Comparison of Techniques for Mitigating the Effects of Illumination Variations on the Appearance of Human Targets

C. Madden[1], M. Piccardi[1], and S. Zuffi[2]

[1] University of Technology, Sydney, Australia
[2] ITC-CNR, Milano, Italy

Abstract. Several techniques have been proposed to date to build colour invariants between camera views with varying illumination conditions. In this paper, we propose to improve colour invariance by using data-dependent techniques. To this aim, we compare the effectiveness of histogram stretching, illumination filtration, full histogram equalisation and controlled histogram equalisation in a video surveillance domain. All such techniques have limited computational requirements and are therefore suitable for real time implementation. Controlled histogram equalisation is a modified histogram equalisation operating under the influence of a control parameter [1]. Our empirical comparison looks at the ability of these techniques to make the global colour appearance of single human targets more matchable under illumination changes, whilst still discriminating between different people. Tests are conducted on the appearance of individuals from two camera views with greatly differing illumination conditions and invariance is evaluated through a similarity measure based upon colour histograms. In general, our results indicate that these techniques improve colour invariance; amongst them, full and controlled equalisation consistently showed the best performance.

1 Introduction

Applications in the computer vision field that extract information about humans interacting with their environment are built upon the exploitation of appearance, shape and motion cues in videos. Appearance (i.e. colour-based) features are increasingly being used because cheaper, higher resolution cameras of good pixel quality are available. However, significant problems still affect the reliable use of appearance features for the analysis of humans in videos, such as the variations in illumination and the articulated nature of humans' geometry. The goal of this paper is to improve the invariance of appearance features such as colour histograms for the global object. This is different from local colour invariants such as CSIFT [2] that describe the object's colours only in a limited spatial neighbourhood. The improvement of colour invariance is investigated through the comparison of data-dependent techniques that compensate for illumination changes. The evaluation of the illumination invariance of these techniques is based upon measuring their ability to remain invariant for a single person under

different illumination, whilst retaining a high degree of discrimination between different individuals.
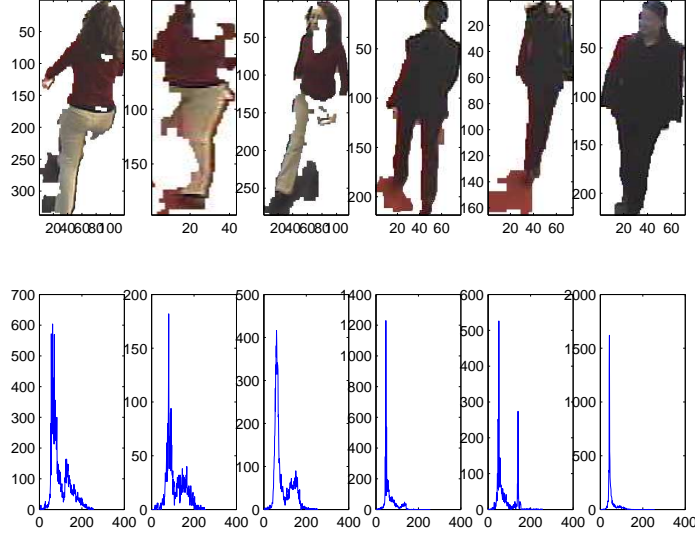
Fig. 1. Sample people of interest and their red histograms under differing illumination

The colour of an object in a camera view is not the intrinsic colour of the object itself, but rather a view-dependent measurement of the light reflected from the object, and the camera sensitivity to that light [3]. By recording the camera response to different wavelengths of light, the colour sensitivity can be estimated and exploited for model-based or empirical camera characterisation [4], [5]; however illumination provides a more difficult challenge. Compensating for illumination changes are broadly classified by Finlayson *et al.* [3] into colour invariants, which seeks transformation of the colours that are illumination independent, or colour constancy, which seeks to estimate the illumination of the scene to extract the intrinsic colours of objects. Whilst accurate models of the illumination of the scene could extract the intrinsic colours of objects, the implementation of this technique is very difficult. In previous work Javed *et al.* [6,7] propose to estimate the intensity transfer functions between camera pairs during an initial training phase. Such functions are estimated by displaying common targets to the two cameras under a significant range of illumination conditions, and modelling correspondences in the targets' colour histograms. However, the authors' assumptions in [6,7] that objects are planar, radiance is diffuse and illumination the same throughout the field of view do not hold in real life. Illumination varies at pixel-level resolution and have first-order effects on appearance. Weiss [8] proposed a method to estimate illumination from a sequence of frames of the same scene. Though the method works well for static objects such

as the background scene, it cannot accurately predict the illumination over 3D moving targets, especially highly articulated ones such as people. Moreover, in these applications segmentation is always affected by a certain degree of error and this adds to the effects of illumination variations and pose changes. Figure 1 shows examples of two such people of interest automatically segmented from the background, and how their red channel colour appearance may alter under differing illumination conditions.

Approaches to colour invariance have had greater success in mitigating the effects of illumination, which Finlayson *et al.* [3] suggest occur because although the RGB values change, the rank ordering of the responses of each sensor is preserved. This implies that the values for a particular colour channel, such as R, will change from illumination source A to source B; however the ordering of those values will remain invariant, as shown in Figure 1. This observation has occurred for what we assume to be typical lighting in human environments, which largely consists of natural sunlight, fluorescent lighting, or incandescent lighting. Other lighting sources are sometimes used, but are rarely used in open common spaces where surveillance occurs, so they are outside the scope of this investigation.

A range of techniques are used to provide colours that are invariant to illumination, with the most common being chromaticity spaces. Chromaticity can be simply derived from the RGB space using the following transformation:

$$r = \frac{R}{R+G+B}, g = \frac{G}{R+G+B}, b = \frac{B}{R+G+B} \qquad (1)$$

This chromaticity vector (r,g,b) has only two independent co-ordinates and is defined such that it is invariant to the intensity of an illumination source. Changes to the illumination will scale the RGB values by a factor s as (sR,sG,sB), leaving r,g,b invariant. If the illumination source changes in spectral output, say from a white fluorescent source to a yellow incandescent source, then a single scale factor is not sufficient to compensate for such a change. A second diagonal space has also been proposed where each sensor response in the R, G, or B channels can be independently derived. This model allows for a shift in illumination intensity as well as a possible shift in the frequency spectrum for that illumination. The response could be modeled using the grey-world representation [9] by using:

$$R^{'} = \frac{R}{R_{ave}}, G^{'} = \frac{G}{G_{ave}}, B^{'} = \frac{B}{B_{ave}} \qquad (2)$$

where R_{ave}, G_{ave}, and B_{ave} denote the means of all the R, G, B values respectively across an entire image.

These common techniques are useful for providing measurements that are invariant to illumination to a degree; however they have difficulty in adequately compensating for the multiple illumination sources that could also be time varying in the case of natural sunlight. These multiple illumination sources also have complicated interplay with the complex 3D surfaces of moving objects, where the effect of illumination in the background, or portions of the background may vary

from its effect upon foreground objects. These chromaticity techniques cannot identify the difference in intrinsic black and white surfaces or differing shades of grey. Moreover, the model in (1) is unstable for dark colours. A different approach to colour invariance considers techniques that perform colour normalisation in a data-dependent fashion, such as histogram equalisation. These techniques do not seek to model the illumination conditions and could therefore be more suitable for uncontrolled scenarios that are predominant in video surveillance. We propose to compare various techniques that transform the RGB data of the object to make the same object more similar under varying illumination conditions, whilst still allowing for the discrimination of differing colours without requiring either training or other assumed scene knowledge.

The four techniques compared are filtering the illumination [10], Section 2, applying histogram stretching to the object histograms, described in 3, and applying equalisation to the object histograms in both a full mode, and a novel controlled equalisation technique [3, 1], Section 4. These techniques have limited computational requirements and are therefore immediately suitable for real time implementation. These methods are compared by applying the techniques at varying parameter levels to a set of 15 tracks acquired from four different individuals within two cameras under differing illumination conditions. After the mitigation techniques have been applied, colour histograms of the object are extracted and compared as described in Section 5. The histograms used are computed in the joint RGB space so as to retain the correlation between colour components. Since the joint RGB space has many possible different values, colours are mapped onto sparse histograms (Major Colour Representations, or MCR's) i.e. histograms retaining only those bins having a non negligible bin count [1]. Then, the similarity of any two histograms is measured based on the Kolmogorov divergence with equal priors [11]. This allows for a comparison of the effects of the mitigation through the comparison of the similarities of the MCR histograms. The results of the similarity measurements of the appearance between 120 track pairs from matching and non-matching individuals are compared in Section 6. This provides a discussion of the ability of these techniques to improve the invariance of the appearance under different illumination, whilst still retaining discrimination over different individuals.

2   Illumination Filtration

This section outlines a technique of homomorphic filtering of the illumination effects from the image based upon the method described by Toth *et al.* [10]. This technique assumes objects consist of Lambertian surfaces and that illumination only changes slowly over space in the image. Toth *et al.* [10] suggests that this low frequency component can be filtered out by converting values to a logarithmic scale then a applying high pass filter, leaving the mid to high frequency details which in practise relate to the reflectance component of the image.

The intensity of the illumination on the surface of the object in the τ-th frame in an image sequence can be modelled as:

$$y_\tau(k) = i_\tau(k) \cdot r_\tau(k) \tag{3}$$

where k is the pixel index in the image, i is the illumination component and r is the reflective component in the image y. If the reflectance component r can be separated from the illumination component i, then it can be used as an illumination invariant representation of the appearance. The slow rate of change of illumination over the space of the image means that it will consist of low frequency components of the image, whilst the reflectance model will consist significantly of mid to high frequency components. Applying the logarithm to **(3)** transforms the multiplicative relationship between y, i, and r into an additive one:

$$log(y_\tau(k)) = log(i_\tau(k)) + log(r_\tau(k)) \tag{4}$$

A high pass filter kernel can then applied to remove the low frequency illumination component i. An exponentiation of the filtered image therefore contains the illumination invariant image consisting of the reflectance information. The parameters of the Gaussian filter applied to remove the illumination relate to the filter size, standard deviation, and a weighting parameter which controls the amount of filtration applied. These parameters are given in this order when the filtration results are presented in Table 1 in Section 6.
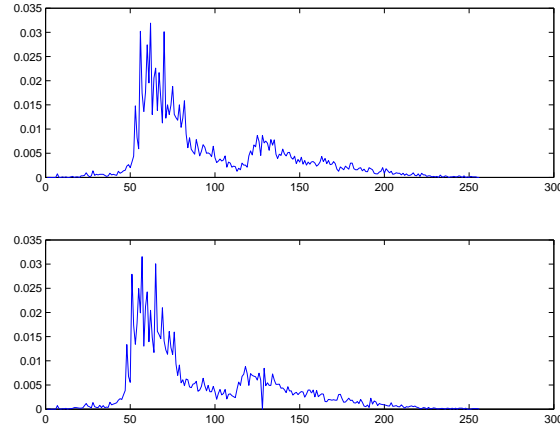

Fig. 2. Individuals R values before and after illumination filtration

3   Histogram Stretching

This section outlines the use of histogram stretching to perform the illumination transformation. This method proposes to stretch the objects histogram sepa-

rately for each of the RGB components to allow for changes in the illumination spectrum. Stretching the histogram should make it appear more similar across a range of illuminations conditions without explicitly modelling those illumination sources. It also preserves the rank ordering of the histogram, which Finlayson *et al.* [3] suggest adds to the success of many of the colour invariance techniques. This technique is demonstrated in Figure 3.
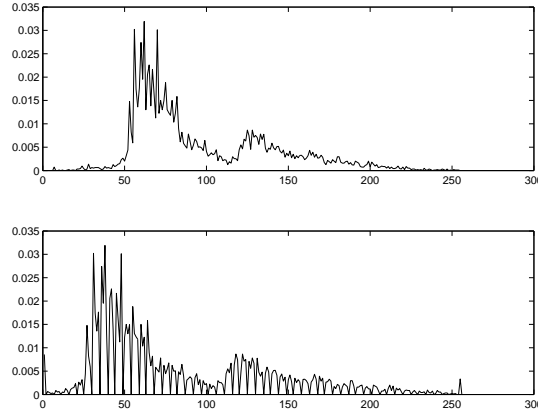
Fig. 3. Histogram Stretching of the Individuals pixels

The key points for histogram stretching are the selection of the upper and lower limits of the output histogram, and the upper and lower limits of the input histogram for each colour channel. Histogram stretching then performs a linear mapping of the input to output values. We maximise the spread of the histogram by choosing the upper and lower limits of the stretched output to be 255 and 0 respectively. We choose the upper and lower limits of the object histogram based upon a single parameter a which denotes the percentage amount of histogram tails to be ignored. The removal of these tail components of the histogram aims to reduce the amount of noise in the input histogram. It is calculated by cumulating the count in each histogram bin from either end until the percentage a is reached.

If we denote the lower input limit as b and the upper input limit as c, then the output of the stretching r' for any given input value in that channel can be calculated as:

$$r' = (r - b) \left(\frac{255}{c - b} \right) \tag{5}$$

This stretching transformation is performed upon each object pixel to generate a new object image which should have a higher tolerance to illumination changes without requiring either training or other assumed scene knowledge. This stretching provides a linear transformation of values so they lie across the

entire histogram, whilst still retaining a similar shape to the original object component. The results of the stretching is presented in Table 1 in Section 6 for a range of a values to explore the effect of changing the amount of the histogram that is ignored.

4   Histogram Equalisation

This section outlines the use of equalisation to perform a data-dependent transformation of an individual's histogram. This method differs from histogram stretching as it can provide a non-linear transformation. First this section explains the application of histogram equalisation as proposed by Finlayson *et al.* [3], before defining the 'controlled equalisation' as we proposed in [1].

Histogram equalisation, also denoted here as full equalisation, aims to spread a given histogram across the entire bandwidth in order to equalise as far as possible the histogram values in the frequency domain. This operation is data-dependent and inherently non-linear as shown in Figure 4; however it retains the rank order of the colours within the histogram. The equalisation process is applied separately in each of the R, G and B colour components to remap their values according to the following transformation functions:
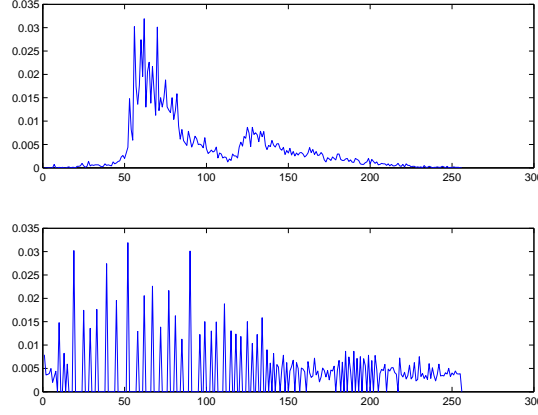
$$T_r\left(i\right) = \frac{255}{N} \sum_{j=0}^{i} p_r\left(j\right) \tag{6}$$

$$T_g\left(i\right) = \frac{255}{N} \sum_{j=0}^{i} p_g\left(j\right) \tag{7}$$

$$T_b\left(i\right) = \frac{255}{N} \sum_{j=0}^{i} p_b\left(j\right) \tag{8}$$

We also introduce a 'controlled equalisation' as described in [1]. This process is based upon equalising a combination of the object pixels and an amount of pre-equalised pixels that is a proportion k of the object size. These pre-equalised pixels effectively 'control' the amount of equalisation such that the pixels are spread to a limited degree within the spectrum instead of being spread fully. Thus although an object should become more matchable under a range of illumination conditions, it is still likely to retain a higher degree of discrimination from objects of differing intrinsic colour. This technique is demonstrated at varying levels of parameter in Figure 5 below.

This equalisation can be formally described by designating the set of N pixels in a generic object as A, and calling B a second set of kN pixels which are perfectly equalised in their R, G, and B components. Note that the parameter k designates the proportionality of the amount of equalised pixels to the amount of pixels in A. From their union $A \cup B$, the cumulative histograms of the R, G, and B components, $p_r\left(i\right), p_g\left(i\right)$, and $p_r\left(i\right)$ for $i = 0 \dots 255$ are computed. A histogram equalisation of the individual colour channels is then derived as shown in 9-11:

Fig. 4. Full Equalisation of the Individuals pixels

$$T_r\left(i\right) = \frac{255}{\left(1+k\right)N} \sum_{j=0}^{i} p_r\left(j\right) \tag{9}$$

$$T_g\left(i\right) = \frac{255}{\left(1+k\right)N} \sum_{j=0}^{i} p_g\left(j\right) \tag{10}$$

$$T_b\left(i\right) = \frac{255}{\left(1+k\right)N} \sum_{j=0}^{i} p_b\left(j\right) \tag{11}$$

These intensity transforms can then be applied to re-map the $R, G,$ and B components in the object's pixels providing the 'controlled equalisation'. The parameter k can be used to control the amount of pre-equalised pixels used, which in turn controls the spread of the object histogram. The results of this technique presented in Table 1 in Section 6 are for a range of k values to explore the effect of changing this parameter.

5 Comparison of Techniques which Mitigate Changes in Illumination

This section outlines the experiment used to compare the techniques that mitigate the effects of changes in illumination upon object appearance. The goal of the experiment is to evaluate the effectiveness of the various techniques by measuring similarities between colour histograms computed over objects extracted from video surveillance videos. To this aim, we used 15 tracks obtained from 4 individuals across 2 cameras with illumination from both natural sunlight and artificial sources. The experiment articulates over various stages of processing whose details are provided in the following.
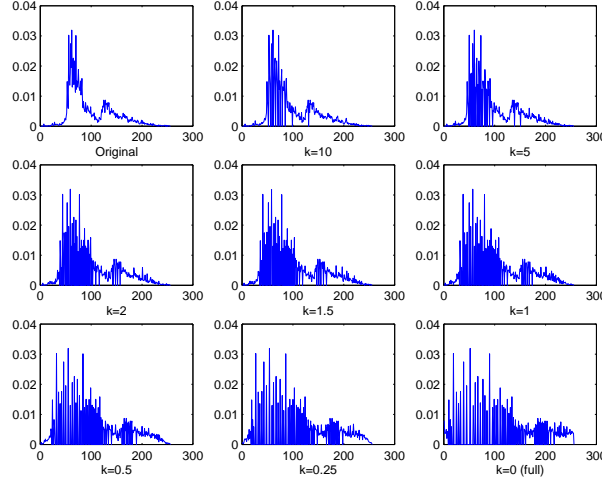
Fig. 5. Controlled Equalisation of the individuals pixels with varying k values

The first stage of processing is to automatically extract the objects from the background in each frame of the videos. We have utilised an adaptive mixture model based upon that derived by Wren *et al.* [12] that quickly provides reasonably segmented objects. All the objects extracted along the frame sequence from a single individual are then manually collected into a single track so as to ensure correct data association.

In the second stage, for each object and in each frame, one of the mitigation techniques is applied in turn and the values of the object's pixels remapped accordingly. In the third stage, the MCR histogram of the object's appearance is computed as described in [1]. An MCR histogram consists of a sparse set of bins mapping the pixels' colour values. Each bin in the RGB space is of spherical shape and has the same radius under a normalised colour distance. The number of such bins is not bounded a priori and the position of their centroids is optimised through a k-means procedure. The MCR histogram is a 3-D, non-parametric representation of an object's colours.

In the forth stage of processing, tracks are considered in pairs. One frame from each track is taken and a similarity measurement, S_f, is computed between their two MCR's based on the Kolmogorov divergence. In a similar way, S_f values are computed for all other possible frame combinations from the two tracks and averaged so as to achieve a similarity measurement at the track level, S_t. A number of track pairs are built for both the cases of two different people (non-match case, or $H0$ hypothesis) or a single person (match case, or $H1$ hypothesis) and all S_t computed and stored for the two cases.

In the fifth stage, the distributions of the S_t values for each of the two hypotheses, $H0$ and $H1$, are statistically modelled by optimally fitting a Gaussian distribution on each. In this way, the two distributions, $p_{H0}(S_t)$ and $p_{H1}(S_t)$,

are simply described by their expected values, μ_{H0} and μ_{H1}, and their standard deviations, σ_{H0} and σ_{H1}. The Gaussian assumption seems to well model the data, with σ_{H0} significantly larger than σ_{H1} (the dispersion of similarity values for different objects is obviously greater than that for different views of a same object). The performance evaluation for the different mitigation techniques is then performed by computing the false alarm rate and the missed detection rate directly from $p(H0)$ and $p(H1)$, assuming $H0$ and $H1$ have equal priors. We derive the similarity value, $S_{t_{th}}$, for which $p_{H0}(S_t) = p_{H1}(S_t)$ as:

$$S_{t_{th}} = \frac{-b - \sqrt{b^2 - 4ac}}{2a} \tag{12}$$

with

$a = \sigma_1^2 - \sigma_2^2$
$b = 2(\mu_1 \sigma_2^2 - \mu_2 \sigma_1^2)$
$c = \sigma_1^2 \mu_2^2 - \sigma_2^2 \mu 1^2 + 2\sigma_1^2 \sigma_2^2 \ln(\frac{\sigma_2}{\sigma_1})$

The false alarm rate, PFA, is then given by the tail $p_{H0}(S_t)$ below $p_{H1}(S_t)$, $(S_t \geq S_{t_{th}})$ and the missed detection rate, PMD, from the tail of $p_{H1}(S_t)$ below $p_{H0}(S_t)$, $(S_t \leq S_{t_{th}})$. By identifying the matching errors from the estimated statistical distributions, the desirable property for the most effective technique is that it will provide the best possible trade-off between false alarm and missed detection rates. The results of the effectiveness of the various illumination mitigation techniques are reported in Table 1.

6   Results

This section discusses the results from the comparison of these data-dependent, rank-preserving techniques which mitigate the effect of illumination changes upon appearance. They are observed for a range of their parameters, including the case of no attempt at mitigation (i.e. leaving the colour values unaltered). These results are based upon the similarity values obtained from 50 matching and 70 non-matching track pairs and are reported in Table 1. The results investigate the effectiveness of the compared mitigation techniques by comparing their estimated PFA, PMD and total error rate.

The results in Table 1 clearly demonstrate that applying histogram stretching actually seems to reduce effectiveness rather than improve upon the case of no mitigation attempt (first row). This occurs through both an undesirable reduction in the similarity of matching objects and an increase in the similarity of differing objects. The application of illumination filtration is suggested to remove highlights upon objects, as well as compensating for general illumination. The results show that varying the filter parameters to increase the size and variation of the Gaussian filter produces some improvement in both the matching of similar colours and that of differing colours with respect to no mitigation attempt. The results for the colour equalisation techniques show the best improvement in matching scores, indicating that as Finlayson et al. suggest [3],

Table 1. Results of similarity measurements for matching and non-matching tracks

Method	Parameter	Matching		Non-Matching		Theoretical Err ($10^{-6}\%$)		
		mean	std	mean	std	PMD	PFA	total
None		0.8498	0.0103	0.2230	0.1202	0.073	0.938	1.011
Equal	Full	0.9095	0.0020	0.2522	0.1437	0.041	3.426	3.467
Equal	0.5	0.9080	0.0018	0.2637	0.1552	0.020	2.114	2.134
Equal	1	0.9116	0.0012	0.2680	0.1599	0.020	3.270	3.290
Equal	2	0.9135	0.0014	0.2726	0.1648	0.042	6.032	6.073
Stretch	0.1%	0.7648	0.0133	0.2070	0.0980	0.043	0.337	0.380
Stretch	1%	0.7452	0.0146	0.1967	0.0891	0.013	0.081	0.093
Stretch	5%	0.7279	0.0145	0.1738	0.0714	0.00001393	0.00007115	0.00008508
Filter	5 1 0.5	0.8496	0.0086	0.2294	0.1243	0.112	1.798	1.910
Filter	7 2 0.4	0.8511	0.0074	0.2262	0.1214	0.039	0.712	0.751
Filter	7 3 0.5	0.8592	0.0072	0.2322	0.1271	0.090	1.765	1.855
Filter	7 2 0.6	0.8632	0.0069	0.2377	0.1326	0.195	4.237	4.432

they are capable of providing colours that are more illumination invariant. Compared to stretching, equalisation reallocates the histogram bins to compress the dynamic of bins with low bin counts and expand that of bins with high bin counts. This non-linear reallocation seems to better compensate for the appearance changes occurring under illumination variations. Whilst full equalisation produces the best overall error rate, this is only marginally lower than that of controlled equalisation, which, at its turn, produces the best similarity between matching objects. Picking the "best" technique between these two would require one to define costs for both a false alarm and a missed detection. Such costs significantly depend on the actual application.

7   Conclusions

Many techniques have been suggested in the literature to compensate for the effects of variable illumination over an object's appearance. In a scenario of video surveillance, explicitly estimating the illumination over 3-D deformable moving targets such as humans simply proves impractical. For this reason, in this paper we have discussed and compared various data-dependent, rank-preserving techniques in an attempt at improving the invariance of a person's appearance across camera views without exploiting any scene knowledge. Results show that some of these techniques can significantly mitigate the effects of illumination variations, almost halving the matching error rate. Therefore, their use seems strongly beneficial for these applications. The histogram stretching technique actually diminish the similarity of matching objects and increase that of differing objects and its use is therefore counter productive. The illumination filtration technique alone provides a marginal improvement in the similarity of an object's appearance under illumination changes, possibly due to its removal of illumination highlights on the object. The equalisation of an individual's colour histograms provides a significant improvement in appearance similarity under

differing illumination. Whilst full equalisation produces the best overall error rate, controlled equalisation produces the best similarity between matching objects. Either technique may suit different surveillance applications depending on their error costs. For instance, as discussed in [1], tracking people across a network of disjoint camera views require the highest possible detection rate in order to avoid costly manual revisions; however false detections are easier to correct. This suggests the use of controlled equalisation as the technique of choice for this scenario.

8   Acknowledgements

References

1. Madden, C., Cheng, E.D., Piccardi, M.: Tracking people across disjoint camera views by an illumination-tolerant appearance representation. Machine Vision and Applications **18** (2007) 233–247
2. Abdel-Hakim, A.E., Farag, A.A.: Csift: A sift descriptor with color invariant characteristics. International Conference on Computer Vision and Pattern Recognition **2** (2006) 1978–1983
3. Finlayson, G., Hordley, S., Schaefer, G., Tian, G.Y.: Illuminant and device invariant colour using histogram equalisation. Pattern Recognition **38** (2005) 179–190
4. Barnard, K., Funt, B.: Camera characterization for color research. Color Research and Application **27** (2002) 153–164
5. Bala, R.: Device characterization. In Sharma, G., ed.: Digital Color Imaging Handbook. CRC Press (2003)
6. Javed, O., Rasheed, Z., Shafique, K., Shah, M.: Tracking across multiple cameras with disjoint views. IEEE Conference on Computer Vision and Pattern Recognition **2** (2005) 26–33
7. Javed, O., Rasheed, Z., Shafique, K., Shah, M.: Tracking across multiple cameras with disjoint views. International Conference on Computer Vision **2** (2003) 952–957
8. Weiss, Y.: Deriving intrinsic images from image sequences. International Conference on Computer Vision **2** (2001) 68–75
9. Barnard, K., Funt, B., Cardei, V.: A comparison of computational colour constancy algorithms; part one: Methodology and experiments with synthesized data. IEEE Transactions in Image Processing **11** (2002) 972–984
10. Toth, D., Aach, T., Metzler, V.: Bayesian spatiotemporal motion detection under varying illumination. European Signal Processing Conference (2000) 2081–2084
11. Zhou, S.K., Chellapa, R.: From sample similarity to ensemble similarity: probabilistic distance measures in reproducing kernel hilbert space. IEEE Transactions on Pattern Analysis And Machine Intelligence **28** (2006) 917 – 929
12. Wren, C., Azarbayejani, A., Darrell, T., Pentland, A.P.: Pfinder: real-time tracking of the human body. IEEE Transactions on Pattern Analysis And Machine Intelligence **19** (1997) 780–785