Colour Transfer by Feature Based Histogram Registration

C. R. Senanayake and D. C. Alexander
Department of Computer Science, UCL
Gower Street, London, WC1E 6BT
{c.senanayake|d.alexander}@cs.ucl.ac.uk

Abstract

This paper addresses the problem of automatically eliminating unwanted colour variation between similar image pairs. We propose a feature-based registration method to align colour histograms without any spatial processing of the image content or assumptions about the scene contents. We use the method for the colour transfer problem. The features are histogram maxima that persist through the scale space, so they are robust to large changes in the sizes of objects in the scene. The algorithm seeks the best matches between features and aligns the histograms via a polynomial warp. We construct a set of image pairs that exhibit variation in object scale and lighting and use it to show that the method produces better colour space alignments than simple alignments of histogram moments.

1   Introduction

A common problem in computer vision is that different sensors acquire different colour responses to an imaged object. This problem occurs because physical factors during the imaging process introduce a variation that differs for each sensor; in addition, it is practically impossible to image an object under perfectly constant lighting conditions at different spatial positions within an imaging environment [3][4][10]. This variation degrades the performance of colour computer vision processes such as object tracking [4]; in addition, the involved nature of calibration routines means that the calibration step is often ignored.

The colour transfer approach offers the potential for automatic alignment of similar colour spaces without manual intervention to perform the calibration. The goal of colour transfer methods is to make the image regions for the same object the same colour; in this paper, we seek to achieve this by aligning the corresponding dense regions of the histograms (clusters). We are only concerned with the situation when the images contain the same or at least highly similar objects. Existing colour transfer methods assume that scale changes between corresponding clusters are small; however, a significant change in object scale is common when the camera moves or when tracking between multiple cameras. We seek to develop a scale invariant cluster alignment algorithm and we pose colour transfer as a two-frame registration problem where our goal is to align the corresponding clusters in two colour histograms in the presence of lighting, automatic camera setting and object scale variation.

Registering colour histogram data presents unique challenges; corresponding clusters may change in shape and size, tear apart or merge in complex and uncharacterised ways under changes in lighting, camera variation and pose. The main contribution of this paper is to introduce a new method for automatically aligning histograms, that is robust to changes of scale of scene objects. The method matches scale invariant histogram features and aligns them using polynomial warps of the colour space.

2 Background

The term colour transfer was introduced to describe the process of aligning the colour histograms of a source and target image [8]. We distinguish colour transfer from the similar problems of colour consistency and colour constancy as follows: Colour consistency methods align colour histograms using correspondences established between objects of known reflectance such as calibration charts. Colour constancy methods estimate the light source and transform the colour histogram of an individual image to a standard reference frame using physical and/or statistical assumptions about the scene. Colour transfer on the other hand does not require objects of known reflectance in the scene and it computes an alignment using the histograms from both images.

The need to correct for sensor response and lighting variation occurs in a wide variety of settings such as 3D reconstruction [3], robotics [10] and special effects processing [8]. At the highest level, colour calibration methods divide into within camera and between camera calibrations. Within camera calibrations remove response variations that occur between different pixels in the same camera due to physical effects such as vignetting. These approaches involve prior physical models and manual calibration steps. Between camera calibrations align the response curves of different cameras. Five broad classes of between camera calibration methods exist in the literature, these are: chart based calibrations [3], calibrations using spatial overlap [9], disjoint view calibrations using non-parametric methods [7], disjoint view calibrations using parametric models [4] and calibrations using training images [10].

Polynomial transformations have been shown [3] highly effective in aligning colour histograms. However, no existing method can automatically establish the necessary correspondences to apply these transformations without colour calibration charts in the scenes. Our method overcomes this problem through a scale-invariant feature detection step that operates on the colour histograms. In addition, our approach involves no knowledge of the layout of objects in the scene and makes no parametric assumptions about the form of the colour data or the number of clusters that are present. FBHR has been developed using the RGB colour space. However, it contains no specific assumptions about the RGB space. We take this approach so the method can be applied to other colour spaces and applications in the future.

3 Method

The method we propose registers a source RGB histogram with a target RGB histogram. Registering the colour histograms of images computes a colour transformation that aligns the colours of a source image with those in a target image. Here, we only consider the case where the two images contain the same set of N single-coloured objects. Each colour
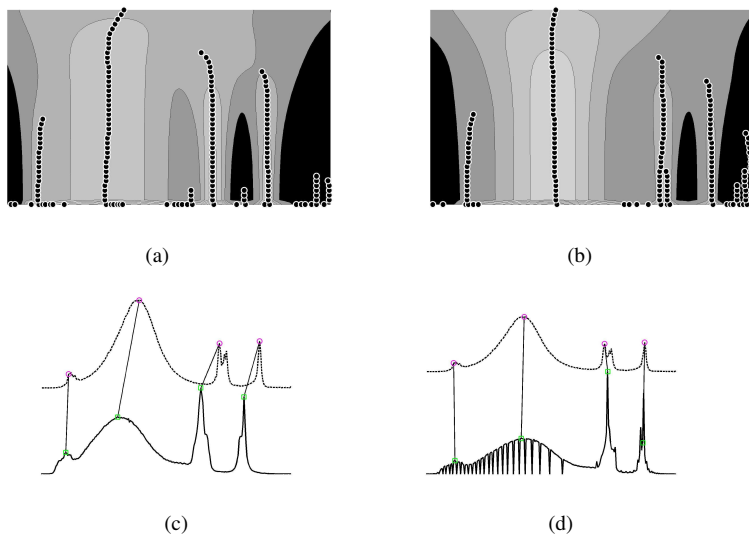
(a)

(b)

(c)

(d)

Figure 1: Example steps in the FBHR algorithm in the green channel when processing the BLSS image pair. 1(a) shows the scale space maxima for the green channel histogram of the source image (2(f)), 1(b) shows the corresponding scale space maxima for the green channel histogram of the target image (2(h)). (Circular dots indicate maxima, scale increases vertically and the intensity scale shows histogram density.) 1(c) shows a view of the detected features and the correspondences between the source histogram (solid line on bottom) and the target histogram (dashed line on top). 1(d) shows the warped source histogram and the target.

Histograms are plotted as lines joining the tops of the histogram bins, the transformed histogram is no longer smooth as it contains zero bins after being stretched by the multiplicative component of the transform.

histogram contains a number of dense regions that correspond to objects of interest. We seek the transform that aligns the clusters that correspond to the same object. The method is designed to handle multiple clusters of potentially different size, no explicit assumption is made about the shape of the distributions or the number of clusters present. We only assume that the source and target images are of the same set of objects.

3.1 Aligning Histogram Moments

Aligning histogram moments is an obvious candidate method for aligning colour distributions. Here, we outline three different moment transformations between a source histogram $\mathbf{S}$ and a target histogram $\mathbf{T}$. We compute an independent transformation in each dimension, we write the distribution in the ith channel as $\mathbf{S}_i$ and for the target distribution as $\mathbf{T}_i$: **1)** Additive alignment of the 1st moment: a scalar shift is computed for each channel $w_i = E(\mathbf{T}_i) - E(\mathbf{S}_i)$, where E is the mean operator; each scalar source data point s_i is transformed to $w_i + s_i$. **2)** Multiplicative alignment of the 1st moment: a scalar gain is computed from each channel $r_i = E(\mathbf{T}_i)/E(\mathbf{S}_i)$ and each source data point

is transformed to $r_i s_i$. **3)** Linear alignment of the 1st and second moment: the first and second moments can be aligned by a gain γ_i and a shift ω_i in each channel. γ_i is computed from the source histogram variance $V(\mathbf{S}_i)$ and target histogram variance $V(\mathbf{T}_i)$ such that $\gamma_i = V(\mathbf{T}_i)/V(\mathbf{S}_i)$. ω_i is computed as $\omega_i = E(\mathbf{T}_i) - \gamma_i E(\mathbf{S}_i)$, each source data point s_i is transformed to $\omega_i + \gamma_i s_i$.

3.2 Feature based histogram registration algorithm

In this section we outline the feature based histogram registration algorithm (FBHR). The steps to compute a colour transformation from a source image to a target image using FBHR are: **1)** Compute one dimensional histograms in each channel for both the source and target images. **2)** Compute the scale space of each histogram and extract salient features. **3)** Reject obvious outlying features. **4)** Match the remaining features. **5)** Compute the coefficients of polynomial transformations to align matching features in each channel. **6)** Transform the source image. **7)** Test for failure in each channel by comparing the transformed and target histogram. **8)** If FBHR fails, revert to a moment based transformation.

The following subsections elaborate on these steps and explain the rationale for taking this approach.

3.2.1 Scale Space features

We aim to match significant maxima in the colour histograms. Colour histogram data is commonly noisy and we avoid false feature generation using scale space smoothing. We track maxima that persist over a scale space. Analysis of the scale space in this way has been termed deep structure analysis [5] . The scale space of a function h can be computed by convolving h with Gaussian kernels $G(.;\sigma)$ of various widths σ. We detect all local maxima at each scale and remove the maxima that persist over less than T scales. We search from low levels of σ upwards. A feature persists between adjacent scales if its position is in adjacent histogram bins at the two scales; feature persistence is affected by histogram bin width and the values of σ. Figure 1 illustrates these features for one example image pair and their subsequent processing.

3.2.2 Rejecting outlier features and Matching

Oversaturation is a very common feature in digital images and leads to spikes at the ends of the histograms in question; we reject these features by thresholding out features less than δ away from the end of the range. Maxima found in histogram bins below a noise floor level γ are rejected as they are unlikely to be reliable.

The matching step seeks a match between the detected source and target histogram feature points. We do this with the following steps: **1)** Generate all matches between the source and target points for this number of matches. If there are the same number of source and target points we attempt to match them all, if either set is smaller we seek to match all the points in the smaller set. **2)** Reject matches that do not preserve rank ordering. In 1D this means that for a set of matched points, both points in each match must be either less than or greater than all other matched pairs. This prevents folding in the transformation. **3)** Pick the match with the minimum global Euclidean cost between points.

3.2.3 Point alignment transforms

The matching step outputs M_i corresponding points (s_{ij}, t_{ij}), $j = 1, ..., M_i$, in each channel $i = 1, ..., N$. We seek the transformation f_i in each channel that minimizes

$$\sum_{j=1}^{M_i} \left| f_i(s_{ij}) - t_{ij} \right|, \tag{1}$$

where $|.|$ is the L2 norm. We use an order d polynomial mapping for f_i so that

$$f_i(s_{ij}) = \alpha_{0i} + \sum_{k=1}^{d} \alpha_{ki} s_{ij}{}^k. \tag{2}$$

Each corresponding pair provides a linear constraint on the α_{ki}. We write them all as a matrix equation

$$\begin{bmatrix} 1 & s_{1,n} & s_{1,n}^2 & \cdots & s_{1,n}^d \\ 1 & \vdots & \vdots & \vdots & \vdots \\ 1 & s_{M,n} & s_{M,n}^2 & \cdots & s_{M,n}^d \end{bmatrix} \begin{bmatrix} \alpha_{0,n} \\ \vdots \\ \alpha_{d,n} \end{bmatrix} = \begin{bmatrix} t_{0,n} \\ \vdots \\ t_{d,n} \end{bmatrix}. \tag{3}$$

Writing this as $\mathbf{A}_i \mathbf{C}_i = \mathbf{D}_i$, we solve for $\mathbf{C}_i$: $\mathbf{C}_i = \mathbf{A}_i^\dagger \mathbf{D}_i$, where $\mathbf{A}_i^\dagger$ is the pseudo-inverse of $\mathbf{A}_i$. This is repeated for each channel.

3.2.4 Failure test

The added flexibility of the polynomial transformation makes them sensitive to mismatches and means that histograms can be poorly aligned when the correspondence step has not returned the correct result. We perform a coarse level of failure testing between the transformed histogram and target histogram in each colour channel. However, when the procedure fails, it tends to fail catastrophically, which is easy to detect. We compute the histograms for the transformed and target data less than δ away from the end of the range, this rejects over saturated regions which cause outlier effects. Then we compute the Bhattachayya coefficient $B_i = \sum_{b=1}^{P} \sqrt{r_b t_b}$ in each channel, where b indexes the transformed histogram bins r_b and the target histogram bins t_b. If B_i is less than a failure threshold F then FBHR has failed in this channel.

4 Experiments and Results

The experiments test the hypothesis that the feature based histogram registration method in section 3.2 transforms the colours of the source image histogram so that they are better aligned with the target histogram than when using alignment of either the 1st, or 1st and 2nd histogram moments (as in section 3.1). We investigate the failure points of the algorithm and simple strategies for mitigating these failures. The test images divide into three sets: 1) images of three distinctly coloured pieces of paper on a desk, 2) images
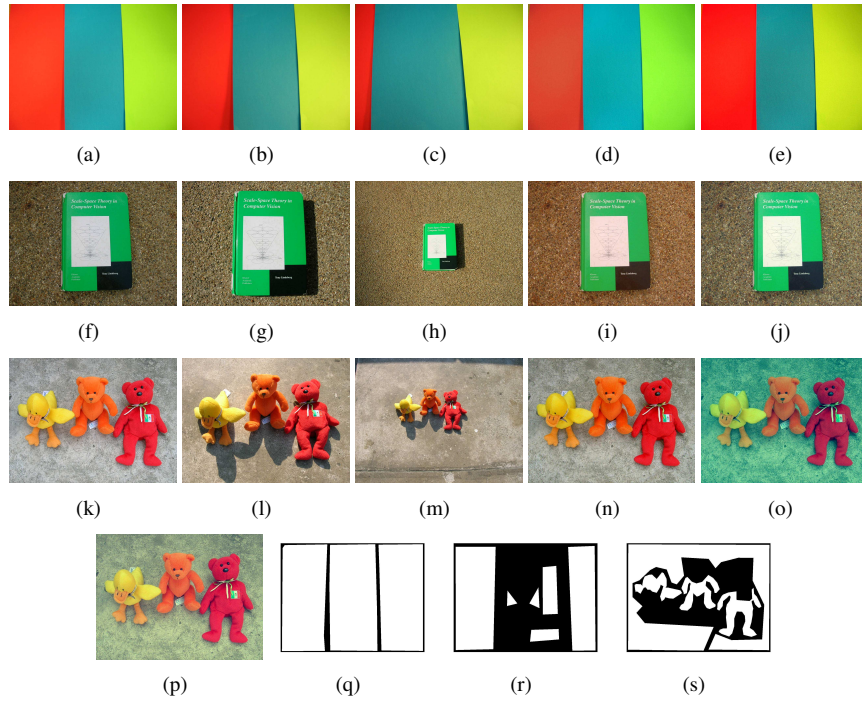
Figure 2: Source and target images from the three datasets, examples of transformed source images and masked regions. Image 2(a) is of paper strips under florescent lighting, 2(b) is taken under similar conditions but with a desk lamp turned on to cause a slight change in colours. Image 2(c) shows rearranged strips with the desk lamp on. Transformed source images are shown for a 1st moment shift 2(d) and FBHR1 2(e) when using source image 2(a) and target image 2(c) (SLSS). Image 2(f) shows the book at large scale in shadow and 2(g) shows it in sunlight, 2(h) shows the book at small scale in sunlight. Transformed images are shown for a 1st moment shift 2(i) and FBHR1 2(j) using source image 2(f) and target image 2(h) (BLSS). Image 2(k) shows the stuffed animals at large scale in the shade, 2(l) shows them in direct sunlight and 2(m) shows them at small scale and in sunlight. Transformed images are shown for a 1st moment 2(n), FBHR1 2(o) and FBHR1 shift fail 2(p) when using source image 2(k) and target image 2(l) (AL). Image 2(q), shows masked regions for 2(a), mask 2(r) shows the regions for 2(f) and 2(s) shows the mask for 2(k); white polygons indicate the regions of interest, black regions are ignored in the mask. For each mask, polygons corresponding to different colour regions of interest can be identified separately.

Table 1: Descriptions of image pairs used and short hand notation to refer to them. (Where possible, we reference the source (Src) and target (Tgt) images in Figure 2.

Code	Image pair description
SL	Paper strips under lighting change at same scale. (Src: 2(a), Tgt: 2(b))
SLS	Paper strips under lighting change and moderate scale change.
SLSS	Paper strips under light change and large scale change. (Src: 2(a), Tgt: 2(c))
BL	Book under lighting change at same scale. (Src: 2(f), Tgt: 2(g))
BS	Book under the same lighting with moderate scale change.
BLS	Book under lighting change and moderate scale change.
BLSS	Book under lighting change and large scale change. (Src: 2(f), Tgt: 2(h))
BLSSS	Book under lighting change and very large scale change
AL	Stuffed animals under a lighting change (Src: 2(k), Tgt: 2(l))
AL2	Stuffed animals occupying small percentage of image under a lighting change
ALSL	Stuffed animals under a lighting change and scale change

of a book and 3) images of stuffed toys. The image sets are constructed to pick source and target images that exhibit either scale, lighting or both scale and lighting changes. In the images of stuffed animals, shadows are a significant source of variation that make the alignment task more difficult. Figure 2 illustrates the types of images present in the image sets and the associated hand marked masks used in the evaluation of the colour transfer results. Table 1 describes the source and target image pairs and introduces a shorthand used to refer to image pairs from this point onwards.

We compare the alignment of moments to the FBHR method using polynomials of order 1, 2 and 3. The scale space was generated over T scales with $\sigma_n = e^{0.1(n-1)}$, $n = 1, \cdots, T$. The exponential increase in kernel scale ensures that blurring occurs more slowly at lower scales which prevents premature breakages in the feature paths that we are tracking. In all cases, we set the over-saturation threshold δ to 3, the noise-floor threshold γ to 0.001, T to 15 and the Bhattachayya failure threshold F to 0.7. The histogram bin width is 1 in all cases.

After alignment, we compute a metric that compares the colours of L corresponding regions marked by hand polygons in all images. The paper strip set is marked up into three regions, the book and stuffed animals each have four marked up regions; examples of marked up regions are given in figures 2(q), 2(r) and 2(s). The metric is the total average Mahalanobis distance between the corresponding regions in the transformed and target images. We denote the RGB pixels extracted from the nth masked region of the transformed image as a 3 column matrix $\mathbf{Q}_n$ and from the target image as a 3 column matrix $\mathbf{W}_n$ where each row in both $\mathbf{Q}_n$ and $\mathbf{W}_n$ corresponds to a RGB pixel extracted from the masked region associated with the image, we compute the Mahalanobis distance in each direction as

$$\phi_n = \sqrt{(E(\mathbf{Q}_n) - E(\mathbf{W}_n))^{-1} \mathbf{C}_{\mathbf{Q}_n} E(\mathbf{Q}_n) - E(\mathbf{W}_n)} \tag{4}$$

and

$$\beta_n = \sqrt{(E(\mathbf{W}_n) - E(\mathbf{Q}_n))^{-1} \mathbf{C}_{\mathbf{W}_n} E(\mathbf{W}_n) - E(\mathbf{Q}_n)}, \tag{5}$$

where $E(\mathbf{Q}_n)$ and $E(\mathbf{W}_n)$ are the mean RGB colours of $\mathbf{Q}_n$ and $\mathbf{W}_n$; $\mathbf{C}_{\mathbf{Q}_n}$ and $\mathbf{C}_{\mathbf{W}_n}$ are the covariance of $\mathbf{Q}_n$ and $\mathbf{W}_n$ respectively. For each pair of corresponding regions we find the
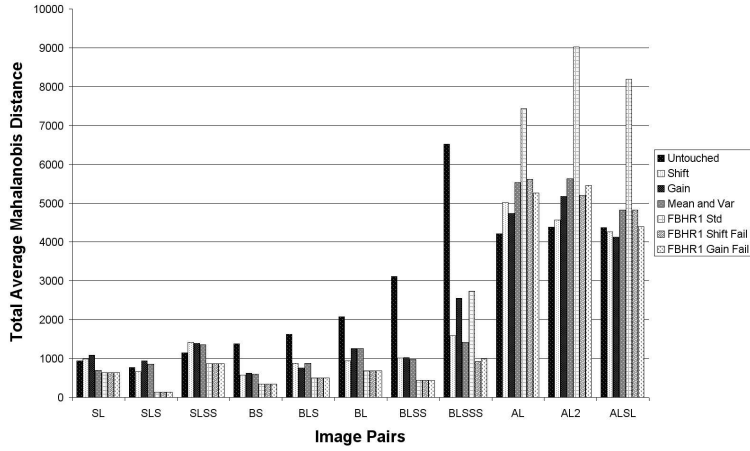
Figure 3: Total average Mahalanobis distances for different image pairs aligned using alignment of moments and FBHR1 and FBHR1 with different failure strategies. FBHR1 shift first applies FBHR1 and then applied a shift alignment of the 1st moment in any channel that fails. FBHR1 applied a gain shift of the 1st moment in any channel that fails.

average of these distances

$$J_n = \frac{\phi_n + \beta_n}{2}, \tag{6}$$

and the overall metric is the sum of the averages of ϕ_n and β_n, $\sum_{n=1}^{L} J_n$. The metric provides us with an overall measure of how closely the corresponding colour clusters are aligned, the measure is chosen over standard overlap measures such as the Bhattachayya coefficient as all clusters do not necessarily overlap after alignment. Overlap measures are dominated by largest overlapping clusters; however, we are interested in how well all the different colours in the scene are aligned. This metric is sensitive to separations between the means of corresponding clusters, but is potentially insensitive to orientation differences between clusters with closely aligned means. Manual inspection suggests that such configurations rarely arise, but we remain wary of this limitation.

Figure 3 summarises the result of applying the different moment alignment strategies and FBHR with two different failure strategies on all different datasets. The FBHR Shift and Gain fail transforms first attempt to apply FBHR in each channel, if a failure is detected they revert to alignment using a shift or gain alignment of the first moments respectively. We see that for the datasets SL,SLS,SLSS,BS,BLS,BL and BLSS no FBHR channel failure occurs and FBHR outperforms alignment using moments. On the BLSSS example the channel failure methods improve over FBHR1 as a channel has failed, this strategy also outperforms aligning moments. For the stuffed animal image pairs, AL and AL2 show worse alignment scores under all transformations and ALSL shows slight improvement when aligning moments. On these images the significant shadowing effects cause cluster variation that cannot be resolved by examining the individual R,G and B histograms alone. FBHR failure modes cause the standard FBHR algorithm to degrade in performance more gracefully. However, only the BLSSS pair shows an improvement over
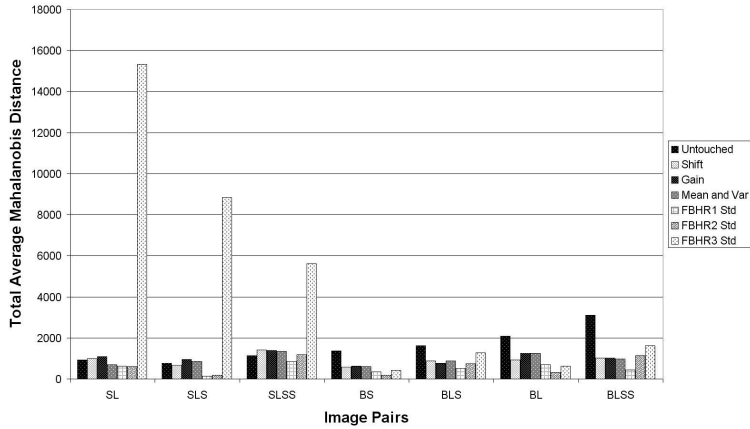
Figure 4: Comparison of moment alignment and FBHR for 1st, 2nd and 3rd order warps for image pair examples that did not detect FBHR failure in any channel.

moment alignment when using this strategy. This tells us that a practical strategy when FBHR failure has been detected is to align the 1st moments of all channels using a shift or gain; applying moment alignment transformations to individual channels on failure does not provide a significant overall improvement.

A further question is how the performance of FBHR compares with polynomial transformations of order 1, 2 and 3. Figure 4 examines FBHR 1, 2 and 3 for the example datasets that did not cause FBHR failure in a channel, we see that FBHR1 outperforms alignment of moments across all examples. Either FBHR1 or FBHR2 are found to perform best. Order 3 fails catastrophically for the strips data set because the histograms have too few features to fit an order 3 polynomial. The book images provide enough features, but the regularizing effect of the lower-order transformations proves advantageous. Order 2 provides little improvement over order 1 so we recommend order 1 by Occam's razor.

The method cannot automatically select the best transform which must be done empirically. However, FBHR 1 has proven robust across a wide range of examples. Note that alignment of the 1st and 2nd moment and FBHR 1 are both linear transformations so results are directly comparable. The fact that FBHR1 produces better alignments validates the feature based registration approach over moment alignment.

5 Conclusions and further work

We have introduced a new algorithm capable of aligning colour histograms that are noisy and include variation due to lighting, scale and small shadow changes (as seen in the book dataset). We have tested the algorithm for aligning a range of colour histograms obtained from pairs of images with systematically introduced variation of lighting, scale and shadowing. The experiments align the 1D individual R, G and B distributions of source and target distributions and have shown quantitative improvements in the 3D RGB colour histogram alignment over the simple methods of aligning moments. We have shown the merits of a feature-based histogram registration approach for aligning multi-modal data

and how to deal with its limitations.

One potential improvement would compute features directly from the 3D histograms rather than the 1D histograms of each colour channel. However, peak detection in 3D is more difficult than in 1D. We also plan to test the algorithm on more complicated scenes and evaluate the performance of standard registration methods such as ICP; RANSAC will be explored as an outlier rejection method. An increased number of colours in the scene will increase the complexity of the matching, we plan to reformulate the matching procedure as a bi-partite matching problem that penalises folding correspondences. The Hungarian algorithm [1] is a candidate method for solving the matching problem in $O(n^3)$ time. We have concentrated on the two-frame histogram registration problem, but plan to extend the method to allow multiple-frame histogram registration; methods to align multiple 3D data scans [2] may provide insight into the best way to do this. Finally, we intend to evaluate whether recently introduced methods for histogram comparison across overlapping and non overlapping clusters such as [6] improve the experimental evaluation.

References

[1] Wikipedia-Hungarian algorithm description. http://en.wikipedia.org/wiki/hungarian_algorithm.

[2] D.W. Eggert, A.W. Fitzgibbon, and R.B. Fisher. Simultaneous registration of multiple range views for use in reverse engineering of CAD models. *Computer Vision Image Understanding*, 69(3):253–272, 1998.

[3] A. Ilie and G. Welch. Ensuring color consistency across multiple cameras. In *The 10th IEEE Conference on Computer Vision (ICCV)*, Beijing, China, October 2005.

[4] K. Jeong and C. Jaynes. Object matching in disjoint cameras using a color transfer approach. *Special Issue of Machine Vision and Applications Journal*, 2006.

[5] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Springer, 1994.

[6] H. Ling and K. Okada. Diffusion Distance for Histogram Comparison. *CVPR*, 1:246–253, 2006.

[7] F. Pitie, A.C. Kokaram, and R. Dahyot. N-Dimensional Probability Density Function Transfer and its Application to Colour Transfer. *ICCV'05 Volume 2*, pages 1434–1439, 2005.

[8] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley. Color transfer between images. *Computer Graphics and Applications, IEEE*, 21(5):34–41, 2001.

[9] G.Y Tian, D. Gledhill, and D. Taylor. Colour correction for panoramic imaging. *Sixth International Conference on Information Visualisation*, pages 483–488, 2002.

[10] L.A. Torres-Mendez and G. Dudek. Color Correction of Underwater Images for Aquatic Robot Inspection. *Lecture notes in computer science*, 3757:60, 2005.