3-2005

A Dynamic Pruning and Feature Selection Strategy for Real-Time Tracking

D. Frank Hsu
Fordham University

Damian M. Lyons
Fordham University

A Dynamic Pruning and Feature Selection Strategy for Real-Time Tracking

D. Frank Hsu and Damian M. Lyons

Computer Vision & Robotics Laboratory
Department of Computer & Information Science
Fordham University,
Bronx, NY 10458, USA
{hsu,lyons}@cis.fordham.edu

Abstract. Automated video tracking is useful in a number of applications such as surveillance, multisensor networks, robotics and virtual reality. In this paper we investigate an approach to tracking based on fusing the output of a collection of video trackers, each attending to a different feature or cue on the target. We show both theoretically and experimentally that the method used to prune the growth of target hypotheses can have a great impact on the trackers performance, and indirectly, change the benefit of using linear score combination as opposed to a non-linear rank combination for fusion. We also show that the rank-score graph defined by Hsu and Taksa can be used to select a subset of features to fuse to reduce classification error.

Index Terms - *Sensor Fusion, Target Tracking, Multisensor Networks, Video Analysis, Sorting/Searching.*

1. INTRODUCTION

The automated tracking of designated targets in a video image is a general problem that has applications in surveillance, multisensor networks, robotics and virtual reality among other areas. It has however, remained a difficult problem to solve [8].

In previous work [5,10], we have proposed an approach to tracking in which the outputs from a collection of trackers, all viewing the same scene, are combined to produce more accurate tracking. In this approach, which we call 'Rank and Fuse' (RAF for short), each tracker produces a ranked list (which includes a rank function and a score function) of matches between a target and candidate targets in the video which are added to a growing pool of target hypotheses. Once the pool reaches a threshold size, the scores (or ranks) assigned to each candidate by each tracker are *fused* together, and the pool is reduced ('pruned') to just the best hypotheses. This hypothesis generation approach is based on the Multiple Hypothesis Tracking algorithm of Reid [1,2,15]. However, RAF differs from most other approaches to fusion for target tracking in using both rank information and score to combine the results from different trackers.

The advantage of MHT is that it defers decision making about correct tracks. At any time, the correct track hypothesis for a target may not be the one rated highest by the tracker, but since all hypotheses are kept, when later measurements support the correct track hypothesis its score will be raised. However, this is clearly of combinatorial complexity; the pool of track hypothesis will grow combinatorially with time as new images are taken and measurements are made. This can quickly surpass computational resources, and destroy the real-time performance of a tracking system, unless steps are taken to reduce the size of the track hypothesis pool. Therefore there are two issues that can be addressed to improve efficiency:

(1) <u>*Pruning the Hypothesis Pool:*</u> A straightforward way to reduce the size of the track hypothesis pool with little decrease in the power of the approach is to remove very low scoring track hypotheses from the pool. Reid [15] proposes this in his seminal work on MHT. Other variations on this include keeping the top n scoring hypotheses, keeping the highest scoring hypothesis derived from the same parent hypothesis in the last n steps [2], and so forth. More sophisticated methods of addressing the complexity issue are possible [22], but pruning has a very low computational cost.

(2) <u>*Fuse a Subset of Features:*</u> There are several reasons to consider using only a subset of the available features in a fusion: It is more efficient computationally to process and fuse a smaller set of features. Some fusion operations, such as a weighted sum, may cause the useful information presented by a subset of features to be outweighed by other features. Finally, it has been observed [5-7,10,13] that when selecting which measurements to combine from a set of feature measurements, better results are obtained when combining measurements that come from processes that have different ranking behaviors.

In this paper, we present first a theoretical analysis of the effect of pruning the hypothesis pool, using the rank-score graph concept of Hsu, Shapiro and Taksa [6], and Hsu & Taksa [7]. We show that as long as the pruning threshold score value p_x is greater than a critical value p_c then pruning produces a much larger variation in ranks in the hypotheses pool. We will propose that this means that the benefit of score-based fusion and rank-based fusion will vary, depending on the pruning threshold. We then present experimental results that demonstrate this effect. Finally we also present an approach for selecting which subset of features to choose.

Section 2 introduces the problem briefly. Section 3 presents a review of literature. Section 4 investigates the effect of a cutoff value in reducing pruning and Section 5 presents the experimental results. Section 6 investigates the use of a subset of features in fusion. Section 7 concludes with a discussion of all these results.

2. THE PROBLEM

We consider a system that tracks multiple moving targets in a video sequence. The output of the system is a

ranked and scored list of tracks for each target. The better the rank and score of a track, the more the tracking system supports the hypothesis that this is the correct track for the target. This ranked list represents the decision the tracking system is making about which track corresponds to the actual target. Now consider a set of such tracking systems, $TR_1..TR_m$, each using different sensing modalities and/or tracking approaches to determine its ranked list of tracks (Fig. 1). Here a ranked list will include a rank function *r* and a score function *s*. The fusion problem we are interested in is to determine how and when to combine the lists to a single list.

The set of *N* track hypotheses generated by a tracker up to some image frame *i* in the video sequence, will be referred to as the pool of track hypotheses, $\mathbf{T}_i$. Let the set of continguous positive integers from 0 to a maximum value *N-1, {0 .. N-1},* be the labels for these hypotheses. We will assume that each tracker agrees on this set, by virtue of a common hypothesis generation stage [5,10] or by the computational generation of a set of composite tracks [11].

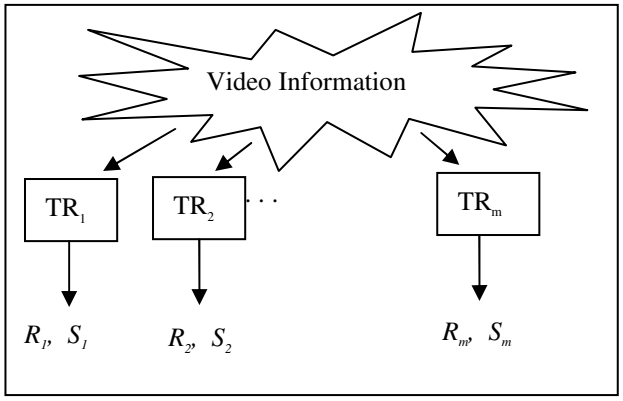


Figure 1: Mutiple Tracker Configuration

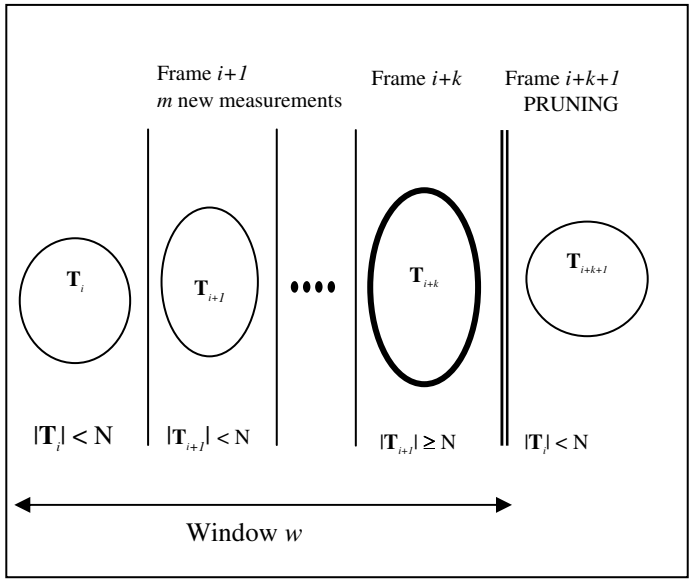


Figure 2: Pruning **T** based on window *w*.

In our RAF implementation [5], on each frame *i* of the video sequence, each tracker *j* makes its own measurements on the image. A common, MHT-like data association step takes the track hypothesis pool generated so far $\boldsymbol{T}_{i-1}$ as input and produces a new set of tracks based on the association of measurements in the image *i* with the existing track hypotheses. A set of score values is associated with each hypothesis in the pool; one value for each tracker. The common association step generates a new track hypothesis pool $\boldsymbol{T}_i$ which is available to all trackers. Each tracker modifies its separate score value on every track in the pool. The pool continues to grow for a time window of size *w* (determined for example by the number of frames, the size of the pool, score characteristics of the pool, etc). This growth process is illustrated in Figure 2.

This growth is exponential in the number of frames and hence needs to be winnowed down to stay within the limits of computational resources. At the end of the time window, the track hypothesis pool is considered as a set of scored and ranked lists, one list per score component, that is, one list per tracker. The rank and score information from each list is fused to generate a single, combined list for the hypothesis pool. The top scoring *q* hypotheses are preserved and the remainder deleted. The track hypothesis pool used at the start of the next time window consists of these top scoring *q* hypotheses.

3. Prior Work

A number of approaches have been developed for target tracking [1,8]. Chief among these have been Multiple Hypothesis Tracking (MHT) [1, 2, 15], Joint Probability Density Association Filter (JPDAF) [1, 14] and Probabilistic MHT (PMHT) [21]. One intuitive approach to improve tracking performance is to consider fusing information from multiple feature measurements [19]. A Bayesian approach to fusion follows naturally from an MHT or JPDAF based approach to tracking. In general it is assumed that the different feature measurements are conditionally independent, and therefore that the conditional probability of an estimated quantity *S* given a collection of image data *I* can be expressed using Bayes rule. In the standard framework for linear estimation, this gives rise to an estimate for S that is a linear score combination of the cues where the combination coefficients are inversely proportional to the variance [17,20]. We refer to this as the *Bayes* fusion score. This approach to fusion, in which the different features are evaluated separately and then combined, has also been called *weak fusion* [16].

However, fusion for tracking can also be viewed as a combination of pattern classifiers at the so-called *measurement level* [23]. This suggests that both rank and voting operations should be evaluated for combination as well as the score combination. These non-linear combinations have been called *strong fusion* [16], and there is evidence that humans employ both weak and strong fusion approaches in perception. Several strong fusion approaches have been proposed for video tracking, e.g., [12,18]. In [5, 10] we proposed an architecture called the "Rank and Fuse" Architecture (RAF) with the purpose of *evaluating* fusion by linear score combination as well as by non-linear rank combination for multiple target tracking in the video surveillance domain. Our preliminary results showed that there were cases where rank combination outperformed

score combination, indicating that an effective multiple-target tracker will need to use both.

4. Effect of Threshold in Pruning

It has been observed in video tracking using MHT [2, 15] that despite the inherent exponential complexity, pruning with a probability cutoff value can be used to quickly remove many low probability hypotheses. Indeed if the scene is relatively uncluttered, and with few targets, then when all track hypotheses are generated, the vast majority will receive a low probability of being a correct target track and few will have a high probability. This distribution of probabilities to track hypotheses is shown in the histogram of Fig. 3(a), where the vertical (ϕ) axis is frequency and the horizontal (p) axis is probability, and there is one point in the area udner the curve $\phi=h_a(p)$ for each track hypothesis.

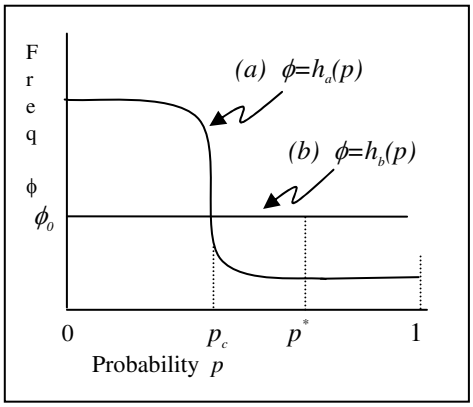
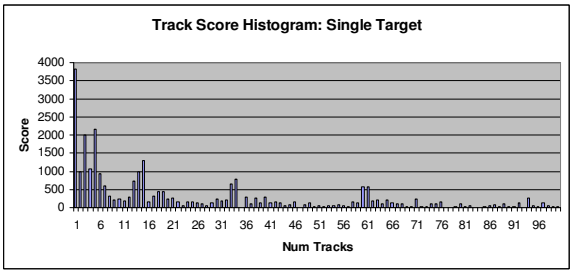
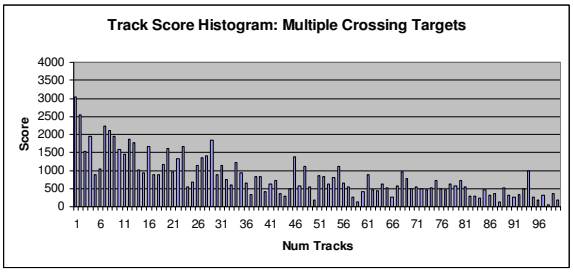


Figure 3: Frequency of Probabilities for N Track Hypotheses in (a) Simple Env. (b) Complex Env.



4(a) Single, Unoccluded Target



4(b) Multiple, Crossing Targets

Figure 4: Example Track Score Histograms for Tracking using Position Feature

However, for a complex scene, a scene with multiple targets and with confusing clutter in the background, the distribution is different: many tracks have roughly the same probability of being correct. This situation is illustrated in the histogram of Fig. 3(b). We have verified that this phenomenon occurs in practice. Using the RAF tracker described in [10] with a target position feature measurement, histograms were recorded to study the distribution of scores to target track hypothesis. Fig. 4 shows two such example histograms.

Fig. 4(a) shows the score distribution of the top 50% of tracks when tracking a single, unoccluded target. 4(b) shows the distribution of the top 50% of tracking when tracking multiple targets that cross each other. Scores are more evenly distributed in 4(b) than in 4(a); the correct choice of target is much less clear cut. Pruning with a probability cutoff value is not quite so straightforward under these circumstances, idealized in 3(b), as we shall demonstrate in the following. In fact, we will show that the selection of a probability cutoff p_x has a much greater effect on the variation in ranks in a complex tracking environment (3(b)) than in a simple tracking environment (3(a)).

4.1. The Rank-Score Graph

Let r be the rank function $r : \{0 .. N-1\} \rightarrow \{1..N\}$ where $r(i)$ is the *rank* of the track i in the ranked output list, where the leftmost element of the list is considered to have rank 1, the next leftmost has rank 2, etc. Let s be the score function, $s : \{0 .. N-1\} \rightarrow \{1..S_{max}\}$ where S_{max} is the maximum score value. The score function $s(i)$ assigns a value, the score, to each track i in the list. The track with highest score is the track with best rank, i.e., with rank equal to 1. We constrain the rank to reflect the score as follows:

$$s(i) > s(j) \quad \Rightarrow \quad r(i) < r(j).$$

There is ambiguity when two tracks have the same score. To resolve this, we add the constraint:

$$s(i) = s(j) \wedge i < j \quad \Rightarrow \quad r(i) < r(j).$$

The score function characterizes how each tracker processes and rates the track hypotheses, with a higher score meaning that the tracker considers that the evidence supports that track hypothesis more than lower scoring hypotheses. Each tracker could use different cue or feature information, or combination of features, or even a different tracking algorithm, as long as there is a composite set of track hypotheses.

Hsu, Shapiro and Taksa [6] and Hsu and Taksa [7] introduce an approach to characterizing the scoring behavior of experts. Consider the following example. A particular expert may assign its scores in linear fashion from highest to lowest. Another expert may habitually give much lower scores to several of its top ranked candidates. Averaging the scores from the two experts will always give the latter expert's top candidates less emphasis. In a situation such as this, where the ranking behavior of the two experts is not the same, using the rank information in place of the score may yield a better combined result [4-6,9,10,13]. Hsu et al. [6,7] characterize the relationship that an expert habitually produces between rank and score as the *rank-score graph* f: $\{1..N\} \rightarrow \Re$, a monotonic non-increasing function f that relates rank and score:

$$f(r(i)) = s(i).$$

As discussed above, the shape of the graph can be different for different trackers and is a characteristic of that tracker's scoring approach. So, in our previous example, the expert who assigns scores in a linearly decreasing fashion will have

a linear rank-score graph (e.g., f_3 in Fig. 5). The expert who habitually assigns high scores to a subset of its top ranked

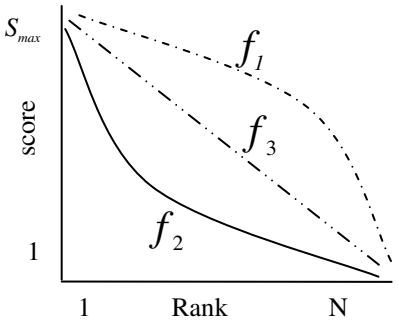
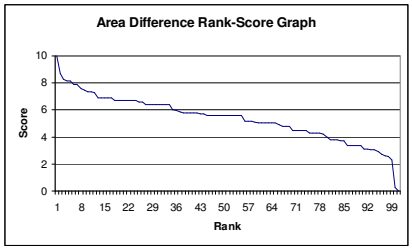


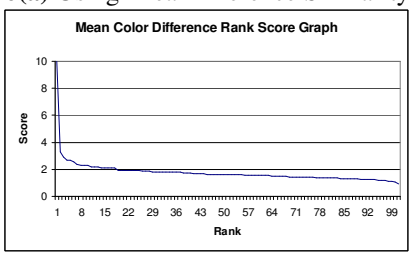
Figure 5: General Classes of Rank-Score Graphs

candidates will have a graph that is not a straight line, but has a low slope for the first few candidates and a higher slope for the remainder. The concave-down graph f_1 in Fig. 5 is an example of this. A third class of scoring behavior is exemplified by f_2 in Fig 5. In this case, the expert habitually gives relatively lower scores to a subset of its top ranked candidates.

The scoring behavior that is captured by a rank-score graph is a characteristic of the choice of cues or feature measurements and tracking algorithm used by the tracker. In Fig. 6 we present two rank score graphs obtained using the RAF tracker described in [10]. In 6(a) the tracker used a similarity measure based on a squared difference of the predicted versus actual area of the target, whereas in 6(b) a squared difference of mean RGB color was used. Both ran for the same time on the same video sequence. Other examples are presented in [10].



6(a) Using Area Difference Similarity



6(b) Using Mean Color Difference Similarity

Figure 6: Examples of Averaged Rank-Score Graphs

4.2. Pruning in Simple and Complex Environments

First we use the information in the distributions 3(a) and (b) to calculate the rank-score graphs for each case. We will assume for this paper that the score distribution is the same as the probability distribution: $s = p = f(n)$. The results of Fig. 4 show that this is a reasonable assumption.

We claim the rank-score graph for 3(a) (simple environment) is concave up and that the graph for 3(b) (complex environment) is linear as shown in Fig. 7. In Fig. 3 we have that frequency ϕ is a function of the probability $\phi = h(p)$. In that case, the rank-score graph f can be expressed as a function of the probability as follows:

$$f^{-1}(p^*) = \int_{p^*}^{1} h(p)dp \,. \qquad (1)$$

Notice that for any $n \in 1..N$, $f_a(n) \le f_b(n)$ since if we cut the graph at p^* on the y-axis in Fig. 7, then

$$f_b^{-1}(p^*) = \int_{p^*}^{1} h_b(p)dp \quad \ge \quad \int_{p^*}^{1} h_a(p)dp$$
$$= f_a^{-1}(p^*). \qquad (2)$$

As p moves from 1 to 0, the change in rank in case (b) is uniform, whereas the change in rank in case (a) slowly increases and increases more rapidly after a certain point. Call this point p_c (see Fig. 7).

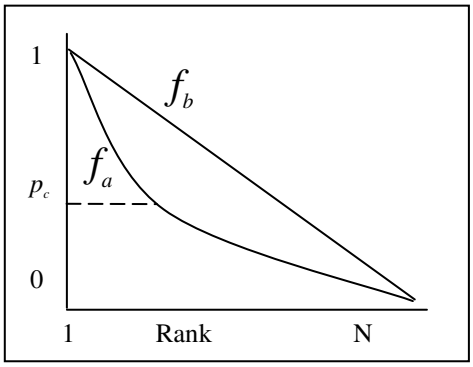


Figure 7: Rank-Score Graphs f_a and f_b superimposed

We can show that as long as $p_x > p_c$ then the effect of selecting a probability cutoff p_x produces a much larger variation in ranks in f_b than in f_a. In fact in Appendix A we show that

$$f^{-1}{}_b(p_x) - f^{-1}{}_b(p_x + h) >> f^{-1}{}_a(p_x) - f^{-1}{}_a(p_x + h) \quad (3)$$

In Appendix B we present an example demonstrating this for two idealized frequency distributions.

The implication of this result is as follows: the choice of threshold value can have a big impact on the number of ranks in f_b than in f_b through the use of rank-score graphs. Hence, it has tremendous impact on the performance of the tracker. Hsu and Taksa [7] show that the form of the rank-score graph determines whether a rank combination or score combination produces a better result. Hence, our theoretical results predict that for a complex environment, the benefit of score based fusions and of rank-based fusions will vary depending on the hypothesis pool pruning threshold.

5. Experimental Investigation

In [5,10] we described the RAF tracking system, which uses color, position and shape to track moving targets. We modified that tracking system to do a comparison of Bayes fusion only versus a combination of Bayes fusion and fusion by rank combination. The two options were evaluated by comparing their result with ground truth data for the video sequence.

5.1. Implementation

Foreground objects are extracted from each frame of the image sequence using the non-parametric background estimation technique of Elgammal et al. [3]. The regions are passed to the three component trackers in the RAF system. Color, location and shape information was collected by applying a tracker-specific measurement:

(a). Color Tracker: $f_{cor}(c_j) = \mu_{rg}(c_j)$, average normalized RGB color of c_j.

(b). Location Tracker: $f_{loc}(c_j)$ = the image location of the centroid of c_j.

(c). Shape Tracker: $f_{sha}(c_j) = area(c_j)$, the area of the image covered by c_j in pixels.

For each frame i in the video sequence, a common MHT based hypothesis generation module associates these measurements with the set of existing track hypothesis $\mathbf{T}_i$. Any track hypothesis which meets the gating criterion for a component c_j is associated with that region. Each of the three trackers applies its similarity function to determine how well the region fits that target hypothesis. A score for the new track hypothesis is generated based on the original hypothesis score and the similarity value.

The pool of track hypotheses grows combinatorially and needs to be pruned to stay within resource limits. The resource limits are represented by a nominal pool size n_T:

$$(\mid \mathbf{T}_i \mid > n_T) \Rightarrow \textit{Prune } \mathbf{T}_i \textit{ down to size } n_T$$

To get the best track hypotheses for each target candidate set, the scores and hence ranks from each of the separate trackers are fused in two ways.

(i). Average rank fusion: Let $r_{k,f}$ be the rank of track hypotheses t_k according to tracker f:

$$S_{k,avrank} = \frac{1}{3}\,(r_{k,cor} + r_{k,loc} + r_{k,sha})$$

We will refer to this as the *RAF* (combination) score.

(ii). Linear score fusion: Let $s_{k,f}$ be the score for t_k by tracker f and $\sigma_{k,f}$ be the variance:

$$s_{k,linscore} = (q_{k,col}\, s_{k,col} + q_{k,loc}\, s_{k,loc} + q_{k,sha}\, s_{k,sha})$$

where

$$q_{k,f} = \frac{\frac{1}{\sigma^2_{k,f}}}{\frac{1}{\sigma^2_{k,col}} + \frac{1}{\sigma^2_{k,loc}} + \frac{1}{\sigma^2_{k,sha}}}$$

We will refer to this as the *Bayes* (combination) score.

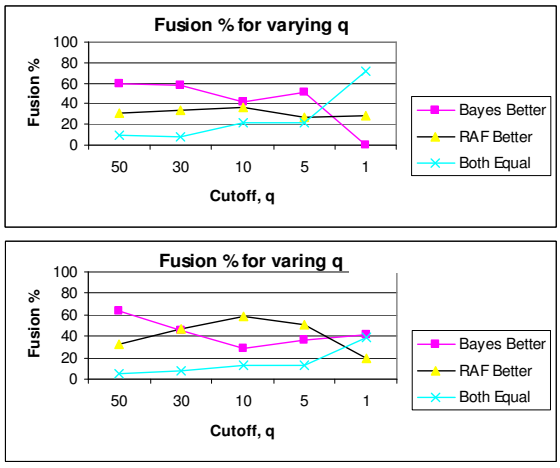
The top scoring track hypotheses for each target are then evaluated against the ground truth data. The evaluation criterion is a performance value calculated as a sum of squared differences between the image location of each component for an image in the sequence in a ground truth target and the component in that image for the track hypothesis. Whichever fusion scores *lower* by this measure is considered the better fusion and this is the one adopted for this target. If both score the same, then the Bayes score was used. Different fusions may be adopted for different targets, and of course, a track hypothesis might have several different fusions used on it over the course of successive pruning events. The image sequence index number and type of fusion used is recorded for each track hypothesis.

Once the fusion calculation is completed, the top scoring track hypotheses for each target are kept, the rest are deleted, and the tracking continues to the next frame and window.

5.2. Results

The RAF tracking system was used to perform the evaluation above on a number of video sequences. The results from two such sequences is presented here. Video sequence 1 and 2 were indoor sequences. Sequence 1 was 50 frames and sequence 2 was 100 frames long at approx. 10 fps. Sequence 1 was of a single, unoccluded moving target. Sequence 2 was of two targets in a complicated mutual occlusion.

Figure 8 shows the results of the comparison for each video sequence. For example, in 8(a) for a cutoff value of 50 (the best 50 hypothesis preserved), the Bayes score produced a result closer to the ground truth in 59% of all fusion operations, the RAF combination produced a better result in 31% of all operations, and in 10% both were equal. The fact that the combination was superior in any fusions at all indicates that rank combinations can improve tracking performance. In Sequence 1, the Bayes score is better for all values of the cutoff above 1. However, in Sequence 2, we note that depending on the cutoff value, sometimes Bayes alone has the high percentage of better fusions and sometimes the RAF combination is better.



(a) Sequence 1 (top) **(b)** Sequence 2 (bottom)

Figure 8: Graph of %Better against cutoff size

6. Selection of Cues for Fusion

If we have a set of tracking systems, as in Fig. 1, each with its own rank-score graph f_k , then we propose that a better fusion result is obtained when a subset of features with different rank-score graphs is combined than when a subset with similar rank-score graphs are combined. Recall that the rank-score graph is monotonic. A concave-up rank-score graph (f_2 in Fig. 5) would assign few ranks to the top scoring tracks and many to the lower scoring tracks, whereas a concave-down rank-score graph (f_1 in Fig. 5) would assign many ranks to the top scoring tracks and few to the lower scoring tracks. We will refer to concave-up and down members of this family as complementary graphs with respect to the ideal rank/score graph f_3 in Fig. 5. Trackers with complementary rank-score graphs should be distinguished from trackers whose output is negatively correlated. The latter is a relationship between the scores the trackers assign to a specific track, whereas the former is a relationship between scoring behaviors, irrespective of the

track being scored. Trackers may be correlated or independent and still have complementary rank-score graphs.

In general each tracking system TR_k of Fig.1 may have a complicated relationship between the score associated with a track hypothesis and the real underlying probability of that track. The result is that the measured rank-score graph for TR_k may differ significantly from the graph predicted from the probability distribution. Figure 9 illustrates two complementary rank-score graphs, f_1, and f_2, versus the actual rank-score graph f_i for the target rich case. We will refer to this actual graph as the ideal graph. Let us consider the effect of pruning the set of track hypotheses at a cutoff probability p_x.

If the actual rank-score graph f_i was known, then this would produce a cutoff at rank r_x of the set of track hypotheses. However, each tracker will have a different, and maybe limited, view of the problem, and may end up with a rank-score graph either of the form of f_1 or f_2. With f_1 a rank cutoff of $r_h > r_x$ is produced. This would be equivalent to a probability cutoff of p_l if the ideal rank-score graph was used. Similarly with f_2 the rank cutoff is $r_l < r_x$ which is equivalent to p_h with the ideal rank-score graph.

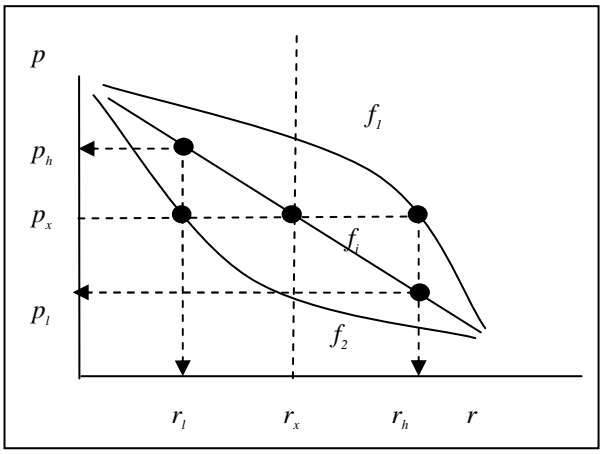


Figure 9: Actual (f_1,f_2) vs. Ideal (f_i) Rank-score Graphs

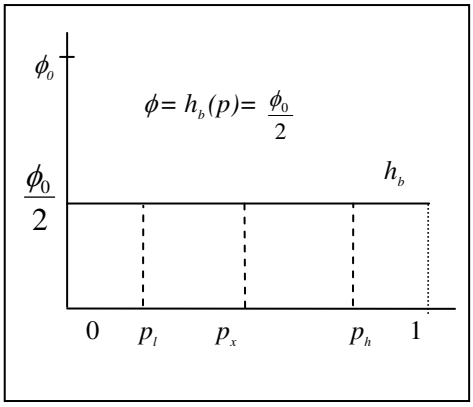


Figure 10: Complex Environment Distribution with Cutoff Points using f_1 (p_l) and f_2 (p_h)

Fig. 10 shows the effect of selecting a probability cutoff p_x when using f_1 or f_2 instead of the ideal rank-score graph f_i for the target rich case. We will take the special case $\phi = h_b(p) = \phi_0/2$ as an example. Fig. 11(a) shows a standard classification Venn diagram. We can use this to quantify the effects of a tracker using a rank-score graph of the form f_1 or f_2 instead of the actual rank-score graph f_i. Let us assume that a probability cutoff p_x is selected for which the False Positives (FP) and False Negatives (FN) are zero, meaning that p_x perfectly separates the track hypotheses into true and false cases (Fig. 11(b)).

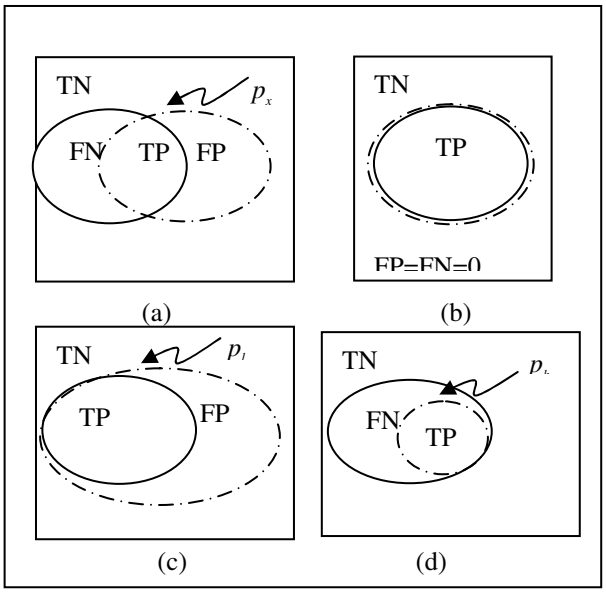


Figure 11: Classification cases: General (a), Ideal (b), using γ_1 (c) and using γ_2(d)

A tracker using p_x and rank-score graph f_1 is cutting at a rank r_h which, as shown before, is equivalent to cutting the actual rank-score graph at $p_l < p_x$. This results in the classification diagram in Fig. 11(c), where the FPs have been increased from zero. From the distribution in Fig. 10 we have that

$$TP = \frac{\phi_0(1-p_x)}{2} \text{ and } FP = \frac{\phi_0(p_x - p_l)}{2},$$

therefore

$$\frac{FP}{TP} = \frac{p_x - p_l}{1-p_x}.$$

And using the rank-score graphs in Fig. 9 we can say

$$\frac{FP}{TP} = \frac{p_x - f_i(r_h)}{1-p_x} = \frac{p_x - (f_i \circ f_1^{-1})(p_x)}{1-p_x}. \quad (4)$$

Similarly, a tracker using p_x and rank-score graph f_2 is cutting at a rank r_l which, as shown before, is equivalent to cutting the actual rank-score graph at $p_h > p_x$. This produces the classification diagram shown in Fig. 11(d) where the FNs have been increased from zero. In this case we have that

$$TP = \frac{\phi_0(1-p_x)}{2} \text{ and } FN = \frac{\phi_0(p_h - p_x)}{2},$$

and

$$\frac{FN}{TP} = \frac{p_h - p_x}{1-p_x} = \frac{f_i(r_l) - p_x}{1-p_x} = \frac{(f_i \circ f_2^{-1})(p_x) - p_x}{1-p_x}. \quad (5)$$

We can conclude from this that when a tracking module employs a rank-score graph that is different from the actual

rank-score graph, it will increase its FPs or FNs such as formula (4) and (5) for f_1 and f_2 respectively. However, it is possible to fuse results from more than one tracking model to alleviate this effect.

If the *complementary* rank-score graphs f_1 and f_2 are added together they produce a rank-score graph that is much more similar to f_i and will hence minimize the false positives and false negatives with respect to f_i. We will define a fused rank-score graph f_* simply as (and there are other definitions possible):

$$f_*(r) = \frac{f_1(r) + f_2(r)}{2}. \qquad (6)$$

And we note that for the case $f_2 = 2f_i - f_1$ (i.e., f_1 is f_2 mirrored in f_i), we have from (4):

$$\frac{FP}{TP} = \frac{p_x - (f_i \circ f_*^{-1})(p_x)}{1 - p_x}$$

$$= \frac{p_x - (f_i \circ f_i^{-1})(p_x)}{1 - p_x} = \frac{p_x - p_x}{1 - p_x} = 0,$$

and similarly from (5):

$$\frac{FN}{TP} = \frac{(f_i \circ f_*^{-1})(p_x) - p_x}{1 - p_x}$$

$$= \frac{(f_i \circ f_i^{-1})(p_x) - p_x}{1 - p_x} = \frac{p_x - p_x}{1 - p_x} = 0.$$

In general, two rank-score graphs won't be perfectly complementary as above, but if their sum is closer to actual rank-score graph, then the FPs or FNs will be reduced. From the expressions above, we can say that

If $(f_i \circ f^{-1}{}_*)(p_x) < (f_i \circ f^{-1}{}_2)(p_x)$
then *FN* will be reduced with fusion, and
If $(f_i \circ f^{-1}{}_*)(p_x) > (f_i \circ f^{-1}{}_1)(p_x)$
then *FP* will be reduced with fusion.

Hence in choosing a subset of features to fuse when tracking in complex scenes, selecting features with complementary rank-score graphs will produce a better result that minimizes false positives and false negatives.

7. Conclusion

In this paper we show theoretically and experimentally that for an MHT based approach to hypothesis generation and pruning in multitarget tracking, the selection of pruning threshold can have a great impact on the tracker's performance, altering the rank-score characteristics of the pool of hypothesis. This change can affect the benefit of using a score based sensory fusion method over a rank-based method.

The experiments reported here used a ground truth file to evaluate the best selection of rank or score fusion operations. However, the ultimate goal is to build an RAF tracking system that select the fusion strategy and apply the optimal fusion operator to each candidate target set on a frame by frame basis. The evaluation in this paper has been a step in this direction. Our next step is to evaluate the uses of the pruning threshold information, and the form of the rank-score graph, as a predictor of which fusion operation will yield the better result.

We also present an approach to selecting which subset of features to fuse for best results. We present a criterion to use to determine whether a fusion between two features will reduce the classification error in tracking. This also uses substantially the rank-score graph concept, and our goal is to also evaluate this experimentally. We will also explore the case where the ideal rank-score graph may not be the simple linear function as in f_3 in Fig. 5 or f_i in Fig. 9. In this case, the concept of complimentary graphs is generalized with respect to a new ideal rank-score graph f_i^*. Moreover the notion of concave-up and concave-down will have to be defined relative to this new ideal rank-score graph f_i^*.

References

1. Bar-Shalom, Y. and Fortmann, T., *Tracking and Data Association*. 1988: Academic Press.
2. Cox, I.J. and Hingorani, S.L. *An Efficient Implementation and Evaluation of Reid's Multiple Hypothesis Tracking Algorithm for Visual Tracking*. *Int. Conf. on Pattern Recognition* (1994) pp.437-442.
3. Elgammal, A., Harwood, D., Davis, L.S.,. *Nonparametric Model for Background Subtraction*. in *Proc. 6th European Conference on Computer Vision*. 2000.
4. Ginn, C., Willett, P., and Bradshaw, J., *Combination of Molecular Simularity Measures using Data Fusion.* Perspectives in Drug Discovery and Design, 2000. **20**: pp. 1-16.
5. Hsu, D.F., Lyons, D.M., Usandivaras, C., and Montero, F. *RAF: A Dynamic and Efficient Approach to Fusion for Multi-target Tracking in CCTV Surveillance*. *IEEE Int. Conf. on Multisensor Fusion and Integration*. Tokyo, Japan; (2003) pp.222-228.
6. Hsu, D.F., Shapiro, J., and Taksa, I., *Methods of Data Fusion in Information Retreival: Rank vs. Score Combination*. 2002, DIMACS TR 2002-58.
7. Hsu, D.F. and Taksa, I., Comparing rank and score combination methods for data fusion in information retrieval, *to appear: Information Retrieval 2004*.
8. Hu, W.; Tan, T.; Wang, L.; Maybank, S., *A Survey on Visual Surveillance of Object Motion and Behaviors*, Part C, IEEE Trans. on Systems, Man and Cybernetics, Volume 34, Number 3, Aug. 2004 pp.334 – 352.
9. Lee, J.H. *Analysis of Multiple Evidence Combination*. *Annual Int. ACM SIGIR Conf. on Research and Development in Information Retrieval*. Philadelphia PA (1997) pp.267-276.
10. Lyons, D., Hsu, D.F., Usandivaras, C., and Montero, F. *Experimental Results from Using a Rank and Fuse Approach for Multi-Target Tracking in CCTV Surveillance. IEEE Intr. Conf. on Advanced Video & Signal-Based Surveillance*. Miami, FL; (2003) pp.345-351.
11. Moore, J.R., and Blair, W.D., *Practical Aspects of Multisensor Tracking* in: Multitarget-Multisensor Tracking (Eds. Y. Bar-Shalom, W.D. Blair) Artech House 2000, pp.1-76.
12. Murphy, T.G. and Lyons, D.M. *Combining Direct and Model-Based Perceptual Information Through Schema Theory*. *IEEE Int. Symp. on Computational Intelligence*

in Robotics & Automation. Monterey, CA; (1997).
13. Ng, K.B., Kantor, P.B., *Predicting the effectiveness of Naive Data Fusion on the basis of system characteristics.* Jour. American Soc. Info. Sc., 2000. **51**(13): pp. 1177-1189.
14. Rasmussen, C., and Hager, G., *Joint Probabalistic Techniques for Tracking Multi-Part Objects*. *Proc. Computer Vision & Pattern Recognition*. Santa Barbara, CA; (1998) pp.16-21.
15. Reid, D.B., *An Algorithm for Tracking Multiple Targets.* IEEE Trans. on Aut. Control, 1979. **AC-24**(6): pp. 843-854.
16. Schrater, P.R. *Bayesian data fusion and credit assignment in vision and fMRI analysis*. *SPIE Int. Symposium on Electronic Imaging Vol #5016*. Santa Clara, CA; (2003).
17. Sharma, R.K., *Probabilistic Model-Based Multisensor Image Fusion*. Ph.D. Diss. Oregon Grad. Institute of Science and Technology: Portland, OR, 1999.
18. Triesch, J., and von der Marlsburg, C. *Democratic Integration: Self-Organized Integration of Adaptive Clues.* Neural Computation 13, 2001, pp.2049-2074.
19. Varshney, P.K., *Special Issue on Data Fusion.* Proceedings of the IEEE, 1997. **85**(1).
20. Vogt, C.C., and Cotrell, G.W., *Fusion via a linear combination of scores.* Information Retrieval, 1999. **1**(3): pp. 151-172.
21. Willett, P., Ruan, Y., Streit, R., *PMHT: Problems and Some Solutions.* IEEE Trans. On Aerospace and Electronic Systems, V38, N3, pp.738-754.
22. Williams, J.L., and Maybeck, P.S., "Cost-function-based hypothesis control techniques for multiple hypothesis tracking," in *Proc. SPIE Signal and Data Processing of Small Targets*, vol. 5428. The International Society for Optical Engineering, April 2004
23. Xu, L., Krzyzak, A., and Suen, C.Y., *Method of Combining Multiple Classifiers and their Application to Handwriting Recognition.* IEEE Trans. SMC, 1992. **22**(3): pp. 418-435.

APPENDIX A: Proof of Formula (3) in Section 4.2

Using the definition in (1), we have

$f^{-1}{}_b(p_x) - f^{-1}{}_b(p_x + h)$

$= \int_{p_x}^{1} h_b(p)dp - \int_{p_x+h}^{1} h_b(p)dp$ *(definition)*

$= \int_{p_x}^{p_x+h} h_b(p)dp$ *(difference)*

$>> \int_{p_x}^{p_x+h} h_a(p)dp$ *(Fig.7, $p_x > p_c$ for same interval h, h_b has greater area)*

$= \int_{p_x}^{1} h_a(p)dp - \int_{p_x+h}^{1} h_a(p)dp$ *(difference)*

$= f^{-1}{}_a(p_x) - f^{-1}{}_a(p_x + h)$ *(definition)*

APPENDIX B: Example.

As an example, suppose that the actual distributions for two cases were $\phi_a(p) = (-\phi_0)p + \phi_0$ and $\phi_b(p) = \frac{\phi_0}{2}$ in Fig. 3. First, we establish the rank-score graphs for each case as in Fig. 7.

$f^{-1}{}_a(p) = \int_p^1 \phi_a(p)dp = \frac{(1-p)(-\phi_0 p + \phi_0)}{2}$

$= \frac{\phi_0}{2}(p-1)^2$

Solving for p and letting $\mu_a = f^{-1}{}_a$, we have

$\sqrt{\frac{2\mu_a}{\phi_0}} = 1 - p$

$p = 1 - \sqrt{\frac{2}{\phi_0}}.\mu_a^{\frac{1}{2}}$

(a parabolic curve) and

$\frac{dp}{d\mu_a} = \frac{1}{\sqrt{2\phi_0\mu_a}}$

$f^{-1}{}_b(p) = \int_p^1 \phi_b(p)dp = (1-p)\frac{\phi_0}{2}$

$= \frac{\phi_0}{2}(1-p)$

Solving for p and letting $\mu_b = f^{-1}{}_b$, we have

$\frac{2\mu_b}{\phi_0} = 1 - p$

$p = 1 - \frac{2\mu_b}{\phi_0}$

(a straight line) and

$\frac{dp}{d\mu_b} = \frac{-2}{\phi_0}$

The rank difference in each graph given a cutoff probability p_x is therefore*:*

$f^{-1}{}_a(p_x) = \frac{\phi_0}{2}(p_x - 1)^2$

$f^{-1}{}_a(p_x+h) = \frac{\phi_0}{2}(1-(p_x+h))^2$

$f^{-1}{}_a(p_x) - f^{-1}{}_a(p_x + h)$

$= \frac{\phi_0}{2}[(1-p_x)^2 - (1-(p_x+h))^2]$

$= \frac{\phi_0}{2}[2h(1-p_x) - h^2]$

$f^{-1}{}_b(p_x) = \frac{\phi_0}{2}(1-p_x)$

$f^{-1}{}_b(p_x+h) = \frac{\phi_0}{2}(1-(p_x+h))$

$f^{-1}{}_b(p_x) - f^{-1}{}_b(p_x + h)$

$= \frac{\phi_0}{2}[(1-p_x) - (1-(p_x+h))]$

$= \frac{\phi_0}{2}[h]$

Inspecting the ratio of these two rank differences, we get:

$\frac{f_b^{-1}(p_x) - f_b^{-1}(p_x+h)}{f_a^{-1}(p_x) - f_a^{-1}(p_x+h)} = \frac{\frac{\phi_0}{2}[h]}{\frac{\phi_0}{2}[2h(1-p_x) - h^2]} = \frac{1}{2(1-p_x) - h}$

and

$\frac{1}{2(1-p_x) - h} > \frac{1}{1-h}$ when $p_x > ½$

> 1 since $h > 0$.