Multi-Frame Moving Object Track Matching Based on an Incremental Major Color Spectrum Histogram Matching Algorithm

Massimo Piccardi, *Member*, *IEEE*, and Eric Dahai Cheng
Faculty of Information Technology, University of Technology, Sydney (UTS), Australia
{cheng, massimo}@it.uts.edu.au

Abstract

In this paper, a Major Color Spectrum Histogram Representation (MCSHR) is introduced to represent a moving object by using a normalized geometric distance between two points in the RGB space. Then, an Incremental Major Color Representation Algorithm is proposed to cope with small pose changes occurring along the track. Finally, a two directional similarity measurement based on the major colors is used to measure the similarity of any two given moving objects in multiple integrated frames. Experimental results show that with a few (4 or 5) frames MCSHR integration, the proposed Incremental MCSHR algorithm can make matching more robust and reliable than single frame matching, especially for small pose changes. The major color representation algorithm based on the introduced color distance can represent moving objects accurately with a limited number of colors and the frequency of each major color. The similarity of a same moving object in two different tracks has improved from 85% to 97% with the number of integrated frames increasing from 1 to 5, while the similarity of two different moving objects has been kept as low as 9% to 19%.

1. Introduction

Matching of moving objects from multiple camera views is becoming more and more important with the increasing use of camera networks in surveillance systems [1-7]. In many cases, the observations of a same object are separate in time and space and single moving objects needs to be tracked across multiple and disjoint cameras. Such a matching is much more challenging than tracking from single or overlapped camera views as it cannot exploit continuous motion information. The assumption in our work is that tracks are available from within single camera views, and the goal is to find correspondences between such tracks.

In this paper, we propose a method to assess the similarity of any given two tracks. First, a color distance based on a geometric distance between two points in RGB space is used to measure the similarity of two different colors. By using a given color distance threshold, all pixels from a moving object MO_i in a given frame t are clustered into a limited number of colors, with each color's frequency defined as the number of pixels in that color. Such colors are then sorted in descending frequency order and the first k used to represent the moving object. We call this histogram the major (or principal) color spectrum histogram representation (MCSHR) of $MO_{i,t}$. Given two arbitrary moving objects, $MO_{i,t}$ and $MO_{j,u}$ a similarity criterion based on the major color representation is used to assess their matching. Furthermore, in order to deal with pose changes along the track, an algorithm for Incremental MCSHR (I-MCSHR) is proposed in this paper. Experimental results proved that with a few (4 or 5) frames MCSHR integration, the proposed incremental algorithm can make matching more robust and reliable than single frame matching, especially for small pose changes.

2. Major color spectrum histogram

2.1. Concept of color distance

In the RGB color space, by using 1 byte to represent each color, there are 16.8 million (16,777,216) colors in total. It is very difficult to compare two objects based on so many possible colors. A common approach to limit the size of this space is that of dealing with the three color components separately, but this dismisses their spatial co-occurrence. In this paper, we introduce a "color distance" between two color pixels based on a normalized geometric distance between the two pixels in the RGB space. Such a geometric distance is defined in equation (1) and exemplified in Fig. 1.

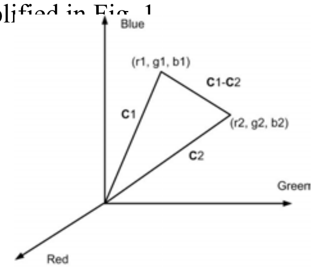


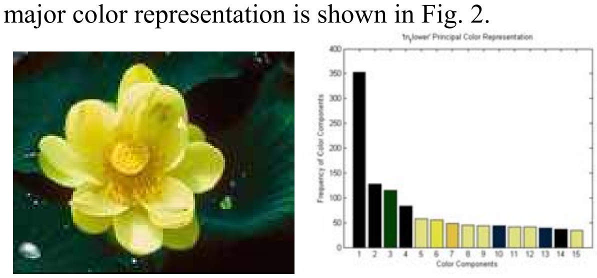
Figure 1. The distance between 2 color pixels in RGB space

$$d(C_1, C_2) = \frac{\|C_1 - C_2\|}{\|C_1\| + \|C_2\|} = \frac{\sqrt{(r_1 - r_2)^2 + (g_1 - g_2)^2 + (b_1 - b_2)^2}}{\sqrt{r_1^2 + g_1^2 + b_1^2} + \sqrt{r_2^2 + g_2^2 + b_2^2}} \quad (1)$$

where C_1 and C_2 are the color vectors shown in Fig. 1.

2.2. Major Color Spectrum Histogram Representation

By using the concept of color distance, we can scale down the possible colors to a very limited number of "Major Colors" (for example, 15 to 100) without losing much accuracy on representing a (moving) object. This statement is true in many practical cases where, for each moving object, there exist a certain number of major (principal) colors, which are retained in the representation, while colors that rarely appear are discarded [8,9]. Colors within a given mutual distance threshold are dealt with as a single color. An example of major color representation is shown in Fig. 2.



(a) 'tn_flower' picture (b) MCR Histogram (threshold: 0.01)

Figure 2. Major Color Spectrum Histogram Representation (MCSHR) of 'tn_flower'

The example picture ('tn_flower') is shown in Fig. 2 (a). In this picture, we can see that the most frequent colors are around dark green-black and yellow values. Fig. 2 (b) shows us the histogram of the major colors under the color distance threshold of 0.01. In the histogram, we can see that there are 4 main dark green-black bins with the highest frequencies (bins 1-4). The number of dark green-black pixels falling in these bins is about 350, 125, 120 and 85 respectively. The yellow colors are distributed in color spectrum bins 5, 6, 7, 8, 9, 11, 12 and 15. The number of pixels of in these bins varies between about 60 and 30. There are also 3 dark green-black bins spread in bins 10, 13 and 14, with counts between 40 and 35.

3. Representation and track matching

3.1. Moving objects similarity measurements

The similarity measurement between two moving objects is based on the major color spectrum histogram of the two moving objects [10, 11]. We assume that there are *M* major colors in the spectrum of moving object A, which can be represented as:

$$MCS(A) = \{C_{A_1}, C_{A_2}, \cdots, C_{A_M}\}. \quad (2)$$

where $C_{A_i}, i = 1,2,\cdots,M$ is the color vector (RGB) of major colors in object A. Object A's color spectrum histogram (i.e. the frequencies) can be represented as:

$$p(A) = \{p(A_1), p(A_2), \cdots, p(A_M)\}. \quad (3)$$

Similarly, the major color spectrum of object B can be represented as following:

$$MCS(B) = \{C_{B_1}, C_{B_2}, \cdots, C_{B_N}\}. \quad (4)$$

where $C_{B_j}, j = 1,2,\cdots,N$ are the color vectors (RGB) of major colors in object B. Object B's color spectrum histogram can be represented as:

$$p(B) = \{p(B_1), p(B_2), \cdots, p(B_N)\}. \quad (5)$$

In order to define the similarity between two moving objects, a *visibility measurement* of major colors C_{A_i} from moving object B's major color $MCS(B)$ is defined as:

$$p(A_i \mid B) = \min\{p(A_i), \sum_{C_{B_j}: d(C_{A_i}, C_{B_j}) < \sigma} p(B_j)\} \quad (6)$$

where $j = 1,2,\cdots,N$. The above equation shows us that the visibility of A from B is given by the sum of histogram values of all major colors in moving object B that are close to the color C_{A_i} (i.e. the color distance between C_{A_i} and C_{B_j} is less than a threshold σ, for example, 0.01, i.e. $d(C_{A_i}, C_{B_j}) < \sigma$). The 'min' operation ensures that $\sum p(A_i \mid B) \le \sum p(A_i)$. So, the similarity of moving object B to moving object A is defined as:

$$Similarity(B \to A) = \frac{\sum_{i=1,2,\cdots,M} p(A_i \mid B)}{\sum_{i=1,2,\cdots,M} p(A_i)} \quad (7)$$

Similarly, the similarity of moving object A to moving object B is defined as:

$$Similarity(A \to B) = \frac{\sum_{j=1,2,\cdots,N} p(B_j \mid A)}{\sum_{j=1,2,\cdots,N} p(B_j)} \quad (8)$$

where $p(B_j \mid A)$ is defined as:

$$p(B_j \mid A) = \min\{p(B_j), \sum_{C_{A_i}: d(C_{A_i}, C_{B_j}) < \sigma} p(A_i)\} \quad (9)$$

where $i = 1,2,\cdots,M$, and $j = 1,2,\cdots,N$.

If both visual objects are the same physical object, both similarities will be high (close to 1.0). Otherwise, both will be low, or at least one will be low (much lower than 1.0).

So, we define the overall similarity between moving object A and moving object B as:

$$Similarity(A, B) = \min\{Simililarity(A \to B), Simililarity(B \to A)\} \quad (10)$$

3.2. Multi-Frame Incremental Major Color Spectrum Histogram Representation

In order to cope with small pose changes occurring along the track, a multi-frame, incremental major color spectrum histogram representation is proposed here as shown in Fig. 3.

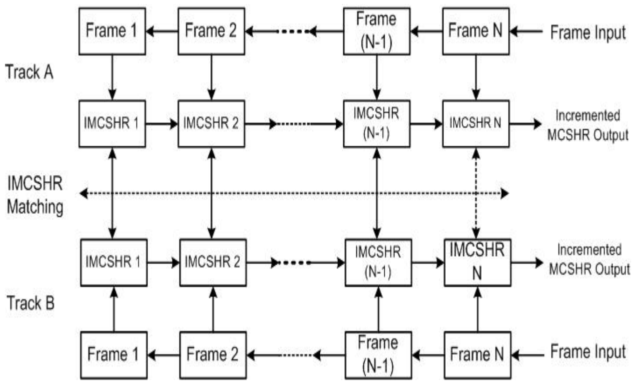


Figure 3. Multi-Frame Incremental Major Color Spectrum Histogram Representation Algorithm

With such a representation, the major colors of frame F_i are computed not only on the frame itself, but on the window of the last N frames $\{F_{i-N}, F_{i-N+1}, \ldots, F_{i-1}, F_i\}$. First, the major colors of frame F_{i-N} are computed as described in Section 2. Then, instead of computing the major colors of frame F_{i-N+1} again from scratch (i.e. starting from an empty table), we compute them starting from those of frame F_{i-N}. We proceed with the other frames in a similar way to eventually obtain the Incremental MCHSR (I-MCHSR) of frame F_i. For those frames with index $j < N$ in the initial part of the track, we restrict the incremental computation to the available frames only $\{F_1, \ldots, F_j\}$. (Alternatively, such frames could be simply dismissed, but we prefer a partial computation to prevent dismissing a significant part of the track).

Thus, in Fig. 3 the matching between I-MCSHR 1 in tracks A and B is actually the single frame MCSHR matching since there is no incremental MCSHR available yet. The matching between I-MCSHR N in tracks A and B is instead an N-frame integrated MCSHR matching. In order to study the effects of the integrated MCSHR on the matching process, the matching were carried out at each stage of MCSHR integration as shown in Fig. 3.

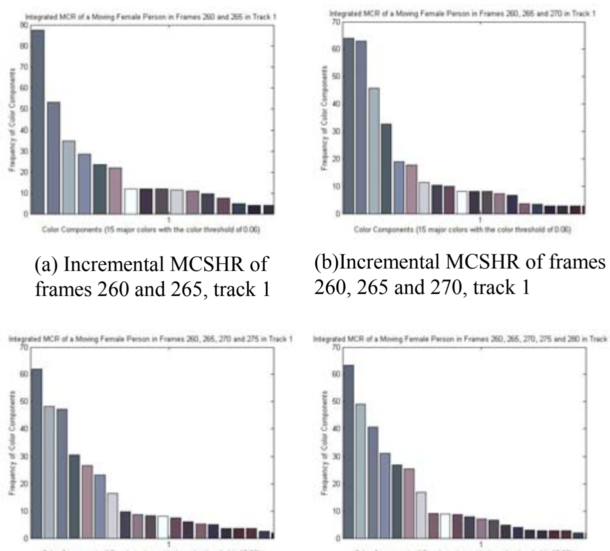
4. Experimental results and analysis

In our experiments, we segment and track moving objects based on [6, 7]. In this paper, we report example results from five typical tracks where three moving objects have been detected and tracked, namely a female subject, a male subject and a white van. In all the experiments reported, the I-MCSHR at frame F_i is obtained by integrating all the previous frames. The segmented moving objects, major color spectrum histograms and experimental results are shown in the following sections.

4.1. Multi-frame Incremental Major Color Spectrum Histogram Representation (MCSHR)

The proposed multi-frame Incremental Major Color Spectrum Histogram Representation is depicted here on data in which a moving female person has been detected and tracked ("track 1", frames 260-280 in steps of five frames). The moving object in these frames and its incremental major color spectrum histograms (I-MCSHRs) are shown in Figs. 4 and 5 respectively.

Figure 4. Frame sequence from track 1



(a) Incremental MCSHR of frames 260 and 265, track 1

(b)Incremental MCSHR of frames 260, 265 and 270, track 1

(c) Incremental MCSHR of frames 260, 265, 270 and 275, track 1

(d) Incremental MCSHR of frames 260, 265, 270, 275 and 280, track 1

Figure 5. Incremental MCSHR for frames in Fig. 4

Fig. 5 shows us that with the increase in the number of integration frames (from 2 to 5), the major color spectrum histogram representation tends to become more stable, in the sense that the actual main colors of the person (approximately, the first half of the bins in the histograms in Fig. 5) are emphasised with respect to the other colors.

Practically, since there always exist errors in the process of moving object detection, together with small pose changes, an integration over a few frames (around 4 or 5) proves helpful for a more accurate and robust moving object representation. Further increase in the number of integrated frames does not necessarily improve the representation accuracy as shown in the following sections 4.2 and 4.3.

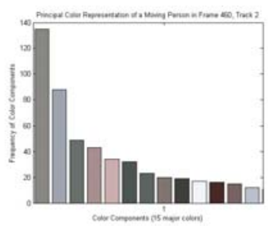
4.2. A same moving person in two different tracks

The first case here reported is from a same person in two different tracks (track 1, frames 260-280, and track 2, frames 460-480 in steps of five frames). A typical frame (460 in track 2), extracted moving object, moving object mask and its major color histogram are shown in Fig. 6.

(a) A moving female person in frame 460, track 2

(b) Extracted moving object

(c) Moving object mask



(d) The MCSHR of a moving female person in frame 460 track 2 (color threshold: 0.06)

Figure 6. A moving female person (frame 460, track 2)

The results for this first case are reported in Table 1 (for a color distance threshold of 0.01, number of MCSHR of 30, similarity threshold of 0.05 and significance threshold of 1.5). The results show us that:

1) With the increase in the number of integration frames from 1 to 5, the similarities increase from 85% to 97%. Hence, the proposed Incremental MCSHR algorithm can make the moving object major color representation more accurate and the matching more robust.
2) Increasing the number of major colors or decreasing the color distance threshold will increase the accuracy of the moving object major color representation, but also increase the amount of calculation at the same time. Thus, we kept the number of major colors for each frame and the color distance threshold as constants (30 and 0.01, respectively, proved adequate values in the experiment). The significance threshold was set to 1.5, and any color with lower significance was ignored.

Table 1. Incremental MCSHR matching results for the first case

Test Case	No of Frames	Frame Index	Track No	I-MCSHR Similarity	Matching Results(1/0)
1	1	260	1	0.8537	1 (Yes)
		460	2		
2	2	260-265	1	0.8913	1 (Yes)
		460-465	2		
3	3	260-270	1	0.9146	1 (Yes)
		460-470	2		
4	4	260-275	1	0.9461	1 (Yes)
		460-475	2		
5	5	260-280	1	0.9700	1 (Yes)
		460-480	2		

As stated above, the number of major colors used in the representation can have great influence on the matching performance. Table 2 gives evidence to that by showing the matching results of a four-frames integration MCSHR with respect to the change in the number of major colors (for color threshold of 0.01 and similarity threshold of 0.05).

Table 2. Incremental MCSHR matching results with respect to the number of major colors used

No of I-MCSHR	4	
Frame Index No	260-275	460-475
Track No	1	2
Similarity (MC:10)	0.7917	
Similarity (MC:15)	0.8997	
Similarity (MC:20)	0.9051	
Similarity (MC:25)	0.9106	
Similarity (MC:30)	0.9461	

The results in Table 2 show that the matching rate has increased from 79% to 95% with the increase of the number of major colors from 10 to 30. We consider this last value as a good trade-off between accuracy and computational load.

4.3. Two different people in two different tracks

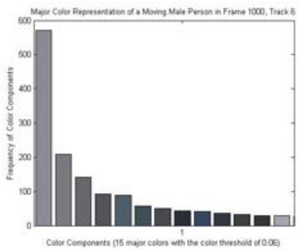
The multi-frame Incremental MCSHR matching results of a moving female person in a track (track 2, frames 460-505) and a moving male person in another track (track 6, frames 100-1045), with color distance threshold of 0.01, number of I-MCSHR of 30, similarity threshold of 0.01 and significance threshold of 1 are shown in table 3. A typical frame (1000 in track 6), extracted moving object, moving object mask and its major color histogram are shown in Fig. 7. The two moving objects in this test case are obviously different individuals. However, the matching test is not trivial as the two people have similar color components and similar size in images.

The test results in Table 3 show us that in all test cases, the similarities are between 0.09 and 0.19, so the two different objects were clearly discriminated.

(a) A moving male person in track 6, frame 1000

(b) Extracted moving object

(c) Moving object mask



(d) The Incremental MCSHR of a moving male person in frame 1000 track 6 (color threshold:

Figure 7. A moving van in frame 1000 track 6

Table 3. Incremental MCSHR Matching Results (Number of MCSHR: 30; Color Threshold: 0.01; Track 2, Frames 460-505/Track 6, Frames 1000-1045)

Test Case	No of Frames	Frame Index	Track No	I-MCSHR Similarity	Matching Results(1/0)
1	1	460	2	0.1067	0 (No)
		1000	6		
2	2	460-465	2	0.0891	0 (No)
		1000-1005	6		
3	3	460-470	2	0.1021	0 (No)
		1000-1010	6		
4	4	460-475	2	0.1054	0 (No)
		1000-1015	6		
5	5	460-480	2	0.1599	0 (No)
		1000-1020	6		
6	6	460-485	2	0.1382	0 (No)
		1000-1025	6		
7	7	460-490	2	0.1325	0 (No)
		1000-1030	6		
8	8	460-495	2	0.1504	0 (No)
		1000-1035	6		
9	9	460-500	2	0.1737	0 (No)
		1000-1040	6		
10	10	460-505	2	0.1888	0 (No)
		1000-1045	6		

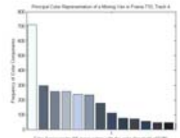
4.4. A same moving van in two different tracks

The multi-frame Incremental MCSHR matching results on a white moving van (track 4, frames 700-745, and track 5, frames 900-945) are shown in Table 4. The typical frames (710 in track 4 and 900 in track 5), extracted moving object, moving object mask and its major color histogram are shown in Figs. 8 and 9.

(a) A moving van in frame 710, track

(b)Extracted moving object

(c) Moving object mask



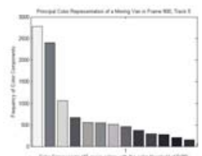
(d) The MCSHR of the moving van (color threshold: 0.06)

Figure 8. A moving van in frame 710 track 4

(a) A moving van in frame 900 track 5

(b)Extracted moving object

(c) Moving object mask



(d) The MCSHR of the moving van (Color threshold: 0.06)

Figure 9. A moving van in frame 900, track 5

Table 4. Incremental MCSHR Matching Results (Number of MCSHR: 30; Color Threshold: 0.01; Track 4, Frame 700-745/Track 5, Frame 900-945)

Test Case	Frame Index No	Track No	I-MCSHR Similarity	Matching Results(1/0)
1	700	4	0.8214	1 (Yes)
(1 frame)	900	5		
2	700-705	4	0.8423	1 (Yes)
(2 frames)	900-905	5		
3	700-710	4	0.8723	1 (Yes)
(3 frames)	900-910	5		
4	700-715	4	0.8348	1 (Yes)
(4 frames)	900-915	5		
5	700-720	4	0.7981	1 (Yes)
(5 frames)	900-920	5		
6	700-725	4	0.7725	1 (Yes)
(6 frames)	900-925	5		
7	700-730	4	0.7628	1 (Yes)
(7 frames)	900-930	5		
8	700-735	4	0.7647	1 (Yes)
(8 frames)	900-935	5		
9	700-740	4	0.7863	1 (Yes)
(9 frames)	900-940	5		
10	700-745	4	0.7994	1(Yes)
(10 frames)	900-945	5		

The test results in Table 4 show us that:

1) With the increase of the integration number from 1 to 3, the matching rate has increased from 82% to 87%
2) With the further increase in the integration number, the matching rate has been kept between 76% and 83%. The variant matching rate in test cases 4 to 10 might due to detection and tracking error, especially in this test, where some people passing by are accidentally detected and partially included in the moving van target.
3) In all 10 matching test cases, the moving objects can still be correctly matched by using a selective matching threshold of 75%.

5. Conclusion

In this paper, an incremental major color spectrum histogram track matching algorithm has been proposed to match tracks from single objects across non-overlapping camera views. First, a color distance based on a normalized geometric distance between two points in the RGB space is defined and used to measure similarity of two different colors. Then, a Major Color Spectrum Histogram representation (MCHSR) is introduced to represent a moving object by its "major colors" and their frequencies. After that, an Incremental MCHSR (I-MCSHR) is proposed to cope with small pose changes and detection errors occurring along the track. A similarity measurement is then used to measure the similarity of any two moving objects. Finally, track matching is based on the post-matching integration of single-frame matching.

Based on our experimental results, the following conclusions can be drawn:

1) Experimental results showed that the proposed representation can describe moving objects accurately with a limited number of colors and their frequencies.
2) Thanks to the incremental representation, the similarity of a same moving object in two different tracks improves from 85% to 97% with the number of integrated frames increasing from 1 to 5, while the similarity of two different moving objects has been kept as low as 9% to 19%.
3) With values of 4 or 5 for the number of integrated frames, the similarities between two different objects have been kept at a very low level, but further increase in this value does not necessarily improve the matching performance.

The proposed incremental MCSHR track matching algorithm has the potential to substantially extend current video surveillance applications by providing accurate tracking across non-overlapping camera views, which is the actual case for real surveillance camera networks. Moreover, to the best of our knowledge, the proposed approach is one of the few in the current literature to tackle the problem of track matching across disjoint camera views [12, 13]. Differently from previous papers, our approach does not require global track matching [12] or rely on a topographic model of the camera network [13].

This research is supported by the Australian Research Council, ARC Discovery Grant Scheme 2004 (DP0452657).

References

[1] T.H. Chang and S. Gong, "Tracking Multiple People with a Multi-Camera System", Proceedings of the 2001 IEEE Workshop on Multi-Object Tracking, 19-26, 2001.

[2] I. Haritaoglu, D. Harwood and L.S. Davis, "W4 Real-Time Surveillance of People and Their Activities", IEEE Trans. on PAMI., 22(8), 809-830, 2000.

[3] L.M. Fuentes and S.A. Velastin, "People Tracking in Surveillance Applications", Proceedings of the 2nd IEEE Workshop on Performance Evaluation of Tracking and Surveillance, 2001.

[4] S. McKenna, Y. Raja, and S. Gong, "Tracking color objects using adaptive mixture models," Image and Vision Computing, vol. 17, pp. 225-231, 1999.

[5] A. Lipton, H. Fujiyoshi, and R. Patil, "Moving target classification and tracking from real-time video," in Proc. Of the IEEE Image Understanding Workshop, 1998, pp. 129-136.

[6] A. Elgammal, D. Harwood, and L. Davis. "Non-parametric Model for Background Subtraction", 6th European Conference on Computer Vision, 2000.

[7] A. Senior, A. Hampapur, Y.-L. Tian, L. Brown, S. Pankanti, and R. Bolle, "Appearance Models for Occlusion Handling", 2nd Int. Workshop on Performance Evaluation of Tracking and Surveillance Systems, 2001.

[8] Zoran Zivkovic and Ben Krose, "An EM-like algorithm for color-histogram-based object tracking," in Proc. IEEE CVPR 2004.

[9] W. Lu and Y. P. Tan, "A Color Histogram Based People Tracking System", Proc. IEEE Int'l Symp. Circuits and Systems, vol. 2, pp. 137-140, 2001.

[10] Y. Chen and E. Wong, "Augmented image histogram for image and video similarity search," in Proc. SPIE Storage and Retrieval for Image and Video Databases," pp. 523-532.

[11] J. Hu and A. Mojsolovic, "Optimum color composition matching of images," in Proc. 15th Int. Conf. on Pattern Recognition," vol. 4, pp. 47-51, Barcelona, Spain, 2000.

[12] O. Kaved, Z. Rasheed, K. Shafique, M. Shah, "Tracking Across Multiple CamerasWith Disjoint Views," in Proc. of the Ninth IEEE Int. Conf. on Computer Vision (ICCV'03), vol. 2, pp. 952-957.

[13] D. Makris, T. Ellis, and J. Black, "Bridging the Gaps between Cameras," in Proc. of the 2004 IEEE CS Conf. on Computer Vision and Pattern Recognition, vol. 2, pp. 205-210.