Detecting Major Segmentation Errors for a Tracked Person Using Colour Feature Analysis

Christopher Madden and Massimo Piccardi
University of Technology, Sydney
Broadway, NSW, 2007 Australia,
cmadden, massimo@it.uts.edu.au *

Abstract

This paper presents a method to identify frames with significant segmentation errors in an individual's track by analysing the changes in the features as they are tracked through the frame sequence. The changes in a person's features between each frame of a given track can give an indication of segmentation errors and illumination changes that are large enough to cause major appearance changes. This paper is found analysing changes in colour features as opposed to the bounding box changes of humans to provide greater accuracy. This was analysed from 26 tracks of 4 different people across two cameras with differing illumination conditions. By fusing two spatial colour features with a global colour feature, probabilities of segmentation error detection as high as 83 percent of human expert identified major segmentation errors are achieved with false alarm rates of only 3 percent. This indicates that the analysis of such features along a track can be useful in the automatic detection of significant segmentation errors. This can improve the final results of many applications that wish to use robust segmentation results or other features from a tracked person.

1. Introduction

Tracking is based upon motion, shape and appearance features [3]. Motion is commonly used as the main feature, with shape and appearance often used to help disambiguate multiple targets. Correct data association is achieved even in the presence of major segmentation errors in some frames due to the continuity of motion as well as partial shape and appearance matching. In this paper we assume that we want to select those frames along the track that are affected by major segmentation errors. This identification of segmentation errors could be useful to improve a range of applications including, but not limited to: a) following single objects as they move across disjoint camera views, where continuous tracking is not possible and matching is enabled by accurate extraction of features such as shape and appearance in each view [4] [5] [9]; b) creating a faithful, synthetic pictorial summary of a tracked object by one or few frames where the object is not affected by major segmentation errors; c) accurate searching in an image or video archives.

This paper provides a detailed look at comparing methods for identifying major segmentation errors along a track. Such errors could be propagated into further applications which are based upon the segmented objects, such as feature analysis. The methods presented and compared overlook minor segmentation errors which commonly occur using most common segmentation techniques [6]. The results presented in this paper are based upon an adaptive Gaussian model similar to that used in the Pfinder project [8]. We use this method because of its speed and reasonable accuracy, and do not attempt to compare it with other segmentation techniques. Significant segmentation errors may occur more frequently than they do in other methods; however this is not a problem if the errors can be identified.

Identifying segmentation errors is made more difficult for articulated objects which can change their shape within constraints, such as humans, as this can lead to appearance changes through self occlusions. A number of hypotheses do hold in a statistical sense for tracked humans: they tend to walk upright, wear clothing differing for the vertical layers relating to the torso and legs, and have an appearance that is often similar for different equatorial views. Illumination provides another challenge as it can vary over time, and in different patterns depending upon camera location and whether it is indoor or outdoor, changing perceived appearance features and the contrast of the object and the background*; however the literature on this subject is small.*

For instance, Erdem *et al.* [1] have tried to identify and overcome segmentation errors for a 3D television application to improve the temporal stability of object segmenta-

*This research is supported by the Australian Research Council under the ARC Discovery Project Grant Scheme 2004 - DP0452657.

tion, rather than identify and remove errors. They achieved their aim by minimising changes in the global colour histogram and turning angle function of the boundary pixels of the segmented object in each frame, which to maximises temporal stability.

We also propose to use changes in colour appearance features to identify segmentation errors, in the form of global, upper, and lower Major Colour Representation (**MCR**) features, which are tolerant to a degree of illumination changes [5]. The upper and lower MCR features aim to represent the colours of the upper and lower regions of a person's clothing. Such regions often consist of a single or small range of colours, where the upper and lower regions are often distinct from each other. As the person is tracked within a single camera over a relatively short period of time, the intrinsic appearance of the features should remain constant. Thus changes in the features of a person between a frame and the majority of the other frames from the same track are due to factors such as segmentation errors, large illumination changes, or tracking errors. If we assume that the majority of the frames within the track are satisfactorily segmented, then the set of frames without major errors can then provide robust features to be used in a variety of applications.

The description of the features used in this paper is separated into two parts, with Section 2 describing a simple bounding box analysis method for determining segmentation errors. Section 3 explores the extraction of colour features relating to the upper and lower clothing colour as well as a global colour feature with Section 3.1 describing how these colour features can be compared along a track to identify segmentation errors, and Section 3.2 describing typical colour feature similarity patterns that identify segmentation errors. Section 4 outlines the statistical analysis of the detection and false alarm rates based upon the usage of these features.

2. Analysing bounding box changes

Bounding boxes have been widely used in the literature to speed up the analysis of objects by creating a simple rectangle model of the object, which can then be used to identify object bounds and when they overlap. Hence it is a good candidate for fast identification of segmentation errors in a person's track. If we assume that an object is correctly tracked either manually, or using one of the many popular motion estimation techniques, such as Zhao and Nevatia's method for tracking in complex scenes [11]. Once accounting for perspective distortion, the changes in the object size are likely caused by large segmentation errors. If the bounding box of any object changes significantly from the expected size in that frame, then this can indicate that the object is either occluded or segmented incorrectly. For an articulated object, such as a person, the position of that person, as well as their size and shape can change within limits. Thus bounding boxes for an accurately segmented person can change due to movement actions such as walking, or from the camera perspective as a person moves toward or away from the camera position.

The expected changes in the bounding box size can be simplified by assuming that the camera frame rate is not slower than a few frames per second, and people are walking upright in the viewed area. These two assumptions generally hold for the video surveillance environment where people are traversing a space viewed by the camera in order to travel from point A to point B. The amount of change in the bounding box size could be learnt per camera to maximise the ability to determine segmentation errors; however for most typical frame rates the allowable amount of segmentation errors is often higher than the changes due to perspective distortion. Although many object statistics could be used to analyse changes in the shape of an object, we found the vertical height of the bounding box for changes remains invariant whilst a person is walking. This still remains sensitive to actions where a person might bend over, or become partially occluded. Typical values of the ratio in vertical size between one frame and the next vary in a small range around one, depending upon frame rate, and the amount of perspective distortion present in the camera view.

Figure 1 shows an example of the ratios of bounding box height obtained between one frame and the next for a track where there is a single frame with a large segmentation error. The error is clearly indicated by the change in ratio value below 0.7. The next ratio is over 1.4, indicating that the next frame is much larger than the erroneous frame as it has returned to the normal size. The 5 sample frames from the track show that the image height of the object is diminishing along the track as the object moves away from the camera. The three middle frames show the frame before the error, the frame with the error and the frame after the error. The change in image height of the individual is evident, as is the loss of legs in this frame. Although this method works well for errors in a single frame or a short run of frames, analysis of gradual increases in segmentation errors remains a problem for this method.

3. Determining the global, upper and lower MCR features

The Major Colour Representation **(MCR)** used in this paper to define the colour features is based upon the method previously developed in [5]. MCR are essentially colour histograms in the joint R, G, B space, built with sparse bins whose position and number is adjusted to fit the pixel distribution. Instead of just using a global colour feature, as in [1] and [5], we propose to add two extra colour features

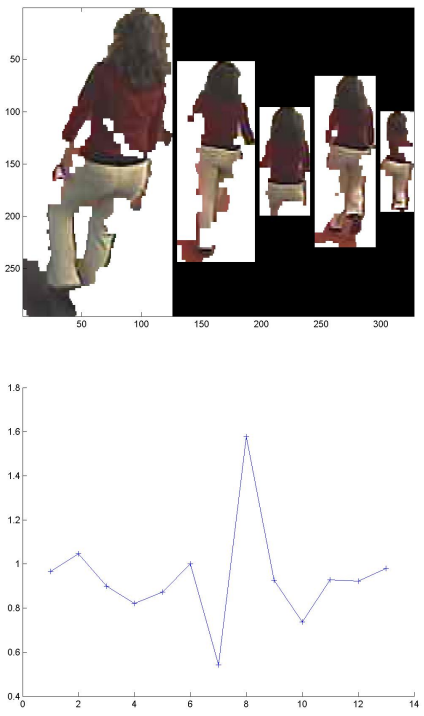

Figure 1. Example of changes in bounding box height ratios where a large segmentation errors occurs and 5 sample frames from the track including the erroneous frame

relating to the upper and lower clothing colours of a person. These features are chosen to represent the often different colours of the clothing on the upper torso, and those on the legs. The narrow spatial aspect of these features also allows for a more sensitive analysis of the spatial positioning of a persons colours. This ensures that changes in the position of the colours can be detected, such as where segmentation errors remove large portions of the object.

Extracting the MCR for each of these three colour features utilises the same process, but analyses different spatial components of the appearance of the segmented object. The process for extracting the MCR's is similar to the method described in detail in [5], that we summarise here:

- A controlled equalisation step performs a data-dependent intensity transform. This spreads the histogram to compensate for minor illumination changes that can be expected within the indoor and outdoor surveillance environments.
- Online K-means clustering of pixels of similar colour within a normalised colour distance δ, as defined in Equation 1, generates the MCR of each spatial region. The cluster centre is the average of the colour values within δ, allowing it to better represent the colour cluster. Due to the movement of colour clusters iteration of cluster improvement and cluster assignment are necessary; however as explained in [5] 3 iterations provides an accurate representation with a minimum of computational cost.

$$\delta\left(C_1, C_2\right) = \frac{\sqrt{\left(R_1 - R_2\right)^2 + \left(G_1 - G_2\right)^2 + \left(B_1 - B_2\right)^2}}{\sqrt{R_1^2 + G_1^2 + B_1^2} + \sqrt{R_2^2 + G_2^2 + B_2^2}} \quad (1)$$

The three MCR features are defined as:

- The global MCR feature represents the colours of the whole segmented object without any spatial information.
- The upper MCR feature represents the colour of the top portion of clothing. This corresponds to the region from 30-40 percent of the person from the top of the objects bounding box as shown in Figures 2 and 3. This narrow band was chosen to ensure that it avoids the inclusion of the head and hair of the object, as well as low necklines, but does not go so low that it includes the belt area, or overlaps with the bottom colour, or pant area.
- The lower MCR feature is aimed to represent the colour of the lower portion of clothing. This corresponds to the region from 65-80 percent of the object from the top of the objects bounding box as shown in Figures 2 and 3. This narrow band avoids the very bottom of the object which can be prone to shadows, or artifacts where the feet touch the ground. It also tries to avoid overlapping with the belt or upper torso area of the person.

The narrowness and positioning of both of the upper and lower MCR regions also allows for them to remain constant under minor segmentation errors that will only have a minimal impact upon a person's features, whilst still remaining sensitive to large segmentation errors. These features also allow for the inclusion of spatial colour features which could possibly identify the difference between people when tracking is incorrect.

Figures 2 and 3 show the upper MCR feature region between the lines toward the top of the person, and the lower MCR feature region between the lines toward the bottom of the person. Figure 2 demonstrates three frames showing frontal and rear views of a person, and a frontal view with a significant segmentation error where the lower half of the person is not found. In this frame the colours within the upper and lower regions change significantly from those

Figure 2. Example of upper and lower regions from three segmentations of one person

Figure 3. Example of upper and lower regions from three segmentations of a second person

in the other two frames. Figure 3 shows two views of a second person where segmentation is arguably reasonable, even if a portion of the head is not correctly segmented in the first frame. The frame shown in the middle is poorly segmented; however a white object above the person is partially included in the same segment. This leads to an added amount of white in the global colours, that is not entirely dissimilar to the pants colour. Such an error leads to the frame having little discernible difference in global colours; however the upper and lower colour regions clearly indicate a degree of change in the spatial positioning of the colours that can be used to identify this poorly segmented frame.

3.1 Comparing colour features between frames

Once extracted, the MCR features can be compared to each other to determine if the features change over time along the track of the object. We begin by assuming that objects are tracked correctly, even though large and sustained changes in human object features may indicate potential data association errors. This analysis of tracking errors is not explored within this paper as the current data involved data that is overly simplified for tracking, and thus tracked manually, allowing us to focus on the full analysis of data from a single individual in each track. Changes in object features along the track are therefore likely to be caused by errors in the identification of foreground pixels, or through other causes such as occlusion, cluttering, or major lighting changes. Minor segmentation errors are common in realistic environments using even the most effective of current techniques [6], and therefore need to be retained to keep a useful number of frames from the track.

Most of the current techniques for evaluating segmentation techniques utilise a ground truth segmentation developed by human experts in a time consuming process [10]. We propose to use an automatic comparison between the frames of a track utilising the global, upper, and lower MCR colour features using the technique described in [5]. *This technique is based upon determining the* Kubek-Liebler *distance [7] between the colour clusters from the MCR of frame A and the MCR of frame B. This produces a similarity value for the three MCR features between the two frames using equation 2.*

$$S(A,B) = 1 - \frac{S_{max} - S_{min}}{S_{max} + S_{min}} \tag{2}$$

This process is used to generate pairwise similarity values for the three MCR features in each frame to every other frame in the track. We apply a statistical analysis of a known training set of the non-matching *H0* or the matching *H1* sets of features to determine Gaussian likelihood functions for given similarities being matching or non matching [2]. Classification can also be obtained by fusing together the matching and non-matching likelihoods of each of the three feature comparisons in an ensemble of classifiers. We assume the features to be conditionally independent and so apply Bayes theorem as:

$$P(H0|s_G, s_U, s_L) = B(P(s_G|H0)\,P(s_U|H0)\,P(s_L|H0)) \tag{3}$$

$$P(H1|s_G, s_U, s_L) = (P(s_G|H1)\,P(s_U|H1)\,P(s_L|H1)) \tag{4}$$

where B is a prior that can be used to bias the operating point of the system.

3.2 Typical patterns of major segmentation errors

Appearance changes are generally caused by either major segmentation errors, where portions of the object are not extracted correctly from the background, by large changes in the illumination conditions, or by tracking errors. Minor segmentation errors of less than 20 percent are overlooked in this process as they commonly occur [6] and there will be a minimum of changes in object appearance. Major errors produce significant changes in the appearance of a single frame; however they often have a different impact upon the segmentation results when considering their difference across the whole track. Three different error patterns are demonstrated in Figure 4, highlighting how the different errors tend to influence the similarity of the proposed features. The first error pattern in Figure 4a shows how large segmentation errors cause a very low similarity in the fused features between that frame and every other frame causing a characteristic "'cross"' in the pairwise comparisons. Where a small number of frames have similar large segmentation errors, such as losing the lower half of the object, these frames

Track 1885-1894

	1885	1886	1887	1888	1889	1890	1891	1892	1893	1894	Avg matching
1885	1	0.72346	0.040269	0.67614	0.75905	0.82743	0.89009	0.63706	0.6335	0.51523	0.67022
1886	0.72346	1	0.02513	0.6977	0.81461	0.84949	0.81277	0.89385	0.76962	0.59259	0.71792
1887	0.040269	0.02513	1	0.026171	0.063725	0.03417	0.043174	0.04205	0.013697	0.030038	0.12984
1888	0.67614	0.6977	0.026171	1	0.66853	0.79567	0.76798	0.69213	0.84459	0.81089	0.69798
1889	0.75905	0.81461	0.063725	0.66853	1	0.82113	0.87647	0.73828	0.65974	0.83645	0.72280
1890	0.82743	0.84949	0.03417	0.79567	0.82113	1	0.85833	0.91382	0.80466	0.57416	0.74789
1891	0.89009	0.81277	0.043174	0.76798	0.87647	0.85833	1	0.87336	0.73598	0.65928	0.75174
1892	0.63706	0.89385	0.04205	0.69213	0.73828	0.91382	0.87336	1	0.69348	0.64218	0.71262
1893	0.6335	0.76962	0.013697	0.84459	0.65974	0.80466	0.73598	0.69348	1	0.91203	0.70673
1894	0.51523	0.59259	0.030038	0.81089	0.83645	0.57416	0.65928	0.64218	0.91203	1	0.65628

a) single large segmentation error

Track 3095-3104

	3095	3096	3097	3098	3099	3100	3101	3102	3103	3104
3095	1	0.000938	0.000945	0.69484	0	0.046527	0.066226	0.008902	0.028491	0.048307
3096	0.000938	1	0.73378	0.003612	0.86109	0.26509	0.28603	0.061421	0.027958	0.14846
3097	0.000945	0.73378	1	0.006682	0.3299	0.28668	0.316	0.055803	0.027105	0.13583
3098	0.69484	0.003612	0.006682	1	0.005538	0.081253	0.078724	0.047608	0.26263	0.08295
3099	0	0.86109	0.3299	0.005538	1	0.34415	0.11527	0.078596	0.060614	0.19886
3100	0.046527	0.26509	0.28668	0.081253	0.34415	1	0.94116	0.65052	0.75248	0.77442
3101	0.066226	0.28603	0.316	0.078724	0.11527	0.94116	1	0.67511	0.87103	0.85858
3102	0.008902	0.061421	0.055803	0.047608	0.078596	0.65052	0.67511	1	0.59428	0.86182
3103	0.028491	0.027958	0.027105	0.26263	0.060614	0.75248	0.87103	0.59428	1	0.90352
3104	0.048307	0.14846	0.13583	0.08295	0.19886	0.77442	0.85858	0.86182	0.90352	1

b) large illumination change

Track 1298-1307

	1298	1299	1300	1301	1302	1303	1304	1305	1306
1298	1	0.79247	0.80477	0.74167	0.29525	0.32224	0.31006	0.093313	0.04554
1299	0.79247	1	0.9232	0.67458	0.66306	0.24841	0.26767	0.2872	0.021861
1300	0.80477	0.9232	1	0.624	0.79788	0.71125	0.72767	0.309	0.17989
1301	0.74167	0.67458	0.624	1	0.69795	0.78452	0.23059	0.28529	0.076366
1302	0.29525	0.66306	0.79788	0.69795	1	0.46039	0.51422	0.79906	0.3297
1303	0.32224	0.24841	0.71125	0.78452	0.46039	1	0.80858	0.74389	0.48604
1304	0.31006	0.26767	0.72767	0.23059	0.51422	0.80858	1	0.72726	0.48639
1305	0.093313	0.2872	0.309	0.28529	0.79906	0.74389	0.72726	1	0.72857
1306	0.04554	0.021861	0.17989	0.076366	0.3297	0.48604	0.48639	0.72857	1
1307	0.006742	0.00811	0.008823	0.030715	0.19158	0.13621	0.085893	0.5694	0.30402

c) gradual illumination change

Figure 4. Three typical error patterns

will tend to be similar to each other, but distinct from the rest of the track.

Figure 4b shows the second error pattern where the portions of the track are self similar, but persistently different from each other. This occurs for large illumination changes, such as switching a light on, and might also be expected for tracking errors, although we have yet to verify such a finding. In this case extracting both sets of feature representations could be useful for manual analysis of the objects track. The bounding box is not directly affected by large illumination changes; however in practise the change in the amount of contrast between the object and background often leads to major segmentation errors as well.

Figure 4c shows the third error pattern, which occurs for gradual illumination changes, such as clouds moving to cover the sun. The error pattern shows that each frame is still likely to match the majority of the rest of the track, however the initial frames could have a large difference in appearance to the frames toward the end. This case is not caused by segmentation errors, so we consider each frame equally suitable to be included in a robust track.

4. Results

The results presented in this section are based upon the comparisons of the four object features of global MCR, upper MCR, lower MCR and fused MCR results based upon a self-comparison of 26 tracks from four people across two cameras, consisting of over 300 frames. These tracks are automatically analysed to identify frames with significant segmentation errors, compared to ground-truth analysis performed by human experts. Examples of typical segmentation and expected errors are given in Figures 2 and 3 in Section 3. Figure 5 shows the clothing worn by the four individuals studied for this experiment, and examples of good segmentation masks. Of the 26 data sets, 5 were used for training the likelihood functions of the non-matching *H0* or the matching *H1* data sets on a frame by frame basis. The remaining 21 used as a testing set for evaluation, with the results given as a ROC curve in Figure 6. This clearly shows that fusing the three features together produces improved results over the use of any single feature.

Figure 5. Four people of interest and automatically segmented masks of good quality

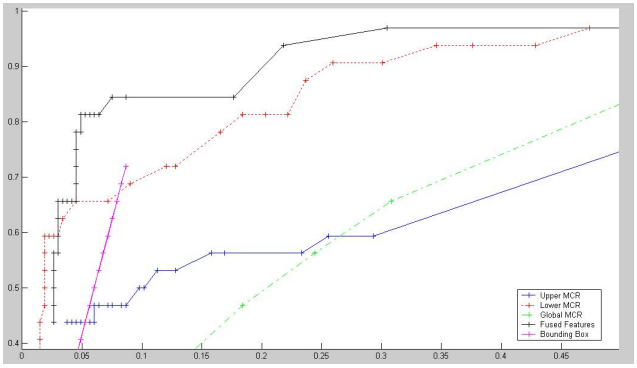

Figure 6. ROC curves of the height, colour and fused feature results

Table 1 gives the probability of detection (*PD*) and probability of false alarm (*PFA*) for each of the MCR features analysed at the selected operating point compared to the expert determined ground truth. A more detailed analysis based upon the individual people who were tracked, as well as the overall quality of the tracks analysed indicates that this method works best with individuals who are not of a uniform colour, and where the overall quality of the segmentation of the tracked object is good; although accuracy is still high under less optimal conditions. The results also

Feature	PD as %	PFA as %
Vertical Bounding Box changes	72	9
Global MCR	66	31
Upper MCR	53	11
Lower MCR	66	5
Fused MCR	84	3

Table 1. PD and PFA values of Bounding Box and MCR features for detecting segmentation errors

indicate that the use of upper and lower MCR features dramatically improves the ability of the system to detect major errors which may not be detected adequately with the global colour or bounding box analysis alone. The major limitation of the bounding box change ratio analysis is that gradually increasing or decreasing segmentation errors are hard to identify, especially where there are multiple concurrent major or minor segmentation errors. This creates a limit to the overall accuracy of such analysis that is not inherent in the fused colour features.

Note that *PD* can be taken close to 100 percent if we can accept a *PFA* of approximately 30 percent. This operating point is of interest if we mean to remove all frames with major errors and the number of frames left by the selection procedure is still sufficient for further processing.

5. Conclusions

A number of factors ranging from the contrast of the person from the background to occlusions and illumination changes, ensure that segmentation will often include a degree of errors, some of which might be very large. Such errors can propagate into a number of subsequent tasks, which range from data association of people in disjointed camera views, to creating a faithful pictorial summary of a tracked object by one or few frames where the object is not affected by major segmentation errors, or accurate searching for the person in image or video archives.

This paper demonstrates an automatic method that can be used to identify significant segmentation errors that occur through the analysis of colour appearance and shape features along an individual's track. It shows that analysing the changes in size of the bounding box of a person throughout their track can identify many large segmentation errors; however it is much less successful where there are multiple major and minor segmentation errors in a track, or gradual errors. The results demonstrate that the upper and lower MCR features, which combine colour and spatial information, can individually identify a degree of the segmentation errors; however when these features are fused with the global colour features, the proposed method identified 84 percent of the major segmentation errors of the 21 test tracks analysed with only 3 percent false alarms. Such low false alarm rates ensures that almost all of the reliable portions of the track can be utilised to improve the robustness of subsequent tasks. Alternatively a detection rate as high as 96 percent of the erroneous frames could remove almost all of the effective frames if the loss of 30 percent of the frames in false alarms will not detrimentally affect the further stages of processing

References

[1] C. E. Erdem, F. Ernst, A. Redert, and E. Hendriks. Temporal stabilization of video object segmentation for 3d-tv applications, 2004. International Conference on Image Processing.

[2] G. Fumera and F. Roli. A theoretical and experimental analysis of linear combiners for multiple classifier systems. *IEEE Transactions on Pattern Analysis And Machine Intelligence*, 27(6):942–956, 2005.

[3] W. Hu, T. Tan, L. Wang, and S. Maybank. A survey on visual surveillance of object motion and behaviors. *IEEE Transactions on Systems, Man and Cybernetics*, 34:334–352, 2004.

[4] O. Javed, K. Shafique, and M. Shah. Appearance modeling for tracking in multiple non-overlapping cameras, 2005. IEEE Conference on Computer Vision and Pattern Recognition.

[5] C. Madden, E. D. Cheng, and M. Piccardi. Tracking people across disjoint camera views by an illumination-tolerant appearance representation. *Machine Vision and Applications*, 1(1):234–778, 2006.

[6] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam. Image change detection algorithms: A systematic survey. *IEEE Transactions on Image Processing*, 14(3):294–307, 2005.

[7] N. sure. Dunno. *Umm*, 1(1):1–2, 5.

[8] C. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland. Pfinder: real-time tracking of the human body. *IEEE Transactions on Pattern Analysis And Machine Intelligence*, 19(7):780–785, 1997.

[9] W. Zajdel and B. Krse. A sequential bayesian algorithm for surveillance with non-overlapping cameras. *Int. Journal of Pattern Recognition and Artificial Intelligence*, 19(8):977–996), year = 2005.

[10] Y. J. Zhang. A review of recent evaluation methods for image segmentation, 2001. International Symposium on Signal Processing and its Applications.

[11] T. Zhao and R. Nevatia. Tracking multiple humans in complex situations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1208–1221, 2004.