Unified Crowd Segmentation

Peter Tu, Thomas Sebastian, Gianfranco Doretto, Nils Krahnstoever,
Jens Rittscher, and Ting Yu

GE Global Research, Niskayuna, NY USA
tu@crd.ge.com

Abstract. This paper presents a unified approach to crowd segmentation. A global solution is generated using an Expectation Maximization framework. Initially, a head and shoulder detector is used to nominate an exhaustive set of person locations and these form the person hypotheses. The image is then partitioned into a grid of small patches which are each assigned to one of the person hypotheses. A key idea of this paper is that while whole body monolithic person detectors can fail due to occlusion, a partial response to such a detector can be used to evaluate the likelihood of a single patch being assigned to a hypothesis. This captures local appearance information without having to learn specific appearance models. The likelihood of a pair of patches being assigned to a person hypothesis is evaluated based on low level image features such as uniform motion fields and color constancy. During the E-step, the single and pairwise likelihoods are used to compute a globally optimal set of assignments of patches to hypotheses. In the M-step, parameters which enforce global consistency of assignments are estimated. This can be viewed as a form of occlusion reasoning. The final assignment of patches to hypotheses constitutes a segmentation of the crowd. The resulting system provides a global solution that does not require background modeling and is robust with respect to clutter and partial occlusion.

1 Introduction

The segmentation of crowds into individuals continues to be a challenging research problem in computer vision [1,2,3,4,5]. The automation of video surveillance systems in public venues such as airports, mass-transit stations and sports stadiums requires the ability to detect and track individuals through complex sites. We identify three challenges that make this problem particularly difficult: *(i) Partial occlusion.* In many crowded scenes people can be partially occluded by others. Monolithic detectors [2,6,7] that model the shape and appearance of an entire person typically fail in such situations and hence cannot reliably detect people in crowded environments. *(ii) Dynamic backgrounds.* When cameras are fixed, statistical background models are commonly used to identify foreground regions [8]. However, this approach fails when the background is dynamic. Further, background modeling is not applicable for moving cameras, such as those mounted on pan tilt devices or mobile platforms. *(iii) Foreground clutter.* The presence of moving non-person objects such as luggage carts, shopping trolleys

D. Forsyth, P. Torr, and A. Zisserman (Eds.): ECCV 2008, Part IV, LNCS 5305, pp. 691–704, 2008.

and cleaning equipment can clutter the foreground of the scene. A robust crowd segmentation algorithm should be immune to foreground clutter without having to explicitly model the appearance of every non-person object.

This paper presents a unified approach to crowd segmentation that effectively addresses these three challenges. The proposed system combines bottom-up and top-down approaches in a unified framework to create a robust crowd segmentation algorithm. We first review a number of relevant approaches.

Low level feature grouping has been used to segment crowds [5, 9]. These approaches take advantage of the fact that the motion field for an individual is relatively uniform and hence tracked corners with common trajectories can be grouped together to form individuals. However, difficulties arise when multiple individuals have similar trajectories. *Monolithic classifiers* capture the shape and appearance space for the whole body using relatively simple learning methods [10,6,7]. The direct application of these classifiers to non-crowded scenes generates reasonable segmentations, however failure modes can occur when partial occlusions are encountered. *Part based constellation models* [11, 12, 13] construct boosted classifiers for specific body parts such as the head, the torso and the legs, and each positive detection generates a Hough-like vote in a parametrized person space. The detection of local maxima in this space constitutes a segmentation. A similar approach [2] uses interest operators to nominate image patches which are mapped to a learned code book. A drawback of these approaches is that the identification of local maxima in the Hough space can be problematic under crowded and cluttered environments - a global approach is required.

The previous approaches can be considered to be bottom-up methods where local context is used. On the other hand, *global approaches* that rely on background segmentation has been proposed in [14, 4]. In [14], Markov Chain Monte Carlo (MCMC) algorithms are used to nominate various crowd configurations which are then compared with foreground silhouette images. However, this form of random search can be computationally expensive. To address this issue an Expectation Maximization (EM) based approach has been developed [4]. In this framework, a hypothesis nomination scheme generates a set of possible person locations. Image features are then extracted from foreground silhouettes and a global search for the optimal assignment of features to hypotheses is performed. The set of hypotheses that receive a significant number of assignments constitute the final segmentation. Reliance on accurate foreground background segmentation is a weakness of both of these approaches.

1.1 Overview of the Unified Approach

In this paper we extend the global EM crowd segmentation framework [4] to use appearance-based features that do not rely on background segmentation. A head and shoulder classifier is used to generate an initial set of hypothesized person locations, a grid of patches are then superimposed on the image. A globally optimal assignment of patches to hypotheses defines the final segmentation. The likelihood of a single patch to hypothesis assignment is evaluated based on local appearance. However, instead of learning an appearance and spatial distribution

model for every possible patch location, we show that a partial response from a monolithic whole body classifier operating solely on a given patch can discriminate between valid and invalid patch assignments. The framework also allows for the inclusion of grouping terms based on low level image cues so that concepts such as uniform motion and intra-garment color constancy can be leveraged. During the *E-step* we estimate a globally optimal assignment of patches to person hypotheses. The *M-step* ensures that globally consistent patch assignments are chosen. This can be viewed as a form of occlusion reasoning.

2 Segmentation

This section provides a detailed overview of the proposed crowd segmentation algorithm. Figure 1 depicts the various stages used to generate the final segmentation of a crowded scene. We assume that the head and shoulders of all detectable individuals can be observed. Hence, an initial set of hypothesized person locations are nominated using a head and shoulders detector (see section 3 for details). These K nominated hypotheses are denoted by $\mathcal{C} := \{c_i\}$. The parameters of this head and shoulders detector are chosen to minimize missed detections, hence many false detections are also generated(see Figure 1a). The scene is partitioned into a set of N rectangular patches $Z = \{z_i\}$, as shown in Figure 1b. The segmentation of the scene into individuals is achieved by a globally optimal assignment of these image patches to the initial hypotheses. The potential assignment of an image patch to a person hypothesis is evaluated using both direct affinity and pairwise affinity terms, as described below.

Let $g_k(z_i)$ denote the affinity associated with the direct assignment of patch z_i to hypothesis c_k. One of the main thrusts of this paper is a novel method for computing this affinity function based on local shape and appearance information - this will be the topic of section 3. Figure 1c illustrates this step for the patch shown in green. The width and color of the arrow connecting the patch to a hypothesis indicates the strength of the affinity. Using camera calibration information and a ground plane assumption, certain direct assignments can be ruled out based on geometric reasoning (shown with black arrows).

Let $g_k(z_i, z_j)$ denote the affinity associated with pairwise assignment of patch z_i and z_j to hypothesis c_k. In this application, pairwise assignment affinity is computed based on the fact that a given individual's clothing often exhibits a certain amount of color and motion constancy. Hence, affinity is based on a similarity measure $sim(z_i, z_j)$ of low-level image cues such as motion fields and color distributions. In this paper we use the Bhattacharya distance measure between the color histograms associated with each patch. Given such a measure of similarity, we define

$$g_k(z_i, z_j) = sim(z_i, z_j) min(g_k(z_i), g_k(z_j)). \tag{1}$$

Figure 1d shows two pairwise assignments. The pair of pink patches have a large degree of pairwise affinity while the pair of blue patches exhibit relatively small pairwise affinity.

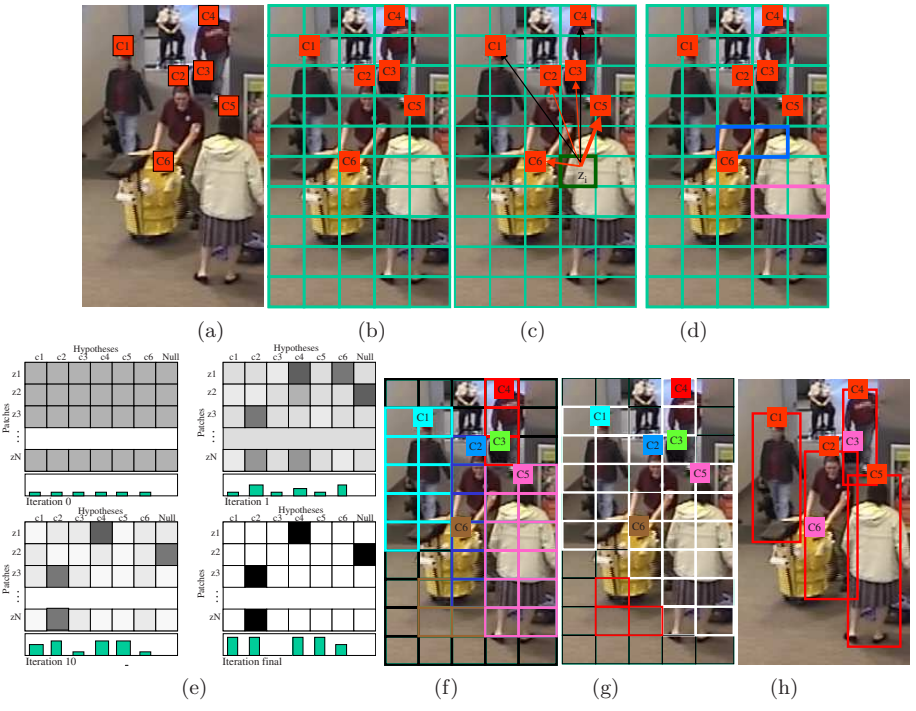

Fig. 1. This figure depicts the different steps of the proposed algorithm. a) An initial set of person hypotheses, b) a partitioning of the scene into a grid of patches, c) an example of the direct association affinity between the green patch and all the hypotheses where the width of the arrow is commensurate with the assignment affinity, d) shows two patches with strong pairwise affinity (pink) and two patches with weak pairwise affinity (blue), e) depicts the soft assign process where patches are assigned to hypotheses, f) shows the assignment of patches to hypotheses after the first *E-step*, g) shows the result of the *M-step* consistency analysis where red patch assignments are deemed to be inconsistent based on occlusion reasoning, h) the final segmentation after multiple iterations of the EM algorithm.

Given an initial set of hypotheses, a patch partitioning and the direct/pairwise affinity measures, a global segmentation can be generated based on an estimate of the optimal set of patch to hypothesis assignments. To achieve this, an EM approach similar to [4] is used. The patch to hypothesis assignments is defined by an assignment vector $V = \{v_i\}$ of length N with $v_i \in [1, \ldots, K]$, where $v_i = k$ implies that patch z_i is assigned to hypothesis c_k. The focus of the EM approach is to model the joint probability of an assignment vector V and the set of patches Z i.e. $p(V, Z; X)$ under the parametrization of a consistency model X (see section 2.1). To this end the log likelihood of the patch assignment is defined as:

$$L(V|Z;X) \propto \gamma_1 \sum_{k=1}^{K} \sum_{i=1}^{N} x_{ik} g_k(z_i)\, \delta_{c_k}(v_i) + \gamma_2 \sum_{k=1}^{K} \sum_{\substack{i,j=1 \\ i \neq j}}^{N} x_{ik} x_{jk} g_k(z_i, z_j)\, \delta_{c_k}(v_i) \delta_{c_k}(v_j)\,, \tag{2}$$

where $\delta_{c_k}(v_i)$ is an indicator function which is one when $v_i = k$ and zero otherwise and x_{ik} is a consistency parameter that is computed during the *M-step* (see section 2.1). During the *E-step*, the consistency parameters are fixed and a distribution for V is computed such that the expectation $\int_V p(V) L(V|Z;X)$ is maximized. It was shown in [4] that a mechanism similar to soft-assign [15] can be used to efficiently perform the *E-step* search. Figure 1e illustrates this iterative process where the distribution of V is parametrized by a matrix of direct assignment probabilities. The element in the i^{th} row and k^{th} column of this matrix is the probability of the assignment of the i^th patch to the k^{th} hypothesis. The sum along each row must be equal to 1 and there can be no negative values. At the first iteration all matrix values are set uniformly. During each iteration of the soft assign process, the matrix probabilities are updated based on the gradient of the expectation function. After a number of iterations, the assignment probabilities are forced to take on binary values and this defines an estimate of the most likely value of V. If a hypothesis receives no patches then it is deemed to be a spurious detection. A null hypothesis is created to allow for the potential assignment of patches to the background and clutter in the scene. In this application the direct assignment affinity between a patch and the null hypothesis is set to a nominal value. The reader is directed to [4] for more detail regarding the *E-step* process.

2.1 Enforcing Consistency

Figure 1f depicts a set of patch to hypothesis assignments that result from one pass of the *E-step*. Note that a certain number of patches are assigned to the spurious hypothesis c_6. In this case, these assignments would seem to indicate that the feet of c_6 are visible but that the head is not. These types of inconsistencies are addressed via the consistency parameters $X = \{x_{ik}\}$ which are computed during the *M-step*. As can be seen from equation 2, when x_{ik} is low, the likelihood of an assignment vector with $v_i = k$ is reduced. For this paper, the primary concern is to enforce consistency based on occlusion reasoning. The following axioms form the basis for such an approach:

- If the person is visible then their head should be visible
- The non-occluded region of a person should be contiguous.

Given the current estimate of V all the patches that are currently assigned to a given hypothesis c_k can be identified. For each patch z_i that is assigned to c_k, a path between it and the head location specified by c_k can be constructed such that the number of patches encountered on the path that are not assigned to c_k is minimal. This process takes advantage of the inherent grid like structure of the patches and can be computed efficiently using dynamic programming. The

value of x_{ik} is set to 1 unless the minimum cost path has a cost that is greater than a threshold, in which case x_{ik} is set to a low value. Prior to the first *E-step*, all the values of X are set to 1. Using this simple process, inconsistencies such as *the legs are visible but the torso is not*, can be identified and addressed before the next iteration of the *E-step*. Figure 1g shows the result of an *M-step* analysis where consistent patch assignments are shown in white and the inconsistent patch assignments are shown in red. By reducing the value of the consistency parameters for the red assignments, their inclusion in subsequent estimates of V will be inhibited.

The EM algorithm operates by iterating between the *E-step* and the *M-step* operations. The process terminates when the estimates of V have converged. Figure 1h shows the final segmentation for this example. By employing a global optimization scheme, the system need not rely solely on local information for making segmentation decisions, which is not the case for many greedy approaches to crowd segmentation. In the next section, the critical question of how to compute the affinity of direct patch to hypothesis assignments will be addressed.

3 Classification

In the previous section a detailed description of the overall crowd segmentation process was given. The focus of this section is to describe how the direct patch to hypothesis affinity function $g_k(z)$ can be computed based on local shape and appearance information. For this purpose we use a whole body monolithic person classifier consisting of a set of weak classifiers selected by boosting. We will show that for certain types of weak classifiers, the whole body classifier response can be computed for a specific patch and that this response can be used to characterize the patch to hypothesis affinity. The section begins with a discussion of the basic whole body classifier followed by details regarding the generation of patch specific partial responses.

3.1 Whole Body Classifier

A sample s is defined as a head location in the image (x_h, y_h) and an associated sample bounding box that covers the entire body region. Given a hypothesized head location, the associated bounding box is computed based on camera calibration information under the dominant ground plane assumption [14]. Given an arbitrary sample, a whole body classifier produces either a positive or negative response. As was shown in [16] a strong classifier (sc) can be constructed from a linear combination of weak classifiers (wc). The weak classifier is chosen by minimizing an expected error function that is based on labeled training data:

$$E(wc) = \sum_{i}^{N} \frac{(1 - wc(s_i) * l_i)p_i}{2}, \tag{3}$$

where s_i is the i^{th} training sample, l_i is its label and p_i is the probability associated with sample s_i. The sample probability distribution is modified in an

iterative fashion so as to emphasize samples that were mis-classified during previous iteration. In this application an iterative site-specific approach is used for learning. Initial training data from the site of interest is manually labeled and a classifier is constructed. It is then applied to additional imagery taken from the site, and the resulting false positives are incorporated into the training data, while correctly classified negative training samples are removed. A new classifier is then constructed and this process is repeated until no new false positive responses are generated. In this manner a series of strong classifiers are constructed which are combined to form a cascaded classifier. The type of weak classifiers that are appropriate for this application is now considered.

3.2 Weak Classifiers

A particular type of weak classifier can be characterized as follows:

$$wc(s; \mathcal{R}(s)), \tag{4}$$

where $\mathcal{R}$ is a region of interest defined relative to the sample bounding box associated with a sample s. If the average values for a set of image statistics are above (or below) a given set of thresholds, then the weak classifier produces a positive or negative response accordingly. Once the type of image statistics have been selected, the weak classifier is essentially parametrized by the relative location and dimensions of its region of interest $\mathcal{R}$. In general the threshold values for selected weak classifiers are determined during the learning phase of the boosting process. If we restrict our hypothesis space to this type of weak classifier, then a patch specific partial response for a whole body classifier can be generated.

Based on the boosting algorithm a strong whole body classifier for the sample s is defined as:

$$sc(s) = \sum_{i=1}^{M} \alpha_i wc(s; \mathcal{R}_i(s)) \tag{5}$$

The basic idea for generating patch specific responses is that each weak classifier will only collect statistics over the intersection of $\mathcal{R}(s)$ and the patch z. Since average statics are used, the thresholds learned during boosting remain valid. However, instead of having a $1/-1$ response, each weak classifier will have its response modulated by the ratio of the areas of $\mathcal{R}(s) \cap z$ and $\mathcal{R}(s)$. Based on this idea, the partial response for a strong classifier with respect to a given patch z and sample s is defined as:

$$sc(s, z) = \sum_{i=1}^{M} \alpha_i wc_i(s, z) \frac{\int_{\mathcal{R}_i(s) \cap z} dx}{\int_{\mathcal{R}_i(s)} dx}, \tag{6}$$

where

$$wc_i(s, z) = wc(s; \mathcal{R}_i(s) \cap z). \tag{7}$$

Note that if the region of interest associated with a particular weak classifier does not intersect with the patch z, then this weak classifier will have no effect on the strong classifier decision.

For a given person hypothesis c_k, a sample s_k can be constructed so that for a particular patch z_i, the direct patch to hypothesis affinity measure can be defined as:

$$g_k(z_i) = sc(s_k, z_i) \tag{8}$$

Figure 2 shows a set of cascaded classifiers that were used to construct the whole body classifier. In this application the image statistic used is the magnitude of the edge responses for pixels that exhibited an orientation similar to the preferred orientation of the weak classifier. Edge magnitude and orientation are calculated using the Sobel operator. Given such a whole body classifier the next question is to determine the appropriate patch size. If the patch is too large, then there is risk of contamination by occlusion. On the other hand, if the patch is too small the ability to discriminate between correct and incorrect patch assignments diminishes. To understand this tradeoff a training set of positive and negative whole body samples was collected. Patches with widths ranging from 0.25W to 1.0W (W = person width) were evaluated across the entire bounding box for each training sample. For each relative patch location, the average number of positively responding strong classifiers from the cascaded whole body classifier was recorded. As shown in Figure 3, when the patch width was reduced below 0.5W the ability to discriminate between positive and negative samples was reduced significantly. Thus for this application a nominal patch width of 0.5W is chosen.

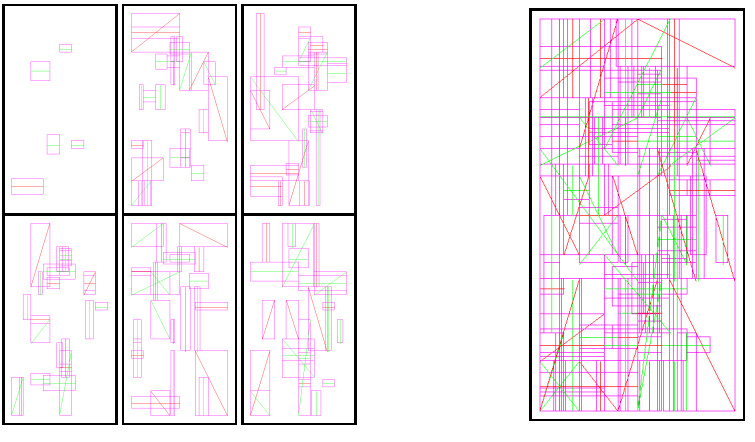

Fig. 2. This figure shows the six strong classifiers that were constructed for the whole body classifier plus all six cascades shown together. Each pink box represents the region of interest for a weak classifier. The line interior to each region of interest depict the weak classifier's preferred orientation. Green lines represent positive features (the average statistic must be above its threshold) and red lines are for negative features (average statistic must be below its threshold).

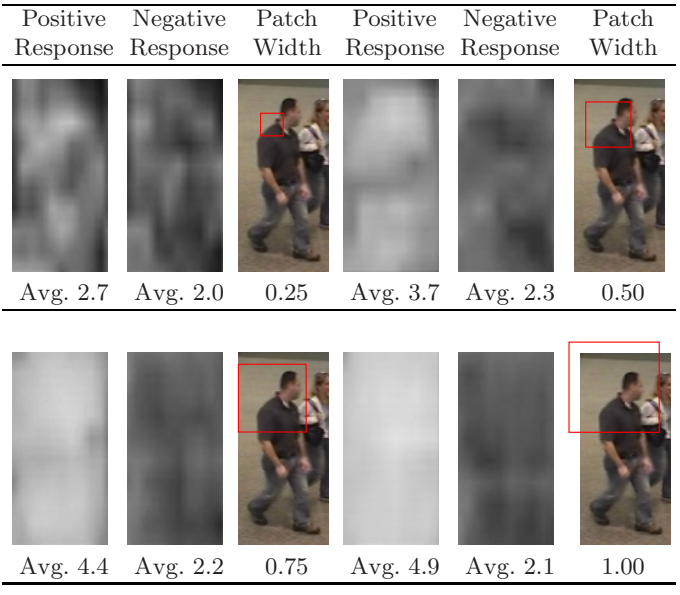

Fig. 3. This figure shows the effect of changing the patch size. Patch size is varied as a function of W = person width. In each case the average number of positively responding strong classifiers from the whole body cascaded classifier is shown as a function of patch location for both positive (person) and negative (non-person) images. Note that when the patch width is reduced below 0.5W the ability to discriminate between positive and negative samples is significantly reduced.

3.3 Hypothesis Nomination

For this application, hypothesis nomination as described in section 2 is achieved using a scanning window approach. For every possible sample, the partial response for the whole body classifier is evaluated based on a patch covering the hypothesized head and shoulder regions. The set of positive responses constitute the initial hypotheses set C.

4 Experiments

Unrehearsed imagery acquired at a mass transit site serves as the source of test imagery for this paper. A whole body classifier was trained for this site. We first illustrate the intermediate steps of our approach on a few representative frames (see Figure 4). The "Initial Hypothesis" column of figure 4 shows the initial set of hypotheses generated by the head and shoulders classifier. Note that while an appropriate hypothesis was generated for each person in each image, several false hypotheses were also generated. The "Single Assignment" column of figure 4 illustrates the direct affinity between each patch and each hypothesis as computed using equation 8. Each patch is color coded based on the hypothesis

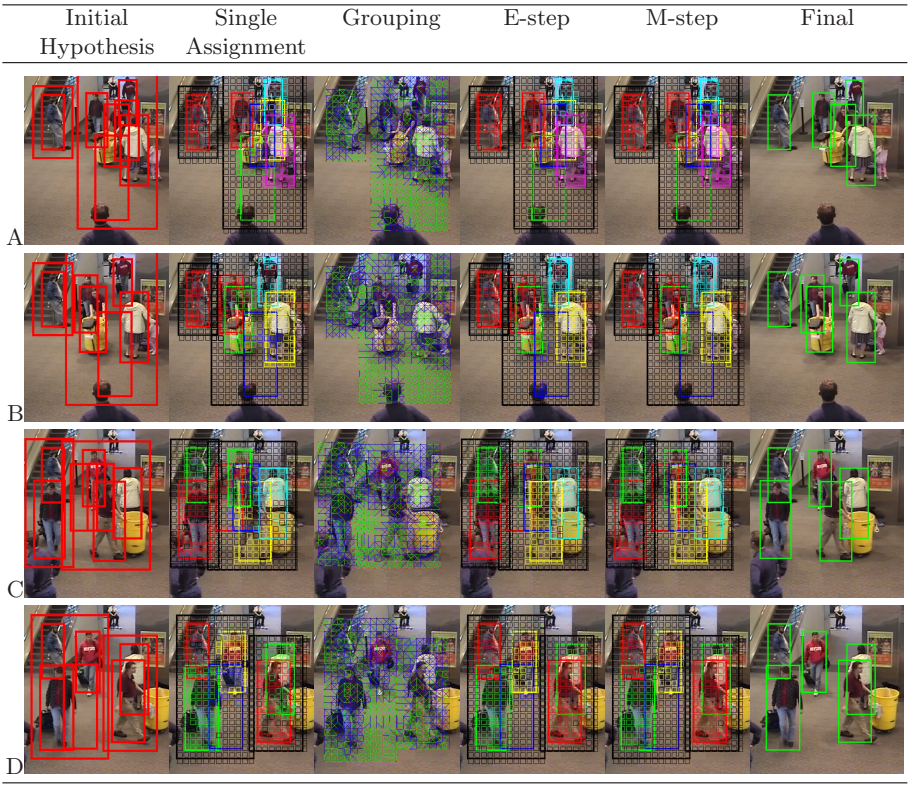

Fig. 4. Six stages of the crowd segmentation process are illustrated for four test images. Overlapping patches of 0.5W are used. However, for clarity smaller patches are shown. The initial hypotheses generated by the head and shoulder classifier are shown in the first column. In the second column, the patches are shown color coded based on their strongest direct assignment as calculated by the whole body classifier. The null hypothesis is shown in black. In the third column, neighboring patches with strong similarity measures based on color constancy are connected by green line segments. The assignment of patches to hypotheses based on the first E-step is shown in the fourth column. The assignment after multiple rounds of both the E and M steps are shown in the fifth column. The final segmentation is shown in the last column.

for which it has the highest direct affinity. Patches that are black have the greatest affinity for the null hypothesis. A significant number of patches have the greatest direct affinity for their true hypothesis, however confusion occurs when multiple hypotheses overlap. An example of this can be seen in row A of the Single Assignment column. In addition, patches that are only associated with false detections tend to have a greater affinity for the null hypothesis.

The "Grouping" column of figure 4 illustrates the effectiveness of the pairwise assignment criteria. For purposes of clarity, only neighboring patches with high similarity measures are shown to be linked in green (blue otherwise). Note that

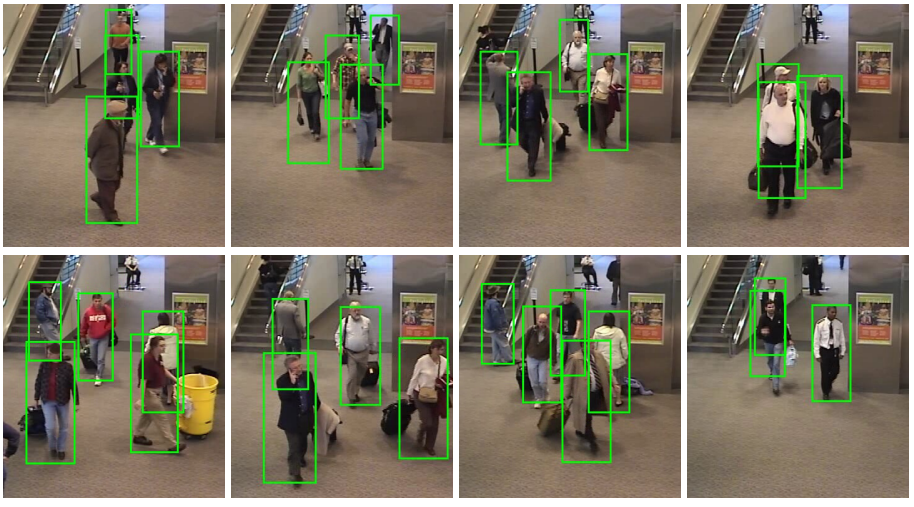

Fig. 5. This figure shows an assortment of crowd segmentation results. Note that the algorithm produces the correct segmentation in case of severe partial occlusion (right column), and in presence of cleaning equipment (bottom left) and a variety of suitcases and bags.

patches associated with the same article of clothing tend to be grouped together. Also, the background often exhibits continuity in appearance and such patches tend to be grouped together.

The "E-step" column of figure 4 shows the patch assignment after the first iteration of the "E-step". Most of the false hypotheses have received very few patch assignments, while the true hypotheses have been assigned patches in an appropriate manner. However, inconsistencies have also been generated. For example, in row A of the "E-step" column, a number of patches have been assigned to the bottom of the green hypothesis. As seen in the "M-step" column, these inconsistent assignments have been correctly removed.

The "Final" column of figure 4 shows the final segmentation. Figure 5 shows similar results from a variety of images. Note that the algorithm is successful when confronted with partial occlusion and clutter such as the janitor's equipment and various suitcases.

The algorithm was also applied to a video sequence (see supplemental material). To measure overall performance 117 frames were processed. The initial hypothesis generator produced 480 true detections, 32 false detections and 79 missed detections. After application of the crowd segmentation algorithm, the number of false detections were reduced by 72 percent at a cost of falsely rejecting 2 percent of the true detections. For purposes of comparison, we applied the Histogram of Oriented Gradients (HOG) [6] to this dataset. Our implementation uses camera calibration information for automatic scale selection. The performance tabulated in Table 1 shows that our crowd segmentation outperformed HOG, arguably this is due to partial occlusion.

Table 1. Comparison of HOG [6] person detector to the proposed crowd segmentation algorithm

	True Detects	Missed Detects	False Alarms
Crowd Segmentation	470	89	9
HOG [6]	387	172	20

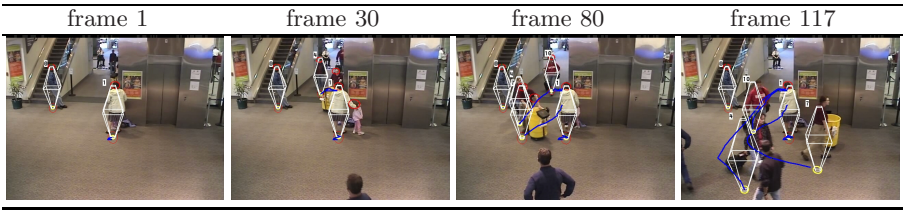

Fig. 6. Four example frames from tracking the results of the crowd segmentation process

The purpose of crowd segmentation algorithms in general is to reliably track the location of people over time. The final segmentation of the sequence previously described was processed by a general-purpose person tracking algorithm. At every time step, locations and estimates of the location uncertainties are projected into the scene ground-plane via an unscented transform. Our tracker processes these detections in a scene ground-plane reference frame, where the dynamical models are intuitively formulated. Our approach to tracking is similar to [17] and [18]. We follow an efficient detect and track approach [19] using a JPDAF filter [20], which has excellent performance in high degrees of clutter while being efficient in the presence of many targets. The tracking results in Figure 6 show the trajectories of all people in the scene.

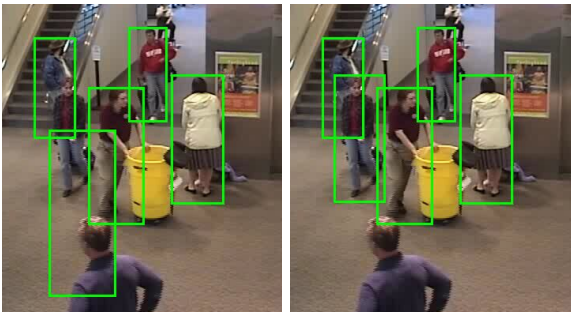

Fig. 7. This figure illustrates the effect of using motion fields in the pairwise patch assignment. The example on the left shows a frame where the algorithm results in both a false positive and a false negative. However, when the motion information from dense optical flow is used the correct segmentation results, as shown on the right.

The results thus far used pairwise patch similarity function based on color constancy as defined in Equation 1. However, this is not always enough as shown in the left image of Figure 7 where the crowd segmentation algorithm resulted in both a false and a missed detection. An experiment was performed where the pairwise patch similarity measures were augmented by the use of a motion consistency measure based on dense optical flow. As can be seen from the right image in Figure 7, this results in a correct segmentation.

5 Discussion

The framework presented in this paper has incorporated many of the strengths of previously proposed crowd segmentation methods into a single unified approach. A novel aspect of this paper is that monolithic whole body classifiers were used to analyze partially occluded regions by considering partial responses associated with specific image patches. In this way appearance information is incorporated into a global optimization process alleviating the need for foreground background segmentation. The EM framework was also able to consider low level image cues such as color histograms and thus take advantage of the potential color constancy associated with clothing and the background. Parametrization of the likelihood function allowed for the enforcement of global consistency of the segmentation. It was shown that these parameters can be estimated during the *M-step* and that this facilitates consistency based on occlusion reasoning.

In the course of experimentation it was found that at various times, different aspects of the crowd segmentation system proved to be the difference between success and failure. For example when confronted with clutter, the appearance based classifiers provide the saliency required to overcome these challenges. However, when multiple people having similar clothing are encountered, the motion field can become the discriminating factor. A robust system must be able to take advantage of its multiple strengths and degrade gracefully when confronted by their weaknesses.

References

1. Munder, S., Gavrila, D.: An experimental study on pedestrian classification. IEEE Trans. on Pattern Analysis and Machine Intelligence 28(11), 1863–1868 (2006)
2. Leibe, B., Seemann, E., Schiele, B.: Pedestrian detection in crowded scenes. IEEE Computer Vision and Pattern Recognition, 878–885 (2005)
3. Leibe, B., Cornelis, N., Cornelis, K., Gool, L.V.: Dynamic 3d scene analysis from a moving vehicle. IEEE Computer Vision and Pattern Recognition, 1–8 (2007)
4. Rittscher, J., Tu, P.H., Krahnstoever, N.: Simultaneous estimation of segmentation and shape. IEEE Computer Vision and Pattern Recognition 2, 486–493 (2005)
5. Brostow, G.J., Cipolla, R.: Unsupervised bayesian detection of independent motion in crowds. IEEE Computer Vision and Pattern Recognition I, 594–601 (2006)
6. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. IEEE Computer Vision and Pattern Recognition, 886–893 (2005)

7. Tuzel, O., Porikli, F., Meer, P.: Pedestrian detection via classification on riemannian manifolds. IEEE Computer Vision and Pattern Recognition (2007)
8. Stauffer, C., Grimson, W.: Adaptive background mixture models for real-time tracking. IEEE Computer Vision and Pattern Recognition 2, 246–252 (1998)
9. Rabaud, V., Belongie, S.: Counting crowded moving objects. IEEE Computer Vision and Pattern Recognition, 705–711 (2006)
10. Viola, P., Jones, M., Snow, D.: Detecting pedestrians using patterns of motion and appearance. International Journal of Computer Vision 2, 734–741 (2003)
11. Fergus, R., Perona, P., Zisserman, A.: A visual category filter for Google images. In: European Conference on Computer Vision, vol. 1, pp. 242–256 (2004)
12. Mikolajczyk, K., Schmid, C., Zisserman, A.: Human detection based on a probabilistic assembly of robust part detectors. In: European Conference on Computer Vision (2004)
13. Wu, B., Nevatia, R.: Detection and tracking of multiple partially occluded humans by bayesian combination of edgelet based part detectors. International Journal of Computer Vision 75(2), 247–266 (2007)
14. Zhao, T., Nevatia, R.R.: Bayesian human segmentation in crowded situations. IEEE Computer Vision and Pattern Recognition 2, 459–466 (2003)
15. Chui, H., Rangarajan, A.: A new point matching algorithm for non-rigid registration. Computer Vision and Image Understanding 89(3), 114–141 (2003)
16. Viola, P., Jones, M.J.: Robust real-time face detection. International Journal of Computer Vision 57(2), 137–154 (2004)
17. Krahnstoever, N., Tu, P., Sebastian, T., Perera, A., Collins, R.: Multi-view detection and tracking of travelers and luggage in mass transit environments. In: Proc. Ninth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS) (2006)
18. Leibe, B., Schindler, K., Gool, L.V.: Coupled detection and trajectory estimation for multi-object tracking. In: International Conference on Computer Vision (ICCV 2007), Rio de Janeiro, Brasil (October 2007)
19. Blackman, S., Popoli, R.: Design and Analysis of Modern Tracking Systems. Artech House Publishers (1999)
20. Rasmussen, C., Hager, G.: Joint probabilistic techniques for tracking multi-part objects. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 16–21 (1998)