MATCHING OF OBJECTS MOVING ACROSS DISJOINT CAMERAS

Eric Dahai Cheng and Massimo Piccardi, Senior Member, IEEE

Faculty of Information Technology, University of Technology, Sydney (UTS), Australia
E-mail: {cheng, massimo}@it.uts.edu.au

ABSTRACT

Matching of single individuals as they move across disjoint camera views is a challenging task in video surveillance. In this paper, we present a novel algorithm capable of matching single individuals in such a scenario based on appearance features. In order to reduce the variable illumination effects in a typical disjoint camera environment, a cumulative color histogram transformation is first applied to the segmented moving object. Then, an incremental major color spectrum histogram representation (IMCSHR) is used to represent the appearance of a moving object and cope with small pose changes occurring along the track. An IMCHSR-based similarity measurement algorithm is also proposed to measure the similarity of any two segmented moving objects. A final step of post-matching integration along the object's track is eventually applied. Experimental results show that the proposed approach proved capable of providing correct matching in typical situations.

Index Terms—Object tracking, major color spectrum histogram representation, disjoint camera views.

1. INTRODUCTION

Computer vision-based tracking of moving objects can be based on coherency of shape, motion and appearance features ([1, 2]). However, in the case of people tracking, shape features are not immediate to exploit since humans are deformable objects. Moreover, when camera views are multiple and disjoint, motion coherency may not be assessed when the tracked objects move across different camera views. Actually, the definition of disjoint camera views in itself implies that the extent of separation between views prevents prediction of the path and timings of single moving objects. In such a case, appearance features are the main cue to reconcile the tracks from separate camera views of a same physical object. Such a scenario is very common in real-life situations where existing camera networks cannot provide full coverage of the monitored space nor measure accurate individual biometrics.

The illumination conditions between disjoint cameras can be significantly different and have great influence on the appearance of moving objects. Illumination effects must be eliminated or at least reduced in order to make the appearance of same object comparable. Color constancy algorithms have been proposed for this in the literature (see [3] for references). However, it is difficult to apply an exact color constancy algorithm in our scenario since little is known about the scene and the objects. Approaches based on intrinsic images aim to separate pure reflectance images from illumination images, with effective methods available for image sequences [4]. In [5], Javed *et al*. proposed to learn the transfer functions between cameras during a learning phase. In this paper, instead we propose to use a simple cumulative color histogram transformation to compensate for the varying illumination conditions. We also make use of an object matching algorithm based on an incremental major color spectrum histogram representation (IMCSHR), a similarity measurement to assess the similarity of any two tracked objects, and a post-matching integration phase to make the whole matching more robust [6]. Differently from previous papers, our approach does not require global information about objects in the surveillance systems [7] nor rely on a topographic model of the camera network [8].

2. MAJOR COLOR SPECTRUM HISTOGRAM

In the RGB color space, using one byte to represent each color yields a total of 16.8 million different colors. It is, in general, very difficult to compare two objects based on so many possible values. By using the concept of color distance [6], we can scale down the number of colors from 16.8 million to a very limited number of "major colors" without losing much accuracy in representing a moving object. For each moving object, a given percentage of major colors are retained in the representation, while colors that rarely appear are discarded [9-11]. Colors within a given mutual distance threshold are dealt with as a single color. An example of such a major color representation is shown in Fig. 1.
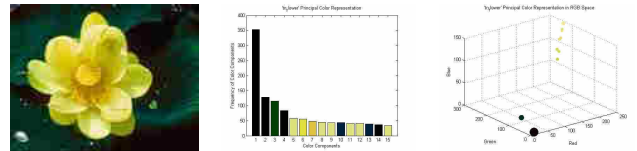
(a) 'tn_flower' picture   (b) MCSHR Histogram   (c) MCSHR in RGB Space

Figure 1 The Major Color Spectrum Histogram Representation (MCSHR) of the 'tn_flower'.

An example picture is shown in Fig. 1 (a), in which we can see that the most frequent colors are around dark green and

yellow values. Fig. 1 (b) shows us the histogram of the major colors under the color distance threshold of 0.01. Fig. 1 (c) shows us the positions of these major colors in the RGB space, with the size of the color spheres proportional to the frequency of each color.

3. MOVING OBJECTS IMAGE PRE-PROCESSING FOR THE DISJOINT CAMERAS

The biggest challenge for the matching of moving objects from disjoint camera views is in the different and varying illumination causing great differences in their appearances. For example, a same object may look very bright or very dark depending on the actual illumination. In order to reduce its effects, a cumulative color histogram transformation algorithm is proposed here. Fig. 2 shows the main steps: first, the bounding box of a moving object is located in the frame. An example of input frame and the located moving object is shown in Figs. 3 (a) and (b). In the second step, the color histograms ($p_r(r)$, $p_g(r)$ and $p_b(r)$.) are calculated for the purpose of cumulative color histogram transformation, in which the background pixels are replaced with virtual pixels of the size of moving object that are equalized in RGB. Then, the three color histogram equalization transforms (T_r, T_g and T_b) are calculated based on the moving object and virtual background and applied to the moving object image. An example of transformed moving object is shown in Fig. 3 (c).
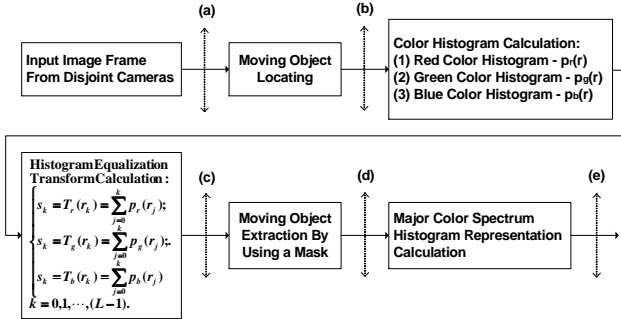

Figure 2 Moving objects image pre-processing for disjoint camera views

By comparing Figs. 3 (b) and (c), we can see that the contrast of moving object has been significantly improved which can prove helpful for the matching process either by computers or humans. After this step, the MCSHR is computed on the object's "blob". The example of the moving object mask and the MCSHR after transformation with color threshold of 0.01 is shown in Figs. 3 (d) and (e) respectively. Please note that throughout the paper, the "blob" of the extracted object has been corrected manually to avoid mixing different aspects such as power of the proposed representation and segmentation errors. In Section 3, we will discuss an approach to mitigate the impact of the unavoidable segmentation errors on the representation.

The corresponding cumulative color histogram transformation and the histograms before and after transformation in red are shown in Figs. 3 (f-h), showing the re-mapped ranges of the red channel.
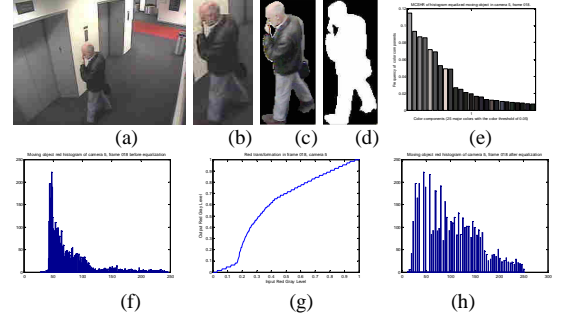

Figure 3 Example of pre-processed moving object (reference: camera 5, frame 018).

4. IMCSHR AND TRACK MATCHING

4.1 Moving Objects Similarity Measurements

In this section, a similarity measurement based on a most-similar color search is proposed to measure the similarity between two moving objects. This algorithm is based on the major color spectrum histogram of the two moving objects. We assume that there are M major colors in the spectrum of moving object A, which can be represented as:

$$MCSHR(A) = \{ C_{A_1}, C_{A_2}, \cdots, C_{A_i}, \cdots, C_{A_M} \} \qquad (1)$$

where $C_{A_i}, i = 1, 2, \cdots, M$ is a major color (RGB) in object A. Object A's major color bin counts can be represented as:

$$p(A) = \{ p(A_1), p(A_2), \cdots, p(A_i), \cdots p(A_M)\} \qquad (2)$$

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

$$MCSHR(B) = \{ C_{B_1}, C_{B_2}, \cdots, C_{B_N} \} \qquad (3)$$

$$p(B) = \{ p(B_1), p(B_2), \cdots, p(B_N)\}. \qquad (4)$$

In order to define the similarity between two moving objects, a subset of $MCSHR(B)$ in equation (3) is firstly defined as:

$$MCSHR^{'}(B / C_{A_i}, \sigma) = \{ C_{B_1^{'}}, C_{B_2^{'}}, \cdots, C_{B_L^{'}} \} \qquad (5)$$

where the distance between $C_{B_j^{'}}, j = 1, 2, \cdots, L$ and C_{A_i} is less than a given threshold σ.

Then, $C_{B_{j/A_i}}$ is defined as the most similar color to C_{A_i} in subset $MCSHR^{'}(B)$ satisfying:

$$C_{B_{j/A_i}} : j = \underset{k=1, \cdots, L}{arg\ min}\{ d(C_{B_k^{'}}, C_{A_i})\} \qquad (6)$$

Differently from [14], we consider only the color of B closest to A instead of the whole subset $MCSHR^{'}(B)$. In this way, we aim to achieve a more selective similarity criterion capable of limiting false matches. The portion of C_{A_i} in object A can be simply calculated as:

$$p_{norm}(A_i) = \frac{p(A_i)}{\sum\limits_{i=1,2,\cdots,M} p(A_i)} \qquad (7)$$

Similarly, the portion of C_{B_j/A_i} in object B can be calculated as:

$$p_{norm}^{[A_i]}(B_j) = \frac{p^{[A_i]}(B_j)}{\sum\limits_{j=1,2,\cdots,N} p(B_j)} \qquad (8)$$

where $p^{[A_i]}(B_j)$ is the frequency of C_{B_j/A_i}. Then, the similarity of color C_{A_i} in object A with its corresponding color C_{B_j/A_i} in object B is defined as:

$$Sim(C_{A_i}, C_{B_j/A_i}) = min\{ p_{norm}(A_i), p_{norm}^{[A_i]}(B_j)\} \qquad (9)$$

The *min()* operator aims to achieve symmetric measurements when searching color matches from A to B and vice versa. However, since the "closest to" relationship is not reciprocal, strictly symmetric measurements are not possible and the *min()* operator not strictly necessary. Then, the similarity of the whole objects A and B in the direction from A to B is then defined as:

$$Sim(A,B) = \sum\limits_{i=1}^{M} Sim(C_{A_i}, C_{B_j/A_i}) \qquad (10)$$

In the same way, the similarity of color C_{B_j} in object B with its corresponding color C_{A_i} in object A is defined as:

$$Sim(C_{B_j}, C_{A_i/B_j}) = min\{ p_{norm}(B_j), p_{norm}^{[B_j]}(A_i)\} \qquad (11)$$

and the object similarity of objects A and B in the direction from B to A is:

$$Sim(B,A) = \sum\limits_{j=1}^{N} Sim(C_{B_j}, C_{A_i/B_j}) \qquad (12)$$

In order to derive a symmetric similarity measurement, the minimum and maximum of equations (10) and (12) are defined as:

$$Sim_{min}(A,B) = min\{ Sim(A,B), Sim(B,A)\} \qquad (13)$$

$$Sim_{max}(A,B) = max\{ Sim(A,B), Sim(B,A)\} \qquad (14)$$

and eventually combined into a single final value, Similarity(A,B). If Simmin(A,B) is less than a given discrimination threshold, ηdiscrim, the similarity of objects A and B is simply defined as:

$$Similarity(A,B) = Sim_{min}(A,B) \qquad (15)$$

The rationale is that in this case the two object similarities between A and B, (14) and (16), are either very asymmetric or both low and for this reason we decide to bound them by their lowest value. Instead, if $Sim_{min}(A,B)$ is above or equal the discrimination threshold, we define:

$$Similarity(A,B) = 1 - \frac{Sim_{max}(A,B) - Sim_{min}(A,B)}{Sim_{max}(A,B) + Sim_{min}(A,B)} \qquad (16)$$

In this case, we are confident that the two visual objects are possibly a same physical one. As a further verification, we choose to check the difference between the maximum and minimum similarities in a ratio form. In (16) we can see that the bigger the difference between maximum and minimum similarity, the less similar are considered the two objects. Such a definition aims to prevent asymmetric, partial matches between two objects and let us set the final

similarity threshold more easily. Eventually, matching is assessed if *Similarity(A,B)* is above the chosen similarity threshold. A multi-frame, incremental major color spectrum histogram matching and post-matching integration were described in [11].

5. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we report example results from four typical tracks from three real disjoint video surveillance cameras installed in the Faculty of Information Technology building, University of Technology, Sydney, where two moving objects have been detected and tracked. The segmented moving objects, major color spectrum histograms and experimental results are shown in the following sub-sections.

5.1 The Matching of the Same Moving Person in Disjoint Camera Views

The test data here reported are from the same person recorded from two disjoint video surveillance cameras tracks (reference: camera 3a, frames 001-019, and camera 5, frames 300-318.), with some of the frames shown in Figure 4. The two cameras are significantly disjoint in both space and time and the person's appearance in the two tracks could not be matched trivially. Moreover, illumination varies significantly also with the object's position within each camera view (unlike assumptions in [10, 12]). However, our representation proves capable of coping with such variations in appearance.
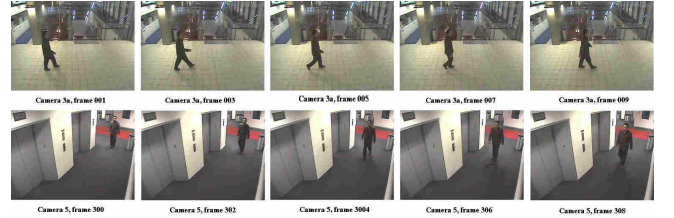

Figure 4 Moving objects from camera 3a, frames 001-009 and camera 5, frames 300-308.

IMCSHR matching results are reported in Table 1 showing that the same moving object in the two disjoint camera views is reliably matched.

Table 1 Results of IMCSHR Matching and Post-Integration

Test Case	Frame No	Camera	Similarity	Matching Results
1	001-005	3a	0.9817	1 (Yes)
	300-304	5		
2	003-007	3a	0.9758	1 (Yes)
	302-006	5		
3	005-009	3a	0.9772	1 (Yes)
	304-308	5		
4	007-011	3a	0.9856	1 (Yes)
	306-310	5		
5	009-013	3a	0.9452	1 (Yes)
	308-312	5		
Integration	001-019	3a		100% (Match)
	300-318	5		

Note: with 90% major colors cut off, color threshold = similarity color

threshold = 0.05, discrimination threshold = 0.4, IMCSHR matching threshold = 0.8, and final integration matching threshold = 80%.

5.2 The Matching of Two Different People from Two Disjoint Camera Views

The test data reported here are from two different people recorded from the same video surveillance cameras (camera *3a*, frames 001-019, and camera *5*, frames 010-022), with some of the frames shown in Figure 5.
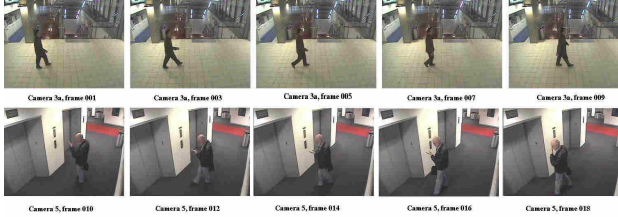

Figure 5 Moving objects from camera 3a, frames 001-009 and camera 5, frames 010-018.

Table 2 Results of IMCSHR Matching for two different people

Test Case	Frame No	Camera	Similarity	Matching Results
1	001-005	*3a*	0.3538	0 (No)
	010-014	*5*		
2	003-007	*3a*	0.7588	0 (No)
	012-016	*5*		
3	005-009	*3a*	0.7224	0 (No)
	014-018	*5*		
4	007-011	*3a*	0.8348	1 (Yes)
	016-020	*5*		
5	009-013	*3a*	0.8075	1 (Yes)
	018-022	*5*		
Integration	001-019	*3a*		40% (No match)
	010-022	*5*		

Note: same parameters as for Table 1.

6. CONCLUSIONS

In this paper, we have proposed a method for matching two objects along their tracks from disjoint camera views. Such views are challenging in that illumination conditions can be very different across cameras and time and the appearance of single objects vary enormously as a consequence. The main contribution of this paper is the proposal of a cumulative color histogram transformation to compensate for the varying illumination conditions across the disjoint views. When integrated in our object matching approach, the matching of a same object improved significantly, proving this representation to be maximally invariant to illumination. At the same time, the transformation scatters colors in a range broader than the original so that two different objects can be more easily discriminated. Experimental results also show that the object matching algorithm can measure the similarity of the two moving objects accurately, and with three to five frames integration, the proposed IMCSHR algorithm can make the matching more robust and reliable

than single-frame matching, especially for small pose changes.

The overall object matching procedure can provide video surveillance applications with the ability of tracking single objects across disjoint camera views, which are predominant in existing surveillance camera networks. Such an ability could prove useful to track assigned individuals (a "watch list") from entry to exit of a building in real time, or as a forensic tool to automatically back-track movements of people from an assigned point in time and space (such as an event of interest). At the moment, we are working on automated estimate of the parameters based on image statistics and considering adding shape invariants to appearance features for more general matching.

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] 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.

[3] G. D. Finlayson and S. D. Hordley, "Color constancy at a pixel," J. Opt. Soc. Am. A., 18(2), 253-264, 2001.

[4] Y. Weiss, "Deriving intrinsic images from image sequences," Proc. of Int. Conf. on Computer Vision, 2001.

[5] O. Javed, K. Shafique, M. Shah, "Appearance modeling for tracking in multiple non–overlapping cameras," in Proc. of IEEE CVPR 2005.

[6] M. Piccardi and E. D. Cheng, "Track Matching Over Disjoint Camera Views Based On An Incremental Major Color Spectrum Histogram", to appear in Proc. IEEE Int. Conf. on Advanced Video and Signal based Surveillance (AVSS 2005).

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

[8] 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.

[9] Liyuan Li, Weimin Huang, I. Y. H. Gu, K. Leman, Qi Tian, "Principal Color Representation for Tracking Persons," in Proc. of IEEE Int. Conf. on Systems, Man and Cybernetics 2003, vol. 1, pp. 1007-1012.

[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, 2000.