Rank-Based Multisensory Fusion in Multitarget Video Tracking

Damian M. Lyons and D. Frank Hsu
Robotics and Computer Vision Laboratory
Department of Computer and Information Science
Fordham University
Bronx NY 10458
{dlyons,hsu}@cis.fordham.edu

Abstract
An attractive approach to improve tracking performance for visual surveillance is to use information from multiple visual sensory cues such as position, color, shape, etc. Previous work in fusion for tracking has tended to focus on fusion by numerically combining the scores assigned by each cue. We argue that for video scenes with many targets in a crowded situation, the splitting and merging of regions associated with targets, and the subsequent dramatic changes in cue values and reliabilities, renders this form of fusion less effective.

In this paper we present experimental results showing that use of cue *rank* information in fusion produces a significantly better tracking result in crowded scenes. We also present a formalization of this fusion problem as a step in understanding why this effect occurs and how to build a tracking system that exploits it.

1. Introduction.

Automated tracking of targets in video remains a difficult problem, especially when dealing with crowded scenes [8]. A video image can be a very rich source of information about a target: image position, image velocity, color properties, shape properties and so forth. Fusing multiple sources of sensory information therefore is an intuitively appealing way to make tracking more robust [26].

Comaniciu et al. [2] partition video tracking into two components: a target location and identification component and a filtering and data association component. Their work falls into the first component, as does much video tracking work [6, 7, 16]. In this paper we are concerned with the *second component* and the way in which it can support the fusion of information from multiple cues. Much of the work in this area has been inspired by Bar-Shalom et al. [1]. Variants of algorithms such as MHT [3] and JPDAF [20] have been applied to video tracking. Under certain assumptions of linearity and Gaussian noise, an optimal Bayesian fusion operation can be derived where a fused estimate is a linear combination of the local estimates and where the combination coefficients are inversely proportional to the variance of the local estimates [22]. As a target is tracked from frame to frame, the differences between the expected value of each cue for that target and the measured value can be collected. For each cue, the variance of this measure is inversely proportional to the 'reliability' of the cue for identifying the target.

This approach works best in tracking a target that is well segmented from its background and does not engage in many occlusions with other targets. If, however, we take a crowded scene, then we expect the targets to be close, perhaps moving as variably sized groups, and with substantial periods of mutual partial occlusion. Under these circumstances the measured cue values for each target change non-linearly, and the variances become less reliable in determining how useful a cue is for identifying the target. Nonetheless, the cues retain some value in identifying the target but the problem we face is how to combine the cues given that we know the cue values are changing in a complicated way due to the splitting and merging of the image regions associated with the targets in these crowded scenes.

Xu et al. [27] divide pattern classifiers into three levels depending on the nature of their output: The abstract level, where the output is the name of the class; the rank level, where the output is a ranked list of classes; and, the measurement level, where the output is a set of scores for each class. In doing multiclassifer fusion, *voting* is an appropriate fusion approach for level 1; *voting* and *rank combination* approaches are possible for level 2; and, *voting*, *rank-combination* and *score-combination* approaches are possible for level 3. The video tracking fusion problem can be treated as an example of this last level. Each multisensory feature or cue can be considered as an expert, classifying image regions to targets. The Bayesian approach to fusion is therefore one example of a score-combination approach, where the classifier output value is used. However, we note that voting and rank-combination are also fusion options for fusing these level 3 classifiers, and that these approaches may be less sensitive to the dramatically changing cue values we see in cluttered scenes as they rely less on the exact score value.

This paper presents experimental results that support the theory that in crowded scenes rank-based fusion helps produce a more accurate track. Section 2 gives a review of related literature. In Section 3, we describe an experiment to evaluate the performance of a rank fusion (average rank) and the Bayes fusion (linear score combination with coefficients inversely proportional to variances). In Section 4 and 5 we formalize the rank and fusion problem, to explain why in certain cases rank-based fusion can improve on a score-based fusion using the concept of the rank-score graph (Hsu et al [11,12]). Section 6 concludes the paper with a discussion of our results and next steps.

2. Literature Review.

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 as

$$P(S \mid I) = \frac{P(I \mid S)P(S)}{P(I)} = \frac{P(S)}{P(I)} \prod_{i=1}^{n} P(f_i \mid S)$$

where f_i , $i=1..n$ are the independent feature or cue measurements. In the standard framework for linear estimation, this gives rise to an estimate for S that is a linear combination of the cues where the combination coefficients are inversely proportional to the variance. We will refer to this in the rest of the paper as *Bayes fusion*. This is a powerful approach and there is evidence that human perception employs it for some tasks [5]. Loy et al. [14] use a particle filter approach to represent multiple target hypotheses. To fuse their multiple visual cues, they employ a weighted sum of cues, where each cue is weighted by a reliability coefficient. (The reliability is also used to allocate computational resources across cues.) Snidaro et al. [23] use an appearance ratio (AR) to determine the reliability of a sensor. The AR value is used to weight the position estimates from a sensor. Triesch and von der Malsburgh [24] again define fusion as a weighted sum of local cue measurement, where each cue estimate is weighted by a reliability coefficient. However, the dynamics of the reliability coefficients are phrased more generally than the inverse variances of the linear estimation case, leading to majority consensus style of fusion.

For video tracking, we can consider that each feature measurement is analyzed by an automated expert, a classifier, that produces an estimate of the probability *P(S|I)* and these local estimates are fused to produce a global estimate [27]. A typical fusion operation in this case is to average the probabilities [13]. This is a weighted score combination. If the outputs of *N* experts are averaged, then the fused error rate can be reduced by a factor of *N* [25] provided the component errors are uncorrelated. Non-linear fusions have been proposed for pattern classifiers, including *voting* and *rank combination* [13,17]. Hsu et al. [11] and Hsu and Taksa [12], using the concept of the rank-score graph, show that, under certain conditions, rank combination outperforms score combination in the fusion of information retrieval systems. Melnik et al. [17] point out that rank combination has an important role even for measurement level classifiers, in that it can be used to normalize the outputs from a set of classifiers that produce very different kinds of outputs.

3. Fusion Comparison Experiments.

In the experiments presented here, we obtained ground truth information for ten video sequences, showing a variety of targets, backgrounds and tracks. The targets are not always separated easily from the background or each other, and are from time to time close enough to each other to cause recurrent partial occlusions. In [10, 15], we describe a tracking system called "Rank and Fuse" tracking (RAF), designed to evaluate score-based, rank-based, and various combinations of these approaches to fusing color, position and shape for video tracking of multiple targets in crowded scenes. The RAF tracking software was modified here to carry out two fusion operations: a score fusion using a Mahalanobis distance and a rank fusion using an average rank distance, both described in more detail below.

Sequence	Description
1	1 moving target, indoors
2	2 slowly crossing targets, indoors
3	1 moving target, outdoors
4	3 moving targets, outdoors, non-adjacent
5	2 quickly crossing targets, outdoors
6	3 moving targets, outdoors, 2 quickly crossing
7	2 adjacent moving targets, outdoors
8	4 moving targets, outdoors, 2 overlapping
9	3 targets moving as a crowd, outdoors
10	7 targets moving as a crowd, outdoors

Table 1: Description of Video Sequences

The tracker was run *twice* on each video sequence. In RUN1 only score fusion was carried out. The top m=30 tracks produced by tracker were evaluated against ground truth using a Mean Sum of Squared distances (MSS distances):

$$\frac{1}{nm} \sum_{j} \sum_{i} (gp_i - tp_{ij})^2$$

where gp_i $i=1..n$ is the ground truth sequence of target centroid image locations and tp_{ij} $i=1..n$ is the *jth* best track's sequence of target centroid image locations.

In RUN2, the tracker was allowed to evaluate *both* fusion operators whenever a fusion needed to be performed. The fusion operator that produced the better MSS distance value on its top 30 tracks at that point in the tracking process was then selected. When the tracker finished, the MSS distance measure was again collected. In addition, the top 30 tracks for each target were examined to count which fusion operators had been used for each fusion.

3.1 Implementation.

Foreground objects are extracted from each frame of the image sequence using the non-parametric background estimation technique of Elgammal et al. [4]. The regions are passed to the three component trackers in the RAF system. Color, location and shape information are collected by applying a tracker-specific measurement:

1. Color Tracker: $f_{cor}(c_j)$, average normalized RGB color of c_j.
2. Location Tracker: $f_{loc}(c_j)$, image location of the centroid of c_j.
3. Shape Tracker: $f_{sha}(c_j)$, 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 $\mathbf{T}_i$. The gating function is that a track hypothesis be within a standard deviation of the predicted position p_k for target k:

$$(p_k - f_{loc}(c_j))^2 < \sigma_k^2$$

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:

$$|\mathbf{T}_i| > f_T n_T \quad \Rightarrow \quad \textit{Prune } \mathbf{T}_i \textit{ down to size } n_T$$

The values $n_T=100$, $f_T=2.5$ are used here.

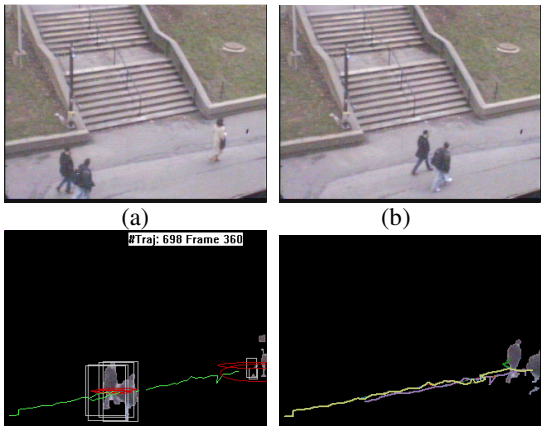
To get the best track hypotheses for each target candidate set, the scores from each of the separate trackers are fused in two ways.

1. Bayes score fusion (*BS*): Let $s_{k,f}$ be the score for t_k by tracker f and $\sigma^2_{k,f}$ be the variance:
$$s_{k,bs} = (q_{k,col} s_{k,col} + q_{k,loc} s_{k,loc} + q_{k,sha} s_{k,sha})$$
where
$$q_{k,f} = \frac{\frac{1}{\sigma^2_{k,f}}}{\frac{1}{\sigma^2_{k,col}} + \frac{1}{\sigma^2_{k,loc}} + \frac{1}{\sigma^2_{k,sha}}}$$
2. Average rank fusion (*AR*): Let $r_{k,f}$ be the rank of track hypothesis t_k according to tracker f:
$$s_{k,ar} = \frac{1}{3} (r_{k,col} + r_{k,loc} + r_{k,sha})$$

In RUN1, only the BS fusion is used. In RUN2, both fusions are evaluated. For each target, for each fusion, the top scoring 30 track hypotheses are evaluated against the ground truth data using the MSS distance measure as described before. 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 is 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. Once the fusion calculation is completed, the top scoring track hypotheses for each target are kept, the rest are deleted, and the tracking continues.

3.2 Results.

The combined MSS Distance average and variance for RUN1 (BS only run), and for RUN2 (mixed BS and AR run), are shown in Table 2. The average MSSD for RUN2 is smaller than that for RUN1 in all cases indicating that on average the tracks produced were closer to the ground truth.



(a) (b) (c) (d)

Figure 1: Example Frames from Video Sequence (a,b); Example of Occlusion during Tracking (c); Display of Final Top Tracks (d).

Seq	RUN1 MSSD Avg.	RUN1 MSSD Var.	RUN2 MSSD Avg.	RUN2 MSSD Var.	t Val.
1	1537.22	694.47	1536.65	695.49	0.1
2	816.53	8732.13	723.13	3512.19	4.62
3	108.89	61.61	108.34	60.58	0.23
4	23.14	2.39	23.04	2.30	0.17
5	201.35	452.42	201.20	450.96	0.02
6	154.76	113.88	151.75	101.06	0.87
7	256.87	83.40	253.94	83.52	1.24
8	96.40	119.22	66.90	12.90	8.1
9	647.31	174.74	622.45	119.24	14.1
10	538.35	605.84	500.90	557.91	6.8

Table 2: MSSD Results (significance > 95% shaded)

Sequence	% BS Better	% AR Better	% Equal
1	58	34	8
2	12	88	0
3	55	45	0
4	83	16	1
5	69	31	0
6	51	49	0
7	27	67	6
8	40	60	0
9	39	54	7
10	36	44	20

Table 3: Fusion Comparison Results for RUN2.

Of course it is possible that the difference in MSS distance measurements and the selection of AR over BS was due to chance. To address this, we calculate the t-test statistic [19] for these two distributions. The lines shown shaded (i.e., sequences 2, 8, 9, and 10) showed a significance level of 95% or greater). Looking back to Table 1, these sequences are those that have a number of crossing targets and resulting partial occlusion). Table 3 shows the breakdown of fusion types in RUN2 for the top 30 tracks for all targets.

3.3 Discussion.

This experiment demonstrates that rank fusion can be valuable in tracking when used at the right

time in the video sequence. The experiment suggests that the right time is when there are visually overlapping targets. The formalization and results presented in the next section are a first step to explain why we see the effects shown in Table 2, namely why in certain cases rank-based fusion can be better than a score-based fusion.

4. Formalizing the Problem

Let us consider a multitarget video tracking module (tracker) the output of which is a list of tracks for each target, with a score associated with each track. The better the score (and rank) of a track, the more the tracker supports the hypothesis that, based on the evidence to this point, this is the correct track for the target. Now let us consider a set of such tracking modules, $TR_1,...,TR_m$, each using different sensing modalities and/or tracking approaches to determine its list of tracks (Fig. 2).

We will assume that each tracker operates on the same pool of track hypotheses. This could be by use of a common hypothesis generation stage [15] (which it is in our case) or by the generation of a set of composite tracks [18]. Let $\boldsymbol{T} = \{0,...,N-1\}$ be the labels for the pool of N track hypotheses generated by the system of tracking modules over a (possibly varying) window of time w.

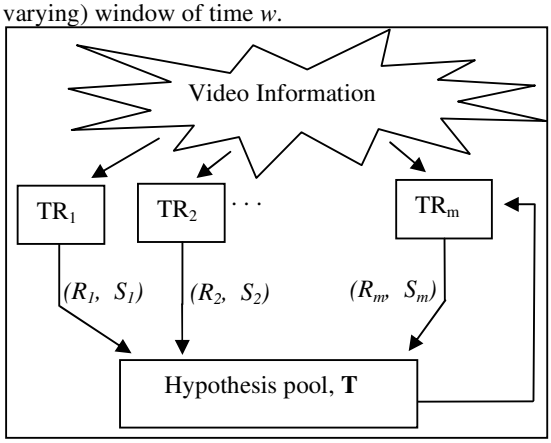


Figure 2: Multiple Tracker Configuration

Our objective is to assign a *fused* score to each hypothesis in the pool by applying a fusion operation to all the tracker scores for that hypothesis.

4.1 Rank and Score Functions.

Consider a single tracker: Let s be the score function for that tracker, $s : T \rightarrow \{1,...,S_{max}\}$ 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 list is the pool of track hypotheses collected during the time window w. Let r be the rank function $r : T \rightarrow \{1,...,N\}$ where $r(i)$ is the *rank* of the track i. The track hypothesis with highest score is the one with best rank, i.e., with rank equal to 1. We need to constrain the rank to reflect the score as follows:

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

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

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

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.

We will denote the combination of score and ranking function as Φ, and the output of each tracker is written $\Phi = (R, S)$, where R is a collection of rank functions r_j, one per target j, and S is a collection of score functions s_j. We will write the rank and score functions for a target j as $\Phi_j = (r_j , s_j)$.

4.2 Relating Score to Probability.

Consider the output of a single tracker in Fig. 2. $P(t_{ij} \mid I)$ is the a-posteriori probability that track hypothesis i is the track of target j given the image information I. Since all trackers share the same image information and agree on the set of track hypotheses, we instead consider $P(t_{ij} \mid \Phi_j)$, the probability that track hypothesis i is the track for target j given that the tracker output is Φ_j. Bayes rule relates this to the likelihood $P(\Phi_j \mid t_{ij})$, that the tracker produces its score and ranking output given that the correct hypothesis for target j is i.

$$P(t_{ij} \mid \Phi_j) = \frac{P(\Phi_j \mid t_{ij}) P(t_{ij})}{P(\Phi_j)} \qquad (1)$$

$P(t_{ij})$ is the a-priori probability of track i being a track for target j. We will assume that all tracks are equally likely for a target. $P(\Phi_j)$ is the a-priori probability of the tracker producing a rank and score combination Φ_j. We will again assume all are equally likely, and hence $P(t_{ij} \mid \Phi_j)$ is directly proportional to the likelihood $P(\Phi_j \mid t_{ij})$. Let K be the constant of proportionality.

The conditional probability of generating Φ_j given that i is the correct track for target j is reflected in the score and rank of i. If the measurements so far indicate that i is the correct track for target j, then i will have a better rank and score. Let f be a function that relates the score of a track, $s_j(i)$, to the probability of that rank and score being produced by a tracker given i as the correct track for target j.

$$P(\Phi_j \mid t_{ij}) = f(s_j(i)) \qquad (2)$$

To identify targets in a video image, a series of measurements are made on the image. These measurements are used to decide which part, if any, of the image corresponds to which, if any, target. We use the term score to refer to the number obtained from the measurements, and we investigate two cases: where the measurements reflect the probability

in a straightforward *linear* fashion, and where there is a less straightforward, *non-linear* relationship.

5. Linear vs. Non-Linear Score Relationship

5.1 Linear Score Relationship.

Consider first the case, where f in (2) above is a *linear function* of score

$$P(\Phi_j \mid t_{ij}) = C_1 s_j(i) + C_2. \quad (3)$$

where C_1 and C_2 are constants. Substituting (3) back into (1) we get

$$P(t_{ij} \mid \Phi_j) = K\ (C_1 s_j(i) + C_2) = C_3\ s_j(i) + C_4$$

In the case where there are m trackers and each produces a rank and score output Φ_{kj} for $k \in \{1,..,m\}$ as in Fig. 2, then we can consider each $P_k(t_{ij} \mid \Phi_{kj})$ as the evidence that tracker k believes that track i is the correct track for target j. We will adopt a simple score fusion operation for the purpose of this outline: Given the input from each of these local "experts" then a better estimate (reduced error [25]) of the probability of t_{ij} can be obtained with a linear combination that averages the component estimates.

$$P(t_{ij} \mid \Phi_{*j}) = \frac{1}{m}\sum_k P_k(t_{ij} \mid \Phi_{kj})_k = \frac{1}{m}\sum_k C_{3k} s_{jk}(i) + \frac{1}{m}\sum_k C_{4k} \quad (4)$$

5.2 Non-Linear Score Relationship.

Now consider the case where the relationship f between the score and probability in (2) is not linear, but is any function g. The only constraint we place on g is that it be monotonic, a weaker constraint than linearity.

$$P(\Phi_j \mid t_{ij}) = g(s_j(i)) \quad (5)$$

Substituting (5) into (1), as we did for (3), we obtain

$$P(t_{ij} \mid \Phi_j) = K\ g(s_j(i)) = C_R\ \ g(s_j(i))$$

We again sum and normalize for a fused estimate:

$$P(t_{ij} \mid \Phi_{*j}) = \frac{1}{m}\sum_k P_k(t_{ij} \mid \Phi_{kj}) = \frac{C_R}{m}\sum_k g_k(s_{kj}(i)) \quad (6)$$

However, now when the scores are summed, it is not the same as summing the probabilities, since in (6) the probabilities may have been transformed by g in a non-linear fashion to yield the score.

1) The Rank-Score Graph. Consider the following example. A particular expert may habitually give very high scores to its top two ranked candidates and very low scores to all the rest. Another expert may habitually assign its scores in linear fashion from highest to lowest. Averaging the scores from the two experts will always give the first expert's top candidates higher 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 [11, 12, 17]. Irrespective of score, all first ranked candidates will be combined with equal weight, and so on.

Hsu et al. [11] and Hsu and Taksa [12] characterize the relationship that an expert habitually produces between score and rank as the *graph of the rank-score function* h: $\{1,\ldots,N\} \rightarrow \Re$, a monotonic function that relates rank and score:

$$h(r_j(i)) = s_j(i) \quad (7)$$

The shape of the graph 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., Fig. 3 (h_2)). The expert who habitually assigns higher scores to a small subset of its top ranked candidates will have a graph that is not a straight line, but has a high slope after the first few candidates and a lower slope for the remainder. The concave-up graph h_1 in Fig. 3 is an example of this. A third class of scoring behavior is exemplified by h_3

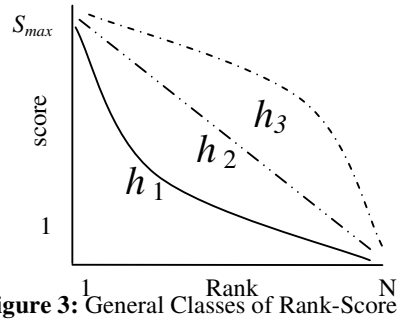


Figure 3: General Classes of Rank-Score Graphs

in Fig 3. In this case, the expert habitually gives higher scores to a larger subset of its top ranked candidates.

2) Fusion by Rank-Sum. We now apply the rank-score concept (7) to (6). We have placed the constraint on g that it be monotonic, that is:

$$s(i) > s(j) \Rightarrow g(s(i)) > g(s(j))$$

Thus if we rank $s(i)$ we will produce the same ranking as if we ranked $g(s(i))$, and a fusion operation based on the rank information of $s(i)$ will be equivalent to one using the rank information of $g(s(i))$. Let h_{p*} be the rank-score function for $P(t_{ij} \mid \Phi_{*j})$, that is, $h_{p*}(P(t_{ij} \mid \Phi_{*j}))$ is the rank of the probability. Let h_{pk} be the rank-score function for $P_k(t_{ij} \mid \Phi_{kj})$ and h_{kj} be the rank-score function for s_{kj}. In that case,

$$\begin{aligned}
&\frac{1}{m}\sum_k h_{pk}^{-1}(P_k(t_{ij} \mid \Phi_{kj})) \\
&= \frac{C_R}{m}\sum_k h_{pk}^{-1}(g_k(s_{kj}(i))) && \text{using (6)} \\
&= \frac{C_R}{m}\sum_k h^{-1}(s_{kj}(i)) && g \text{ monotonic} \\
&= \frac{C_R}{m}\sum_k r_{kj}(i) && \text{using (7)} \\
&= h_{p*}^{-1}(P(t_{ij} \mid \Phi_{*j}))
\end{aligned}$$

The quantity $\frac{1}{m}\sum r(i)$ above is the *average rank* operator of Section 3.1. Thus, although the linear combination of probabilities is no longer equal to the linear combination of scores, the linear combination of the ranks of the probabilities is equal to the linear combination of the ranks of the scores.

6. Conclusion

This paper has reported experimental evidence that the use of rank information in fusion for video tracking produces a significantly better result for a crowded video sequences. It is easy to understand why rank combinations might be overlooked. Given the scores, rank follows with just an ordering operation by sorting the score values, so how can it possibly provide any additional information beyond score? The insight is that when the results being combined are very different, rank combinations outperform score combinations. Our theoretical results capture this effect. In tracking applications without much clutter or target occlusions/crossings for crowded video scenes, there is a simple relationship between the feature measurements and the probabilities. We expect score combinations to work well in these cases. In applications with repeated partial occlusions such as in the video sequence presented here, the effect of the occlusions and crossing is to muddle the relationship between the feature measurements and probabilities in a non-linear fashion. We expect rank combinations to operate better under those circumstances.

Our next step is to verify this application of our theory. We will determine whether indeed the rank combinations are *best in the temporal and spatial vicinity* of occlusions/crossings. In [9], the authors present a dynamic hypotheses pruning strategy for real-time tracking using the RAF system we developed [10,15]. Our ultimate objective with the RAF tracker is to be able to automatically determine which fusion operation is most appropriate given the target and environment conditions, and in this way to construct a tracker that is adaptive, efficient and robust.

References

1. Bar-Shalom, Y. and Fortmann, T., *Tracking and Data Association*. 1988: Academic Press.
2. Comaniciu, D., Ramesh, V., Meer, P., Kernel-Based Object Tracking. IEEE PAMI V25 #5 May 2003 pp.564-577.
3. 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) 437-442.
4. Elgammal, A., Harwood, D., Davis, L.S.,. *Nonparametric Model for Background Subtraction*. in *Proc. 6th European Conference on Computer Vision*. 2000.
5. Fine, I., and Jacobs, R., *Modeling the combination of Motion, Stereo, and Vergence Angle Cues to Visual Dept.* Neural Computation, 1999. **11**: pp. 1297-1330.
6. Gavrila, D., *The Visual Analysis of Human Movement: A Survey.* Comp. Vis. & Image Understanding, 1999. **73**(1): p 82-98.
7. Haritaoglu, I., Harwood, D., and Davis, L. *W4: Who, When, Where, What: A Real-time System for Detecting and Tracking People*. *3rd Int. Conf. on Face and Gesture Recognition* (1998) pp.877-892.
8. Hu, W.; Tan, T.; Wang, L.; Maybank, S., *A Survey on Visual Surveillance of Object Motion and Behaviors* Systems, Man and Cybernetics, Part C, IEEE Transactions on ,Volume: 34 , Issue: 3 , Aug. 2004 Pages:334 – 352.
9. Hsu, D.F., and Lyons, D.M., *A Dynamic Pruning Strategy for Real-Time Tracking.* To appear: IEEE 19th Int. Conf. on Advanced Information Networking and Applications, 2005.
10. 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.
11. Hsu, D.F., Shapiro, J., and Taksa, I., *Methods of Data Fusion in Information Retreival: Rank vs. Score Combination*. 2002, DIMACS TR 2002-58.
12. Hsu, D.F. and Taksa, I., Comparing rank and score combination methods for data fusion in information retrieval, *to appear: Information Retrieval 2004*.
13. Kittler, J., and Alkoot, F., *Sum versus Vote Fusion in Multiple Classifier Systems.* IEEE PAMI, 2003 **25**(1) pp110-115.
14. Loy, G., Fletcher, L., Apostoloff, N., and Zelinsky, A. *An Adaptive Fusion Architecture for Target Tracking*. *Proceedings of the 5th Int. Conf. on Face and Gesture Recog.* Washington DC (2002).
15. 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.
16. Lyons, D.M., *Discrete-Event Modeling of Misrecognition in PTZ Tracking.* IEEE Intr. Conf. Advanced Video & Signal-Based Surveillance, July 21-22, 2003, Miami Beach FL.
17. Melnik, O., Vardi, Y., Zhang, C-H., *Mixed Group Ranks: Preference and Confidence in Classifier Combination.* IEEE PAMI V26, N8, August 2004, pp973-981.
18. 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.
19. Press, W., et al. *Numerical Recipes in C.* Cambridge University Press 2002.
20. 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.
21. 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).
22. Sharma, R.K., *Probabalistic Model-Based Multisensor Image Fusion*, Ph.D. Diss. 1999, Oregon Grad. Inst. Science & Tech.: Portland, OR.
23. Snidaro, L., Foresti, G., Niu, R., Varshney, P. *Sensor Fusion for Video Surveillance*. in *7th Int. Conf. on Information Fusion*. 2004. Stockholm Sweden 25. Triesch, J., and von der Marlsburg, C. *Democratic Integration: Self-Organized Integration of Adaptive Clues.* Neural Computation 13, 2001, pp.2049-2074.
24. Triesch, J., and von der Marlsburg, C. *Democratic Integration: Self-Organized Integration of Adaptive Clues.* Neural Computation 13, 2001, pp.2049-2074.
25. Tumer, K., Ghosh, J., *Linear and Order Statistics Combiners for Pattern Classification.* In: A. Sharkey Ed., Combining Neural Nets, Springer-Verlag 1999, pp.127-162.
26. Varshney, P.K., *Special Issue on Data Fusion.* Proc. IEEE 1997. **85**(1).
27. 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.