Matching Moving Objects by Parts with a Maximum Likelihood Criterion

Eric Dahai Cheng and Massimo Piccardi

Faculty of Information Technology, University of Technology, Sydney (UTS),
PO Box 123, Broadway NSW 2007, Australia

Email: {cheng, massimo}@it.uts.edu.au

Abstract

In this paper we present an algorithm for matching the appearance of two moving objects based on a matching-by-parts approach and a maximum likelihood criterion. We assume that the two moving objects to be matched are first extracted from videos by a preliminary foreground extraction-tracking step and our goal is that of matching their appearances. To this aim, we first consider the matching between single frames, one from each track. In order to increase the ability of discriminating between two different physical objects while keeping the matching rate of a single physical object high, each object is divided into N parts and then parts are matched in pairs. The appearance of each part is represented by a colour histogram (called MCSHR for short in the following) and a histogram similarity measure is used to compare two parts. The single-frame matching result is then obtained by fusing the similarities of each part matching. Later, our track matching algorithm extends the single-frame matching along the objects' tracks by a post-matching integration algorithm. Experimental results presented in this paper show that the proposed similarity measurement is accurate at the single-frame level.and that the post-matching integration makes the overall matching more robust and reliable.

Keywords: Moving object tracking, object matching by parts, maximum likelihood criterion, major colour spectrum histogram representation, colour distance, similarity measurement.

1 Introduction

Robustly tracking a single object throughout a network of cameras is an important function for effective video surveillance of wide areas [1-4]. However, in most real-world camera networks it is not possible to track a moving object through a continuity of overlapping camera views. Instead, most often the object has to completely exit from the view of a certain camera before it can reappear under the view of a different one. This common scenario is often referred to as disjoint camera views, where observations of a single object are disjoint in time and space to a certain extent. In order to allow tracking in such a scenario, single-camera views of a same object must be matched across neighbouring cameras.

In the following, we assume that each object is extracted and tracked within each single camera view by a preliminary foreground extraction-tracking step, and that the relevant information (the object's blob in each frame) is available – hereafter we call such sequence of blobs *track* for simplicity. Our goal is then that of matching tracks from disjoint views by using some objects' appearance features. To this aim, in this paper we present an algorithm for appearance matching based on a matching-by-part approach and a maximum likelihood criterion. First, we choose the two tracks to compare and consider the first frame in each. We compare the blobs from these two frames by dividing each blob into N parts, and orderly comparing parts in pairs. Each pair matching provides a similarity measurement, or matching belief, bounded between 0 and 1. The N results from part matching are then fused by an average rule and compared against a threshold set based on a maximum likelihhod criterion to provide the results at the frame level. The single-frame matching is repeated for following frames in the tracks and, eventually, such results are integrated to obtain the overall matching result between the two tracks.

The appearance representation is based on a colour histogram with sparse bins. A colour distance based on a geometric distance between two points in the RGB space is first introduced to measure the similarity of any two colours. By using the colour distance and a given threshold, pixels from each part are clustered into a limited number of bins, with each bin's frequency defined as the number of pixels falling into that bin. Such bins are then sorted in descending frequency order and a chosen percentage of them (in our work, 90%) is chosen as major colours to represent the part's appearance. We call this histogram the major colour spectrum histogram representation (MCSHR). A criterion is then defined to assess the similarity, bounded between 0 and 1, of the MCSHRs of two given parts.

To date, the problem of matching the appearance of objects across disjoint camera views has been addressed in relatively a few papers in the literature; [5] and [6] are notable examples. In both [5] and [6], the matching basd on appearance is reinforced by the use of priors based on statistics on travelling times between cameras acquired during a learning stage. The main problem that we identify with such an approach is that matching is more prone to fail in anomalous cases, which are instead those of interest for surveillance. For instance, if people remain in a blind area for long time in order to carry out activities such as tampering or stealing, their re-appearance under camera views will occur outside of statistical timing windows. For this reason, our approach deliberately avoids the use of time-based priors. Moreover, unlike [1], [5], [6], we use a part-based matching that prevents false matches between people with similar overall colours but with different spatial distribution.

2 Maximum Likelihood Criteria for Moving Objects Matching by Parts

2.1 Feature Space and Maximum Likelihood Criteria

The raw feature vectors in the observation space of the two matching moving objects are the major colours of the divided parts, shown in the following equations:

$$X_1 = [X_{11}, X_{12}, \cdots, X_{1N}] \tag{1}$$

$$X_2 = [X_{21}, X_{22}, \cdots, X_{2N}] \tag{2}$$

where X_{1i} and X_{2i} are major colour vectors of the *ith* divided parts in moving objects one and two, and *N* is the number of divided parts. Since the major colour vectors are multiple dimensional vectors, their distributions are very difficult to estimate. Therefore, the similarity between two matching objects is used as an observation variable (or one dimensional space) in the process of deriving an optimum matching structure based on maximum likelihood criteria.

The hypothesis test assumes:

$$\begin{aligned} H_0 &: sim = \mu_0 + n, \\ H_1 &: sim = \mu_1 + n. \end{aligned} \tag{3}$$

where n is the error noise that produced in the process of major colour similarity calculation, and based on experience of our experiments, we believe that the noise has Gaussian probability distribution with zero mean and variance σ_N^2, i.e. $n \in N(0, \sigma_N^2)$; μ_0 and μ_1 ($\mu_1 > \mu_0$) are the average similarities when H_0 (objects are two physically different objects) and H_1 (objects are a single physical object) are true. For simplicity, we assume that μ_0 and μ_1 are constant and that variations are to be blamed on the noise component.

The above assumptions can be validated by testing the data reported in Tables 1 (for μ_0) and 2 (for μ_1) in Sections 5.1 and 5.2, respectively. In this case, $\mu_0 = 0.4638$ and $\mu_1 = 0.7843$ and assumption $\mu_1 > \mu_0$ is verified. The standard deviations are $\sigma_0 = 0.046$ and $\sigma_1 = 0.056$. In the following, since their difference is small, we treat them as a same value.

Thus, the probability distribution function of *sim* under the hypothesis of H_0, is shown in equation (4).

$$p_0(sim) = \frac{1}{\sqrt{2\pi}\sigma_N} \exp[-\frac{(sim(X_1, X_2) - \mu_0)^2}{2\sigma_N^2}] \tag{4}$$

Similarly, the probability distribution function of ***sim***, under the hypothesis of H_1, is shown in equation (5).

$$p_1(sim) = \frac{1}{\sqrt{2\pi}\sigma_N} \exp[-\frac{(sim(X_1, X_2) - \mu_1)^2}{2\sigma_N^2}] \tag{5}$$

The likelihood ratio (LRT) is calculated as follows:

$$\Lambda(sim(X_1, X_2)) = \frac{P(sim(X_1, X_2) \mid H_1)}{P(sim(X_1, X_2) \mid H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \tag{6}$$

Taking the natural logarithm of both sides of equation (6) to obtain the log LRT:

$$\ln \Lambda(sim(X_1, X_2)) = \frac{(2sim(X_1, X_2) - \mu_0 - \mu_1)(\mu_1 - \mu_0)}{2\sigma_N^2} \underset{H_0}{\overset{H_1}{\gtrless}} \ln \eta \tag{7}$$

Equation (7) can be simplified as:

$$sim(X_1, X_2) \underset{H_0}{\overset{H_1}{\gtrless}} (\frac{\mu_1 + \mu_0}{2} + \frac{2\sigma_N^2 \ln \eta}{(\mu_1 - \mu_0)} = \lambda) \tag{8}$$

The above equation shows the optimum structure of the matching detector, in which the optimum threshold is the function of the average similarities - μ_0, μ_1, and the variations of similarity - σ_N^2. In the sense of maximum likelihood criteria, in order to minimize the total error (detection and false alarm) η should be 1, so the optimum threshold in equation (8) becomes $\lambda = (\frac{\mu_1 + \mu_0}{2})$.

2.2 Matching Performance Evaluation

Just a s a corollary, we show in the following that the matching performance can be easily evaluated in terms of the probability of detection - P_D and false alarm rate - P_{fa} - as a function of the average similarities, μ_0, μ_1, and variance σ_N^2.

The probability density functions (pdf) of matching objects under H_0 and H_1 described in equations (4) and (5) are shown in Fig. 1.

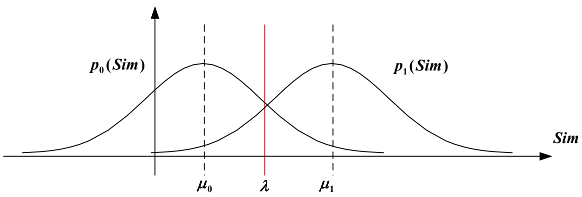


Figure 1: Similarity probability density functions.

In Fig. 1, the probability of false alarm matching - P_{fa} is the area under function $p_0(sim)$ above the detection threshold - λ, i.e.

$$\begin{aligned}P_{fa} &= \int_{\lambda}^{\infty} p_0(sim)dsim \\ &= \int_{\lambda}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_N} \exp[-\frac{(sim(X_1,X_2)-\mu_0)^2}{2\sigma_N^2}]dsim = \alpha\end{aligned} \quad (9)$$

The probability of the detection or correct matching - P_D - is the area under function $p_1(sim)$ above the detection threshold - λ, i.e.

$$\begin{aligned}P_D &= \int_{\lambda}^{\infty} p_1(sim)dsim \\ &= \int_{\lambda}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_N} \exp[-\frac{(sim(X_1,X_2)-\mu_1)^2}{2\sigma_N^2}]dsim = \beta\end{aligned} \quad (10)$$

Thus, equations (9) and (10) show that the probabilities of correct matching and mismatching are simple functions of the average similarities, μ_0, μ_1, and variance σ_N^2.

3 Major Colour Spectrum Histogram

3.1 Concept of Colour Distance

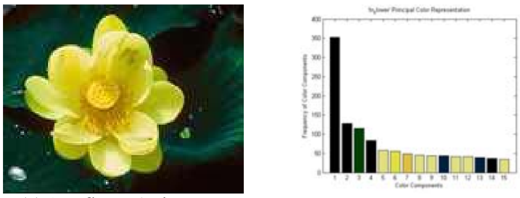
In this section, we first introduce the concept of "colour distance" between two colour pixels in the RGB space based on a normalized geometric distance between the two pixels. Such a geometric distance is defined in equation (11):

$$d(C_1,C_2) = \frac{\|C_1 - C_2\|}{\|C_1\|+\|C_2\|} = \frac{\sqrt{(r_1-r_2)^2+(g_1-g_2)^2+(b_1-b_2)^2}}{\sqrt{r_1^2+g_1^2+b_1^2}+\sqrt{r_2^2+g_2^2+b_2^2}} \quad (11)$$

C_1 and C_2 are the colour vectors. The smaller the colour distance, the more similar are the two colours.

3.2 Moving Object Major Colour Representation

By using the concept of colour distance, we can scale down the possible colours to a very limited number of "major colours" (for example, several hundreds) without losing much accuracy on representing a moving object. For each part of a moving object, a given certain percentage of major colours are retained in the representation, while colours that rarely appear are discarded [7, 8]. Colours within a given mutual distance threshold are dealt with as a single colour. An example picture ('tn_flower') is shown in Fig. 2 (a) in which we can see that the most frequent colours are around dark green-black and yellow values. Fig. 2 (b) shows that the histogram of the major colours (under the colour distance threshold of 0.01) seems a faithful representation of the image's colours and their frequencies.



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

Figure 2. The Major Colour Spectrum Histogram Representation (MCSHR) of the 'tn_flower'.

4 Single-Frame Matching and Post-Matching Integration Algorithm

4.1 Moving Objects Parts Similarity Measurements

In this section, a similarity measurement based on a most-similar-colour searching algorithm is proposed to measure the similarity between two corresponding parts of moving objects. This algorithm is based on the major colour spectrum histogram of the two corresponding divided parts of the two moving objects. We assume that there are *M* major colours in the spectrum of a divided part in a moving object *A*, which can be represented as:

$$MCSHR(A) = \{C_{A_1}, C_{A_2}, \cdots, C_{A_i}, \cdots, C_{A_M}\} \quad (11)$$

where $C_{A_i}, i = 1,2,\cdots,M$ is the colour vector (RGB) of the major colours in the divided part in object *A*. The major colour frequencies of the divided part in object A's can be represented as:

$$p(A) = \{p(A_1), p(A_2), \cdots, p(A_i), \cdots p(A_M)\} \quad (12)$$

The major colour spectrum histogram of object *B* can be represented similarly. In order to define the similarity of the corresponding divided parts of two moving objects, for each color C_{A_i} in *A*, a corresponding colour, $C_{B_j|A_i}$, is searched in *B* as:

$$C_{B_j|A_i} = \arg\min_{k=1,\cdots,L} \{d(C_{B_k^{'}}, C_{A_i}) < \sigma\} \quad (13)$$

$C_{B_j|A_i}$ is the closest colour to C_{A_i} within a threshold, σ, and $p^{[A_i]}(B_j)$ its frequency. Then, the similarity of C_{A_i} and $C_{B_j|A_i}$ is defined as:

$$Similarity(C_{A_i}, C_{B_j|A_i}) = \min\{p(A_i), p^{[A_i]}(B_j)\} \quad (14)$$

The similarity of the divided part in object A and its corresponding part in object B is obtained by adding up over $i = 1,2,\cdots,M$:

$$Similarity(A,B) = \sum_{i=1}^{M} Similarity(C_{A_i}, C_{B_j|A_i}) \qquad (15)$$

In a similar way we can obtain *Similarity*(*B,A*) that generally differs from *Similarity*(*A,B*) since the colour pairs defined by (13) may be different in the two directions. However, if A and B are the same physical object, these two similarities would be approximately symmetric. Therefore, in the final matching criterion we give importance to the symmetry of *Similarity*(*A,B*) and *Similarity*(*A,B*). We first define:

$$Similarity_{\min} = \min\{Similarity(A,B), Similarity(B,A)\} \qquad (16)$$

$$Similarity_{\max} = \max\{Similarity(A,B), Similarity(B,A)\}\,. \qquad (17)$$

Then, we combine them into a single final value, $Similarity_{A,B}$:

$$Similarity_{A,B} = \begin{cases} Similarity_{\min} & if\ Similarity_{\min} < \eta_{discrim} \\ 1 - \dfrac{Similarity_{\max} - Similarity_{\min}}{Similarity_{\max} + Similarity_{\min}} & otherwise \end{cases} \qquad (18)$$

If $Similarity_{min}$ is lower than a discrimination threshold, $\eta_{discrim}$, we bound $Similarity_{A,B}$ to it. Instead, if $Similarity_{min}$ is above or equal the discrimination threshold, we choose to check the difference between the maximum and minimum similarities in a ratio form for asymmetry. The bigger the difference between the maximum and minimum similarities, the lower is $Similarity_{A,B}$. Eventually, matching is assessed if $Similarity_{A,B}$ is above an assigned similarity threshold.

4.2 Similarity at the Whole-Object Level

Once obtained a similarity value, bounded between 0 and 1, for each pair of divided parts, the values for all the N part pairs need to be combined in order to obtain a single matching result at the whole-object level. For this, one can choose amongst popular fusion techniques such as the product rule, average rule or weigthed average rule [9]. The product rule suffers from the famous "curse of product" and should be used only in the case of complete statistical independence between the values to be fused. In our application, some degree of correlation instead certainly exists (two adjacent parts may share parts of a same piece of clothes and thus be materially correlated; the body shape deformates along the sequence, hence blob parts map on different bodily parts along frames) and therefore we cannot use the product. The weighted average rule requires a very well informed estimation of weights to be likely to outperform the (unweighted) average rule [9]. Therefore, we chose to use the latter in our approach. Equation (19) provides the required similarity at the whole-object level.

$$sim(X_1, X_2) = \frac{1}{N}\sum_{i=1}^{N} sim(X_{1i}, X_{2i}) \qquad (19)$$

4.3 Single-Frame Matching and Post-Matching Integration Algorithm

In the track matching algorithm, we consider the same number of frames from each track. Moving objects from corresponding frames in Track One and Track Two are matched based on similarity of their major colour spectrum, and the matching results are given as a binary decision.

The second step is the multi-frame post-integration, normalization, and thresholding. The advantages of this algorithm are:

- The single-frame matching is based on the major colour spectrum histogram and two direction similarities measurements, which makes the single-frame matching very accurate.
- The final conclusion is made based on the statistical average of single-frame matching. So, no detailed feature errors are carried forward after this stage, which makes the track matching conclusion more reliable than single frame matching.

5 Experimental Results and Analysis

In our experiments, we report example results from three typical tracks from the PETS 2001 dataset where moving objects have been detected and tracked. The segmented moving objects, major colour spectrum histograms and experimental results are shown in the following sections.

5.1 Matching-by-parts of Two Different Moving Objects

The first case reported here are from two different persons (track 1, frames 0400-0412 and track 2, frames 2150-2162), with two sets of typical extracted moving objects and object masks shown in Figure 3.

In the test, the moving objects are equally divided into seven parts along the vertical direction. The results from matching-by-parts at the single frame level and post-matching integration alomg the track with 90% of major colours, colour threshold of 0.01, discrimination threshold of 0.35, and matching threshold of 0.6241 are shown in Table 1. While other thresholds are empirical, the matching threshold, λ, is calculated as in equation (8) based on Tables 1 and 2. The results in Table 1 shows that all seven cases are correctly discriminated, with similarities at the whole-object level between 0.41 and 0.55, and the post-integration rate of 0%. Thus, the two tracks are reliably discriminated.

(1a) MO and mask in frames 0400 and 2150

Figure 3. Moving objects from track 1, frames 0400-0412 and track 2, frames 2150-2162.

Table 1. Matching similarities. (PETS 2001 dataset 1, frames 0400-0412 and 2150-2162, Color distance: 0.01, discrimination threshold: 0.35, MCSHR cut off: 90%, Number of divided parts: 7).

Test Case	Frame No	Part 1	Part 2	Part 3	Part 4	Part 5	Part 6	Part 7	Similarity (mean)	Matching Results
1	0400	0.6441	0.1823	0.1640	0.3388	0.8584	0.9225	0.0455	0.4508	0 (No)
	2150									
2	0402	0.7267	0.8094	0.2492	0.2133	0.2217	0.8907	0.7669	0.5540	0 (No)
	2152									
3	0404	0.2639	0.2999	0.1933	0.2234	0.3176	0.8233	0.7481	0.4099	0 (No)
	2154									
4	0406	0.1724	0.1889	0.1135	0.3267	0.8131	0.8657	0.6451	0.4465	0 (No)
	2156									
5	0408	0.2542	0.7222	0.2024	0.1705	0.3207	0.8100	0.9404	0.4886	0 (No)
	2158									
6	0410	0.2361	0.2152	0.1739	0.2480	0.7907	0.7070	0.7414	0.4446	0 (No)
	2160									
7	0412	0.9398	0.1681	0.0705	0.7958	0.6600	0.2402	0.2933	0.4525	0 (No)
	2162									
Post-Integration	0400-0412									0% (No)
	2150-2162									

5.2 Matching-by-parts of a Single Moving Object in Two Different Tracks

The test data reported here is from the same person in two different tracks (track 1, frames 2040-2052, and track 2, frames 2150-2162 in steps of five frames). The extracted moving object and moving object mask in typical frames (2048 in track 1, and 2156 in track 2) are shown in Fig. 4. The results from matching-by-parts at the single frame level and post-matching integration alomg the track with 90% of major colours, colour threshold of 0.01, discrimination threshold of 0.35, and matching threshold of 0.6241 are shown in Table 2. The results in Table 2 show us that in all seven cases, similarities were between 0.70 and .87, proving that the proposed matching-by-parts MCSHR algorithm offers an accurate appearance representation and similarity measurement. The post-integration of the seven individual matching cases is 1.0, thus the two tracks are reliably matched.

(a) MO and mask in frame 2048 (b) MO and mask in frame 2156

Figure 4. Moving objects from track 1, frames 2040-2052 and track 2, frames 2150-2162.

Table 2. Matching similarities. (PETS 2001 dataset 1, frames 0400-0412 and 2150-2162, Color distance: 0.01, Discriminate threshold: 0.35, MCSHR cut off: 90%, Number of divided parts: 7).

Test Case	Frame No	Part 1	Part 2	Part 3	Part 4	Part 5	Part 6	Part 7	Similarity (mean)	Matching Results
1	2040	0.2681	0.6239	0.8946	0.8581	0.7363	0.6118	0.9360	0.7041	1 (Yes)
	2150									
2	2042	0.7146	0.7568	0.9188	0.9344	0.9517	0.3098	0.6402	0.7466	1 (Yes)
	2152									
3	2044	0.7817	0.9093	0.9493	0.8878	0.9280	0.8268	0.8428	0.8751	1 (Yes)
	2154									
4	2046	0.8171	0.6491	0.8222	0.9308	0.8707	0.7943	0.7867	0.8101	1 (Yes)
	2156									
5	2048	0.6718	0.7190	0.8564	0.9283	0.9626	0.7997	0.3050	0.7490	1 (Yes)
	2158									
6	2050	0.8139	0.6633	0.6983	0.9550	0.8936	0.8258	0.7009	0.7930	1 (Yes)
	2160									
7	2052	0.7588	0.5820	0.8044	0.9618	0.9216	0.7687	0.8879	0.8122	1 (Yes)
	2162									
Post-Integration	2040-2052									100% (Yes)
	2150-2162									

6 Conclusions

In this paper, a matching-by-parts algorithm based on maximum likelihood criteria has been proposed. Based on our experimental results, the following conclusions can be drawn:

1) The proposed moving object matching-by-parts algorithm shows both good invariance and discrimination.
2) The assumptions made in the model in (3) are well validated by results reported in Tables 1 and 2. This allows formal derivation of the matching threshold, , λ.
3) Thanks to the post-matching integration, potential single-frame matching errors do not affect the overall matching result and robustness and accuracy are increased.

The proposed moving objects track matching-by-parts algorithms can significantly extend current video surveillance applications by providing them with the capability of tracking single objects across disjoint camera views which is the actual case for many real-world surveillance camera networks.

7 Acknowledgements

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

8 References

[1] M. Piccardi and E. D. Cheng, "Track Matching Over Disjoint Camera Views Based on an Incremental Major Color Spectrum Histogram", IEEE Int. Conf. on Advanced Video and Signal based Surveillance, Como, Italy. (AVSS 2005)

[2] 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.

[3] 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.

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

[5] O. Kaved, 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.

[6] 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.

[7] 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.

[8] 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.

[9] F. Roli, G. Fumera, "A theoretical and experimental analysis of linear combiners for multiple classifier systems, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 27, Issue 6, Jun 2005, Pages 942 - 956