking.[7,8] In this paper we investigate this issue by applying the criterion of *minimal-redundancy-maximal-relevance* (*mRMR*),[5] which has been typically used for classification purposes, to online feature selection in discriminant visual tracking. The improvements to tracking robustness offered by information theoretic feature selection is demonstrated experimentally using the single object tracking algorithm of Collins et al.[1] Furthermore, these feature selection algorithms are also compared in the context of single object tracking using an existing multi-object tracking algorithm, the Competitive Attentional Correlation Tracker using Shape and Feature Learning (*CACTuS-FL*).[9]

This paper is organized as follows: section 2 provides an overview of the relevant information theory and its application in the visual tracking literature, section 3 describes the implementation of our novel feature selection scheme as well as the discriminative multi-object tracking system that is used to test it, in section 4 real-world data sets are used to quantify the improvement in tracking robustness that is provided by our new feature selection, and we conclude with a summary of our findings in section 5.

Further author information:
Victor Stamatescu: E-mail: Victor.Stamatescu@unisa.edu.au, Telephone: +61(0)883023781

2. FEATURE SELECTION USING MUTUAL INFORMATION

Feature selection serves as a pre-processing step in the detection stage of many visual tracking systems.[10] Its intended goals may include a reduction in the dimensionality of the data, removal of redundant features, and in the case of classification tasks, the retention of features that are relevant to class labels. This section outlines common feature selection approaches based on information theory and summarizes their existing applications to discriminant visual tracking.

2.1 Information Theory

Mutual Information has been used in feature selection to measure dependencies between features. By treating the response of two features as discrete random variables A and B, their *Mutual Information* is:

$$I(A;B) = \sum_{a \in A, b \in B} p(a,b) log_2 \frac{p(a,b)}{p(a)p(b)} \ , \tag{1}$$

where $p(a)$ and $p(b)$ are the marginal probability mass functions (PMFs) of the two feature responses, and $p(a,b)$ is their joint PMF. If the random variables A and B are independent, then $p(a,b) = p(a)p(b)$, so that $I(A;B) = 0$. Furthermore *Mutual Information* is non-negative ($I(A;B) > 0$) and symmetric ($I(A;B) = I(B;A)$). $I(A;B)$ is also called *information gain*, because it quantifies the reduction of the uncertainty in A by having knowledge of B, and vice-versa:[11]

$$\begin{aligned} I(A;B) &= H(A) - H(A|B) \\ &= H(B) - H(B|A) \ , \end{aligned} \tag{2}$$

where $H(A)$ is the *entropy*, which is a measure of the uncertainty in A:

$$H(A) = -\sum_{a \in A} p(a) log_2 \, p(a) \ , \tag{3}$$

and $H(A|B)$ is the *conditional entropy*:

$$H(A|B) = -\sum_{a \in A, b \in B} p(a,b) log_2 \, p(a|b) \ . \tag{4}$$

Furthermore, *Mutual Information* can be expressed using Kullback-Leibler (KL) divergence:

$$I(A;B) = KL[p(a,b)||p(a)p(b)] \ , \tag{5}$$

where $KL[p(x)||q(x)] = \sum_{x \in X} p(x) log_2 \frac{p(x)}{q(x)}$ is a metric for the distance between two distributions p and q. Finally, the *Symmetrical Uncertainty* (SU)[12] provides a normalized version of *Mutual Information*:

$$SU(A;B) = 2 \, [\frac{I(A;B)}{H(A) + H(B)}] \ , \tag{6}$$

such that SU has range $[0,1]$.

2.2 Information Theoretic Feature Selection

By framing discriminant visual tracking as a two-class (target/background) classification problem, the most discriminant subset of features $\mathbf{X}^*$ can be prescribed using the *infomax space*,[13] which has also been called *Max-Dependency*.[5] This is the subset of features that maximizes its *Mutual Information* $I(\mathbf{X};C)$ with the two class labels: background ($C = 0$) and target ($C = 1$)), where $\mathbf{X}$ is a subset of the K candidate (input) features X_k, $k = 1, ..., K$. Using Eq. 2, this may be viewed as minimizing the uncertainty about which class is responsible for the observed features $H(C|\mathbf{X})$, which in turn relates to the minimization of the Bayes classification error.[6, 13]

Given the difficulty in estimating the multi-variate joint PMFs required to compute $I(\mathbf{X}; C)$, a useful decomposition that approximates the *infomax space* is given by the *maximum marginal diversity* (*MMD*):[7,8,13]

$$
\begin{aligned}
I(\mathbf{X}; C) &\approx \sum_{k=1}^{N} I(X_k; C) \\
&= \sum_{k=1}^{N} \sum_{i=0}^{1} \sum_{x \in X_k} p(X_k = x; C = i) log_2 \frac{p(X_k = x; C = i)}{p(X_k = x) p(C = i)} \\
&= \sum_{k=1}^{N} \sum_{i=0}^{1} p(C = i) \sum_{x \in X_k} p(X_k = x | C = i) log_2 \frac{p(X_k = x | C = i)}{p(X_k = x)} \\
&= \sum_{k=1}^{N} \sum_{i=0}^{1} p(C = i) KL[p(X_k = x | C = i) || p(X_k = x)] \ ,
\end{aligned}
\tag{7}
$$

where N is the number of features to be selected. This shows that each *marginal diversity* $I(X_k; C)$ is the weighted average of the distance between the class-conditioned feature response distributions $p(X_k = x|C = i)$ and their mean $p(X_k = x)$, where the weights are given by the class priors $p(C = i)$. Hence the most discriminant features, according to *MMD*, are those whose class-conditioned feature response distributions are well separated from each other. Given the need to choose some N features, the *MMD* algorithm simply involves ordering features according to the *marginal diversity* of each feature, and then selecting the top N features from this list.

The underlying assumption in approximating $I(\mathbf{X}; C)$ by $\sum_{k=1}^{N} I(X_k; C)$ in Eq. 7 is that the *Mutual Information* between a new candidate feature and the set of previously selected features do not provide additional classification power.[4,6,13] While this assumption has been demonstrated to hold for features extracted using band-pass (e.g. Gabor[14] or wavelet) filters from natural images, it does not apply in general.[6]

To address this problem a number of heuristics exist that provide a closer approximation of the *infomax space* without explicitly computing its joint multivariate distributions. One such example is the computationally efficient *predominant correlation* filter of Yu & Liu.[15] This algorithm applies a threshold to feature-class *Symmetrical Uncertainty* (SU_c) values to select relevant features and then discards those considered to be redundant by comparing SU_c to their SU with other selected features. Another is the *mRMR*[5] forward search method, which is considered state of the art.[6] For a *first-order* incremental search, where a set containing $m - 1$ selected features ($\mathbf{S}_{m-1}$) already exists and a new m^{th} feature X_j from those remaining ($\mathbf{X} \backslash S_{m-1}$) is to be added, *mRMR* simultaneously maximizes the relevance of the selected features to a target class while minimizing redundancy between features:

$$
\underset{X_j \in \mathbf{X} \backslash S_{m-1}}{max} [I(X_j; C) - \frac{1}{m-1} \sum_{X_i \in S_{m-1}} I(X_j; X_i)] \ . \tag{8}
$$

Peng et al.[5] showed that *mRMR* and *infomax space* are equivalent for this type of *first-order* feature selection.

2.3 Related Work

Mutual Information has been used in online feature selection for discriminant visual tracking, however, to our best knowledge, this study is the first to implement *mRMR* in full for this purpose. For instance, Alvarez-Santos et al.[16] used the *minimal redundancy* aspect of *mRMR* to perform offline feature selection and applied this to a person-tracking mobile robot. Leung and Gong[17] used the *Mutual Information* between features to select, in an online manner, reliable features for tracking. In their particle filter approach to tracking multiple objects, Cui et al.[18] evaluated the discriminability of features by way of the *Mutual Information* between features and multiple class labels. Hong and Han[19] also used a particle feature approach in which feature weights were computed by maximizing the *Mutual Information* between the target model and query features. Mahadevan and Vasconcelos[7,8] used *MMD* to define bottom-up *saliency.* This approach provided an optimum way in which to select the most discriminant features from a set of *band-pass* filters.[6]

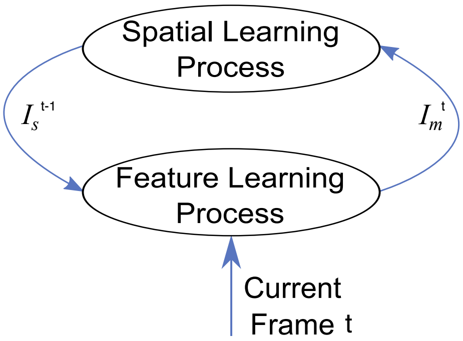

Figure 1: The feedback mechanism between online feature learning and spatial learning allows *CACTuS-FL* to autonomously focus attention to regions of locally correlated saliency to selected image features.[21] I_S^{t-1} is the learnt image of the object that is produced in the tracking stage of the algorithm by the Shape Estimating Filter (SEF). I_m^t is the measured, fused likelihood ratio (LR) map produced in the detection stage of the algorithm.

3. ONLINE FEATURE SELECTION USING MUTUAL INFORMATION

This section describes the implementation of two online feature selection schemes based on *MMD* and *mRMR*. These schemes are tested across three discriminant visual tracking algorithms: our implementation of the original algorithm by Collins et al.[1] (hereafter called *Collins*), a modified version that uses more advanced input features (hereafter called *Collins+*), and the multi-object tracking algorithm *CACTuS-FL*.[9] The *Collins* and *CACTuS-FL* algorithms are similar in terms of object detection, but differ in their general approach to object tracking.

3.1 Tracking

The *Collins/+* algorithms use a simple mean shift tracker[3] to carry out single object tracking. This takes the detection maps computed from N selected features as its input. The mean shift process is applied separately to each map and this generates, via gradient ascent, N estimates of the object position, which are combined to give the final estimate of 2D position in the current frame.

CACTuS-FL, on the other hand, was designed for multi-object tracking and uses multiple sub-trackers called Shape Estimating Filters (SEFs).[20] These sub-trackers operate simultaneously, competing with each other across the video frame and attempting to track all salient objects. The aim of this is to make the tracking more robust by *explaining away* all distracting clutter in the scene. Each SEF tries to track and detect a single object by learning its state model, which includes a probabilistic representation of the object shape. Single object tracking is carried out by each SEF using a factorised state-space made up of two dimensional PMFs that correspond to shape, position and velocity, all of which are linked through a hierarchical model. A Bayesian update scheme is applied to perform prediction and measurement of shape, position and velocity at each time-step of a video sequence.

3.2 Detection

CACTuS-FL performs object detection using an online feature learning process that is based on that of *Collins*. The *Collins* algorithm uses linear colour channel (RGB) combinations:[1]

$$\mathbf{X} \equiv \{w_1 R + w_2 G + w_3 B \mid w_1, w_2, w_3 \in [-2, -1, 0, 1, 2]\} \ , \tag{9}$$

where $\mathbf{X}$ is the set of candidate features and w_1, w_2, w_3 are weights for each colour channel. After discarding $(w_1, w_2, w_3) = (0, 0, 0)$ and redundant combinations $(w_1', w_2', w_3') = k(w_1, w_2, w_3)$), Eq. 9 yields 49 input features. The *Collins+* and *CACTuS-FL* input features consist of a subset of only 9 colour channel combinations, obtained by using $w_* \in [-1, 0, 1]$ in Eq. 9, together with a spectral saliency feature,[23] a motion history image feature,[24] four Gabor[14] filter features, and two color opponency (Red-Green, Blue-Yellow) features.

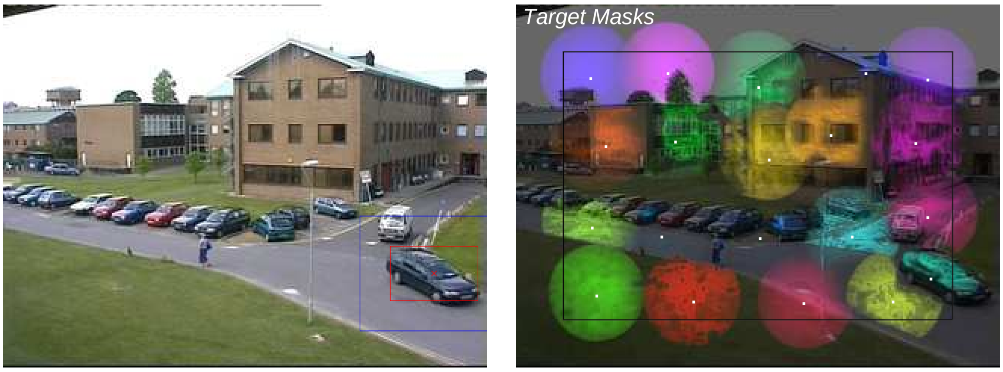

Figure 2: Target and local background regions for the *Collins/+* (*left*) and *CACTuS-FL* (*right*) tracking algorithms, shown for the second frame of the *car* video sequence.[22] The target region for the Collins+ tracker is indicated by the red rectangle, while that of the *CACTuS-FL* tracker this is defined using the probabilistic pixel weight mask shown in cyan (over the *car*). The local background region for the *Collins/+* tracker is defined by the area between the blue and red rectangles, while that of the *CACTuS-FL* tracker is defined using the complement of the probabilistic target mask, computed over an appropriately chosen rectangular region based on the position of the tracker (white dots).

All of the discriminant tracking algorithms tested here compute a Likelihood Ratio for each feature: $LR = F^t(u)/B^t(u)$, and this is propagated back into each feature map to generate a set of LR (detection) maps at every frame t. The numerator and denominator of LR are the class-conditional feature response PMFs: the normalized target region feature response histogram ($F^t(u)$) and the normalized local background region feature response histogram ($B^t(u)$), where u is a bin corresponding to a range of feature response values.

In *CACTuS-FL*, F^t is computed by using the learnt object image from the previous frame I_s^{t-1} as a pixel weighting mask and feature values from the current frame Z^t according to:

$$F^t(u) = \frac{\sum_{\boldsymbol{i}} I_s^{t-1}(\boldsymbol{i})\,\delta\left(Z^t(\boldsymbol{i}) - u\right)}{\Sigma_u}, \tag{10}$$

where $\boldsymbol{i}$ is a given pixel index, δ is the Dirac delta function and $\frac{1}{\Sigma_{\boldsymbol{u}}}$ is a normalization constant. B^t is generated in a similar way, using pixel weighting mask of $1 - I_s^{t-1}$ over an appropriately chosen patch of the frame. Using the learnt image I_s^{t-1} to define the target and local background regions leads to a more precise extraction of class-conditioned feature response, which has been shown to reduce model pollution and mitigate the effects of model drift,[21] thereby making the tracking more robust. This idea is illustrated conceptually in Figure 1 and shown in practice in the *right* plot of Figure 2. For the *Collins/+* algorithms the target and local background feature response are extracted in a less flexible manner using *a priori* defined rectangular binary masks, as illustrated in Fig. 2.

While the LR maps in *Collins/+* are passed to the tracking process separately, the output of the CACTuS-FL detection stage is a single fused LR map that is then passed to its SEF. This LR map fusion is normally implemented as a weighted sum,[9] where the weights are the scores used to rank features for feature selection (see below). In this study however, an un-weighted summation of likelihood maps is used instead in order to decouple the effects of feature selection and fusion on the overall tracking performance.

3.3 Online Feature Selection

Online feature selection involves choosing, in each frame t, a subset of N features to be used in the tracking stage. The value of N is chosen to be 6 in this study, however it should be noted that this choice is somewhat arbitrary, and, based on our experience, other similar (e.g. 5, 4) values may provide comparable performance. The *baseline* feature selection algorithm used in *Collins/+* ranks features according to the *Variance Ratio*[1] (VR)

metric, calculated using the log likelihood ratio $L = log(LR)$ and class-conditioned feature response distributions F and B:

$$VR(L; F, B) = \frac{var(L; (F + B)/2)}{var(L; F) + var(L; B)} \ , \tag{11}$$

where, for a given PMF p, the variance of L with respect to p is: $var(L; p) = \sum_u p(u)L^2(u) - [\sum_u p(u)L(u)]^2$. In *Collins/+* features are ranked according to their *VR* scores and the corresponding top 6 *LR* maps are chosen as inputs to the mean shift tracker.

In the *CACTuS-FL baseline* algorithm the feature score is given by the product of *VR* and a *similarity score* that is based on the Bhattacharyya coefficient:[25]

$$B = \sum_u \sqrt{F_m^t\,(u)\,F_s^{t-1}\,(u)} \ , \tag{12}$$

where B rewards temporal consistency between the measured target region feature response F_m^t measured for the current frame and a target region feature response learnt up to the previous frame F_s^{t-1}. Once the *similarity score* is computed, the learnt feature response is updated:

$$F_s^t\,(u) = \frac{F_s^{t-1}\,(u)\,F_m^t\,(u)}{\Sigma_u} \ , \tag{13}$$

where $\frac{1}{\Sigma_u}$ is a normalization constant, so that F_s^t is the posterior learnt target PMF for the current frame. Each feature is ranked according to its combined score ($VR \times B$). The corresponding top 6 *LR* maps are averaged to yield the combined *LR* map that serves as input to the SEF.

The information theoretic *MMD* and *mRMR* algorithms provide two alternatives to the *baseline* feature selection algorithms described above. The application of *MMD* involves calculating *marginal diversity* $I(X_k; C)$ for each feature, ranking them according to this and then selecting the top 6 *LR* maps that correspond to those features. The *marginal diversity* $I(X_k; C)$ is computed according to Eq. 1. The joint PMF $p(X_k = x; C = i)$ required for this is built by concatenating the two un-normalized class-conditional feature response 1D histograms and normalizing the resulting 2D histogram. The marginal PMFs $p(X_k = x)$ and $p(C = i)$ are then simply computed by integrating all bins along the C and X_k axes of $p(X_k = x; C = i)$, respectively.

The *marginal diversity* scores are also used in *mRMR*, which proceeds as a forward search in which features that satisfy Eq. 8 are sequentially removed from the set of remaining features ($\mathbf{X} \setminus S_{m-1}$) and included in the set of selected features (S_{m-1}), which initially contains the feature with the largest *marginal diversity.* The *Mutual Information* between two features $I(X_j; X_i)$ is computed according to Eq. 1 and also requires building a joint PMF and its marginals. The region chosen to fill this joint PMF is the rectangular patch of feature map from which both the target and local background feature response are extracted.

4. EXPERIMENTAL EVALUATION

4.1 Qualitative Evaluation

Fig. 3 and 4 illustrate how feature selection works in practice by applying *Collins+* and *CACTuS-FL*, respectively, to the *car* video sequence.[22] In Fig. 3 the plots on the *left* show the evolution, with frame number, of *VR* (for the *baseline* algorithm) and of *marginal diversity* ($I(X_k; C)$) (for the *MMD* algorithm). In Fig. 4, the *CACTuS-FL baseline* feature selection scheme uses combined scores of $VR \times B$, so a scheme using only *VR* is also shown for comparison, along with a scheme using *marginal diversity.* Both of *VR* and *marginal diversity* measure how discriminant a given feature is. Large temporal fluctuations in these scores tend to indicate changing local background conditions, which tend to take place in the frames just before and after the tracker loses the object (indicated by the vertical red lines).

The plots on the right in Fig. 3 and 4 show the features selected by each scheme, which are the 6 top ranked features at each frame. Comparing Fig. 3 to Fig. 4 suggests that the overall effect of using the learnt object shape to extract class-conditioned feature responses is an enhanced temporal consistency in the choice of features: for

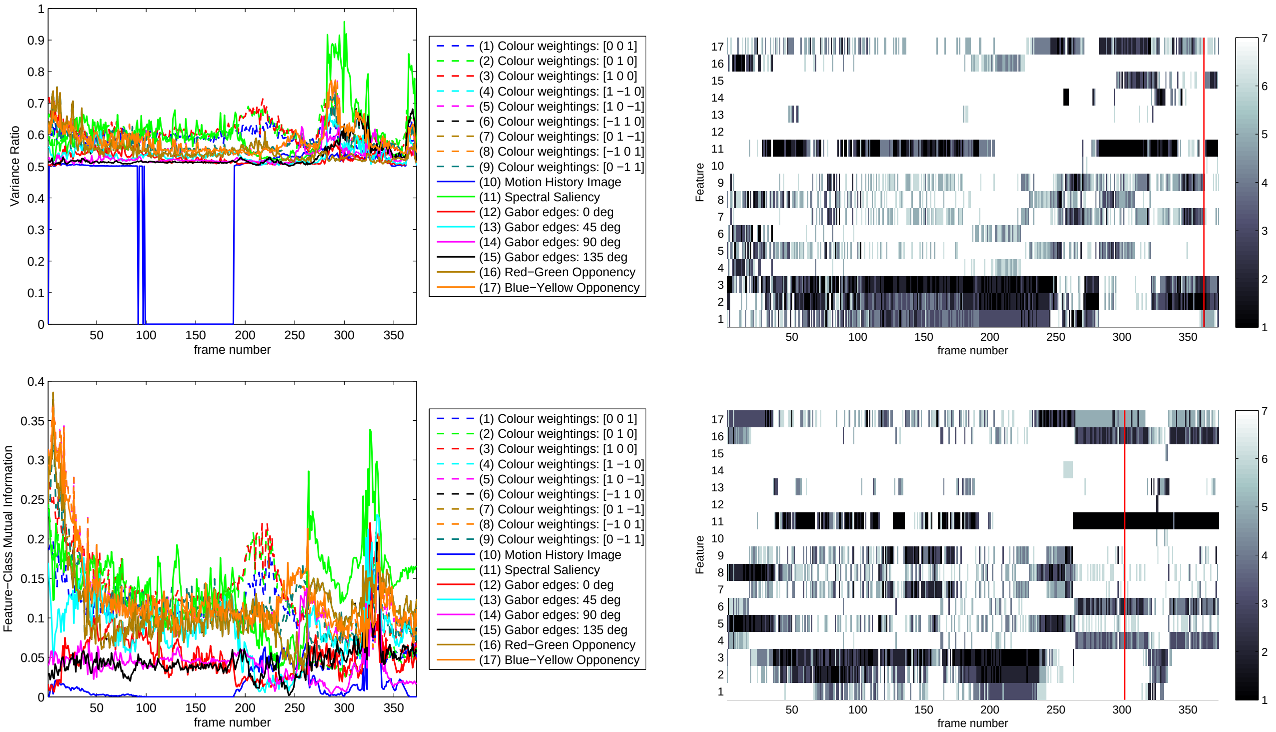

Figure 3: Feature selection based on the *baseline* scheme, which uses *Variance Ratio* (*VR*) (*top* plots), and a scheme using *maximum marginal diversity* (*MMD*) (*bottom* plots), applied in the *Collins+* tracker for the *car* sequence.[22] The plots on the *left* show the evolution of *VR* and *marginal diversity* ($I(X_k; C)$) scores, respectively, as functions of the frame number. The plots on the *right* show the top 6 ranked features at each frame whose corresponding likelihood ratio (LR) maps are the inputs to the mean shift tracker.[3] The candidate feature index indicated on the *y-axis* correspond to the numbers listed in the legend on the *left*. The vertical red lines indicate the frames at which the tracker is considered by the VOT2013 framework[22] to have stopped tracking the target.

all feature selection schemes the plots on the *right* show a more consistent selection across the different frames in Fig. 4 than in Fig. 3. Fig. 4 also shows that for the *CACTuS-FL* algorithm there is a clear differentiation in the preferred features of the *baseline* ($VR \times B$) and the VR/MMD feature selection schemes, with colour combination features being preferred by both *VR* and *MMD*. These features tend to be more correlated with each other than other types of features, which may worsen the overall tracking performance (see below and Table 3).

4.2 Quantitative Evaluation

The tracking performance given by online feature selection with *MMD* or *mRMR* is evaluated across the *Collins*, *Collins+* and *CACTuS-FL* visual tracking algorithms against their *baseline* performance. These comparisons use 8 of the 16 videos used in the Visual Object Tracking Challenge VOT2013,[22] which were recorded under changing background, lighting conditions and/or object appearance. The video sequences together with a tracking performance evaluation kit are publicly available at `http://votchallenge.net/vot2013/`. Each video contains a single object of interest to be tracked. The selected videos were those whose camera motion (camera jerk) is within the limits of the *CACTuS-FL* motion model. Tracking performance is evaluated through the VOT2013 evaluation kit and is based on the *robustness*[22] score, which is defined as the number of times that a tracking algorithm lost track of the object of interest and had to be re-initialized. The tracking algorithm is considered to have lost the track when the spatial overlap between tracker predicted bounding box and the ground truth bounding box drops to zero.[22]

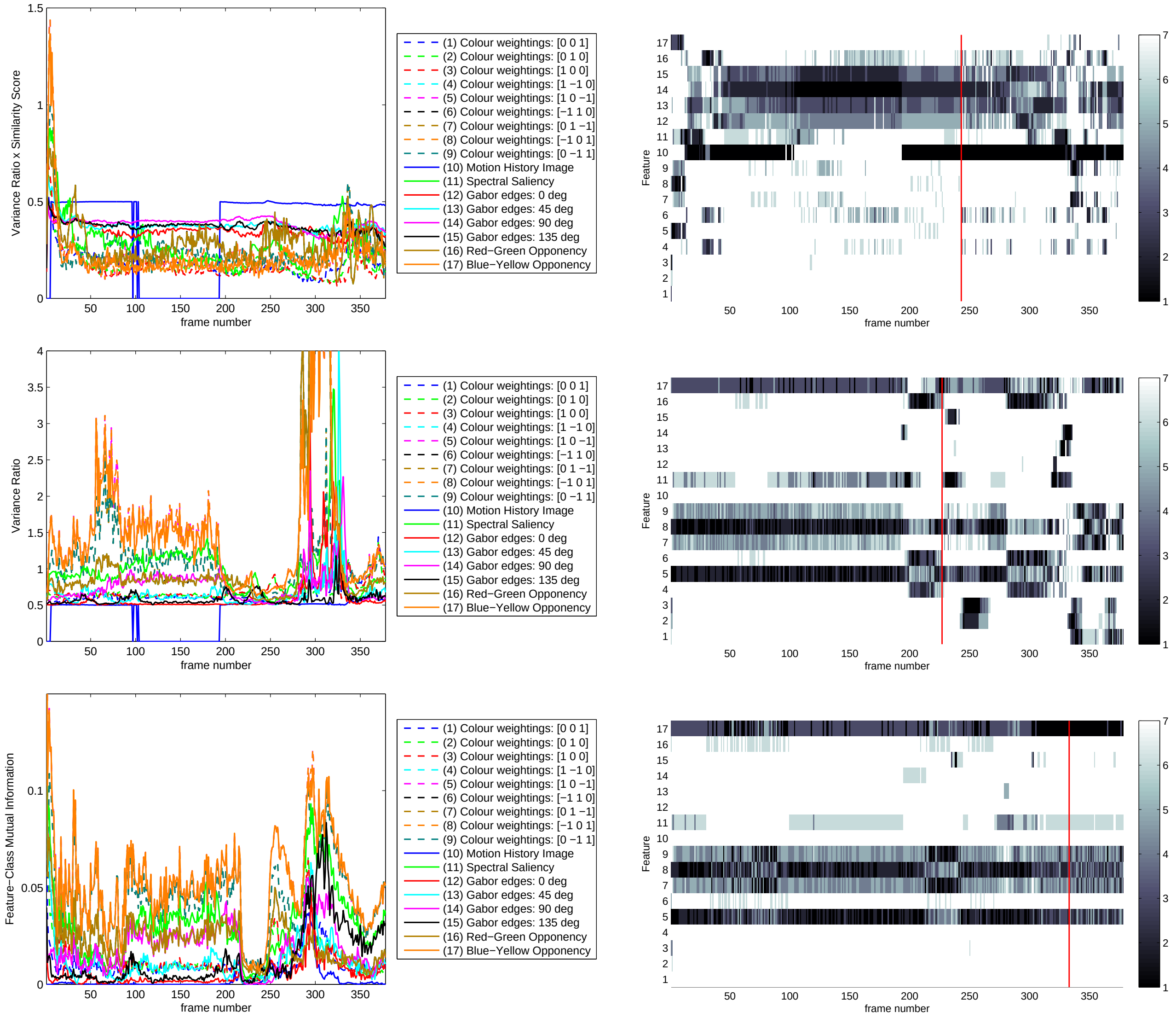

Figure 4: Feature selection based on the *baseline* scheme, which uses *Variance Ratio* × *Similarity Score* ($VR \times B$) (*top* plots), a scheme that uses only the *Variance Ratio* (VR) (*middle* plots), and a scheme that uses *maximum marginal diversity* (MMD) (*bottom* plots), applied in the *CACTuS-FL* tracker for the *car* sequence.[22] The plots on the *left* show the evolution of $VR \times B$, VR and *marginal diversity* ($I(X_k; C)$) scores, respectively, as functions of the frame number. The plots on the *right* show the top 6 ranked features at each frame whose corresponding likelihood ratio (LR) maps are averaged and this becomes the input for the shape estimating filter.[20] The candidate feature index indicated on the *y-axis* correspond to the numbers listed in the legend on the *left*. The vertical red lines indicate the frames at which the tracker is considered by the VOT2013 framework[22] to have stopped tracking the target.

Tables 1, 2 and 3 provide the *robustness* scores for the *Collins*, *Collins+* and *CACTuS-FL* tracking algorithms, respectively. The *robustness* scores for each video sequence (columns) are shown in each table for three online feature selection algorithms (rows): *baseline*, *MMD* and *mRMR*. Table 3 provides additional results for a feature selection scheme using only *VR*, provided becuase the *CACTuS-FL baseline* scheme uses combined scores of $VR \times B$. Columns on the far right provide the total *robustness* score for all 8 video sequences and a particular feature selection scheme.

Tables 1 and 2 show that when applied to *Collins* and *Collins+*, both *MMD* and *mRMR* lead to enhanced tracking performance with respect to the *baseline*. The improvement offered by both *MMD* and *mRMR* is more pronounced for *Collins+*, which uses a wider variety of input feature algorithms. *mRMR* yields the best results for both *Collins* and *Collins+*, with the latter showing a striking degree of improvement, so much so that the *mRMR robustness* score of 7 is comparable to the results achieved using *CACTuS-FL* in Table 3.

The *mRMR* feature selection scheme does not improve on the *CACTuS-FL baseline* performance, although it should be noted that in absolute terms, this *baseline* already provides good tracking performance that is competitive against what is achieved by recent state of the art single-object tracking algorithms.[9] These results show that using the learnt object shape to extract the class-conditioned feature response benefits all of the tested feature selection schemes (compare Tables 2 and 3). It is interesting to note that when applied in *CACTuS-FL*, both the *VR* and *MMD* feature selection schemes perform poorly compared to the *baseline* and *mRMR*. This shows that incorporating a feature *similarity score* to reward temporal consistency leads to better overall performance in the *baseline* algorithm than what is achieved with *VR* alone or with *MMD*. Taking the redundancies between features into account by applying *mRMR* then improves on the performance of *MMD*, matching that of the *baseline*.

Table 1: The tracking *robustness*[22] of *Collins* for different online feature selection algorithms using videos from the Visual Object Tracking Challenge VOT2013.[22]

Algorithm	**car**	**david**	**diving**	**face**	**gymnastics**	**iceskater**	**jump**	**woman**	**total**
baseline (*VR*)	3	10	**0**	**5**	3	4	**0**	5	30
MMD	3	**5**	2	6	2	**0**	1	5	24
mRMR	**2**	**5**	5	**5**	**1**	**0**	1	**3**	**22**

Table 2: Tracking *robustness*[22] of *Collins+* for different online feature selection algorithms using videos from the Visual Object Tracking Challenge VOT2013.[22]

Algorithm	**car**	**david**	**diving**	**face**	**gymnastics**	**iceskater**	**jump**	**woman**	**total**
baseline (*VR*)	1	12	4	7	2	2	1	6	35
MMD	3	4	3	6	2	**0**	1	4	23
mRMR	**0**	**3**	**0**	**2**	**0**	**0**	**0**	**2**	**7**

Table 3: Tracking *robustness*[22] of *CACTuS-FL* for different online feature selection algorithms using videos from the Visual Object Tracking Challenge VOT2013.[22] Given that, in the case of *CACTuS-FL*, the *baseline* feature selection scheme uses the combined score of *Variance Ratio* × *Similarity Score* ($VR \times B$), the *robustness* obtained with a feature selection scheme that uses only *Variance Ratio* (*VR*) is also provided for comparison.

Algorithm	**car**	**david**	**diving**	**face**	**gymnastics**	**iceskater**	**jump**	**woman**	**total**
baseline ($VR \times B$)	**1**	**0**	**0**	**0**	**0**	**0**	**0**	4	**5**
VR	4	3	3	**0**	2	2	1	5	20
MMD	**1**	2	3	**0**	3	2	**0**	3	14
mRMR	2	**0**	**0**	**0**	**0**	**0**	1	**2**	**5**

5. CONCLUSION

In this paper we have investigated experimentally the problem of online feature selection in discriminant visual tracking. In particular, we have sought to approximate the *infomax space* of features, which provides optimum classification in the Bayes classification error sense.[13] To this end, we have applied the *minimal-redundancy-maximal-relevance*[5] (*mRMR*) criterion in full to the problem of choosing the subset of features that best discriminate between a target and its local background. Experiments were conducted on real world data sets using a single-object discriminant tracking algorithm which was run separately for two input feature libraries. The results showed that *mRMR* offers improved tracking robustness with respect to alternative online feature selection schemes that use the *Variance Ratio* metric or *maximum marginal diversity* (*MMD*), neither of which account for redundancy among the input features. The level of improvement is dependent on the input feature library, and a more advanced library of features allowed the *Collins+* algorithm to reach a similar level of performance to that achieved by the advanced multi-object tracking system *CACTuS-FL* using the same library.

ACKNOWLEDGMENTS

We would like to thank Dr Dmitri Kamenetsky for suggestions that helped us to improve this paper.

REFERENCES

[1] Collins, R., Liu, Y., and Leordeanu, M., "Online selection of discriminative tracking features," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **27**, 1631–1643 (October 2005).

[2] Matthews, L., Ishikawa, T., and Baker, S., "The Template Update Problem," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **26**(6), 810–815 (2004).

[3] Comaniciu, D., Ramesh, V., and Meer, P., "Real-time tracking of non-rigid objects using mean shift," in [*Conference on Computer Vision and Pattern Recognition (CVPR'00), IEEE Computer Society*], 142–149, IEEE (June 2000).

[4] Vasconcelos, N. and Vasconcelos, M., "Scalable discriminant feature selection for image retrieval and recognition," in [*Conference on Computer Vision and Pattern Recognition (CVPR'04), IEEE Computer Society*], 770–775, IEEE (June 2004).

[5] Peng, H., Long, F., and Ding, C., "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **27**, 1226–1238 (August 2005).

[6] Vasconcelos, M. and Vasconcelos, N., "Natural image statistics and low-complexity feature selection," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **31**, 228–244 (February 2009).

[7] Mahadevan, V. and Vasconcelos, N., "Saliency-based discriminant tracking," in [*Conference on Computer Vision and Pattern Recognition (CVPR'09), IEEE Computer Society*], 1007–1013, IEEE (June 2009).

[8] Mahadevan, V. and Vasconcelos, N., "Biologically inspired object tracking using center-surround saliency mechanisms," *IEEE Trans. Pattern Anal. Mach. Intell* **35**(3), 541–554 (2013).

[9] Wong, S., Gatt, A., Kearney, D., Milton, A., and Stamatescu, V., "A competitive attentional approach to mitigating model drift in adaptive visual tracking," in [*The 29th International Conference on Image and Vision Computing New Zealand (IVCNZ'14)*], 1–6, ACM (November 2014).

[10] Mangai, U., Samanta, S., Das, S., and Chowdhury, P., "A survey of decision fusion and feature fusion strategies for pattern classification," *IETE Technical review* **27**, 293–307 (July 2010).

[11] Cover, T. and Thomas, J., [*Elements of Information Theory, 2nd Edition*], Wiley-Interscience (2006).

[12] Press, W., Flannery, B., Teukolsky, S., and Vetterling, W., [*Numerical recipes in C*], Cambridge University Press, Cambridge (1988).

[13] Vasconcelos, N., "Feature selection by maximum marginal diversity: optimality and implications for visual recognition.," in [*Conference on Computer Vision and Pattern Recognition (CVPR'03), IEEE Computer Society*], 762–769, IEEE (June 2003).

[14] Itti, L., Koch, C., and E., N., "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **20**, 1254–1259 (1998).

[15] Yu, L. and Liu, H., “Feature selection for high-dimensional data: A fast correlation-based filter solution,” in [*Proceedings of the Twentieth International Conference on Machine Learning*], 856–863 (2003).

[16] Alvarez-Santos, V., Pardo, X., Iglesias, R., Canedo-Rodriguez, A., and Regueiro, C., “Feature analysis for human recognition and discrimination: Application to a person-following behaviour in a mobile robot,” *Robotics and Autonomous Systems* **60**, 1021–1036 (August 2012).

[17] Leung, A. and Gong, S., “Online feature selection using mutual information for real-time multi-view object tracking,” in [*AMFG’05 Proceedings of the Second international conference on Analysis and Modelling of Faces and Gestures*], 184–197, Springer-Verlag (October 2005).

[18] Cui, P., Sun, L., and Yang, S., “Adaptive mixture observation models for multiple object tracking,” *Science in China Series F Information Sciences* **52**, 226–235 (2009).

[19] Hong, K. and Han, K., “A multiple feature based particle filter using mutual information maximization,” in [*Intelligent Robots and Computer Vision XXVIII: Algorithms and Techniques*], 1–9, SPIE (January 2011).

[20] Wong, S. and Kearney, D., “Relating image, shape, position, and velocity in visual tracking,” in [*Proc. SPIE 7338, Acquisition, Tracking, Pointing, and Laser Systems Technologies XXIII*], SPIE (May 2009).

[21] Gatt, A., Wong, S., and Kearney, D., “Combining online feature selection with adaptive shape estimation,” in [*25th International Conference of Image and Vision Computing New Zealand (IVCNZ), 2010*], 1–8, IEEE (November 2010).

[22] Kristan, M. et al., “The visual object tracking vot2013 challenge results,” in [*IEEE Workshop on Video Object Tracking (in conjunction with ICCV)*], 98–111, IEEE (November 2013).

[23] Hou, X. and Zhang, L., “Saliency Detection: A Spectral Residual Approach,” in [*Conference on Computer Vision and Pattern Recognition (CVPR’07), IEEE Computer Society*], 1–8, IEEE (2007).

[24] Yin, Z. and Collins, R., “Moving Object Localization in Thermal Imagery by Forward-backward MHI,” (2006).

[25] Bhattacharyya, A., “On a measure of divergence between two statistical populations defined by their probability distributions,” *Bulletin of the Calcutta Mathematical Society* **35**, 99–109 (1943).