University of Kentucky Master's Theses   Graduate School

2007

OBJECT MATCHING IN DISJOINT CAMERAS USING A COLOR TRANSFER APPROACH

Kideog Jeong
University of Kentucky, kideog@vis.uky.edu

Recommended Citation

Jeong, Kideog, "OBJECT MATCHING IN DISJOINT CAMERAS USING A COLOR TRANSFER APPROACH" (2007). *University of Kentucky Master's Theses*. 434.
https://uknowledge.uky.edu/gradschool_theses/434

ABSTRACT OF THESIS

OBJECT MATCHING IN DISJOINT CAMERAS
USING A COLOR TRANSFER APPROACH

Object appearance models are a consequence of illumination, viewing direction, camera intrinsics, and other conditions that are specific to a particular camera. As a result, a model acquired in one view is often inappropriate for use in other viewpoints. In this work we treat this appearance model distortion between two non-overlapping cameras as one in which some unknown *color transfer function* warps a known appearance model from one view to another. We demonstrate how to recover this function in the case where the distortion function is approximated as general affine and object appearance is represented as a mixture of Gaussians. Appearance models are brought into correspondence by searching for a bijection function that best minimizes an entropic metric for model dissimilarity. These correspondences lead to a solution for the transfer function that brings the parameters of the models into alignment in the UV chromaticity plane. Finally, a set of these transfer functions acquired from a collection of object pairs are generalized to a single camera-pair-specific transfer function via robust fitting. We demonstrate the method in the context of a video surveillance network and show that recognition of subjects in disjoint views can be significantly improved using the new color transfer approach.

KEYWORDS: Computer Vision, Wide-Area Multi-Camera Surveillance Network, Object Matching, Color, Appearance Model

Kideog Jeong

November 27th, 2006

OBJECT MATCHING IN DISJOINT CAMERAS
USING A COLOR TRANSFER APPROACH

By

Kideog Jeong

_____
Director of Thesis

_____
Director of Graduate Studies

_____

RULES FOR THE USE OF DISSERTATIONS

Name   Date

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

THESIS

Kideog Jeong

The Graduate School

University of Kentucky

2006

OBJECT MATCHING IN DISJOINT CAMERAS
USING A COLOR TRANSFER APPROACH

_____

THESIS

_____

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science in the
College of Engineering
at the University of Kentucky

By

Kideog Jeong

Lexington, Kentucky

Director: Dr. Christopher Jaynes, Associate Professor of Computer

Science

Lexington, Kentucky

2006

ACKNOWLEDGEMENTS

Table of Contents

List of Tables

List of Figures

List of Files

jeong.pdf

Chapter 1

Introduction

Persistent tracking of moving objects through multiple cameras is a challenging problem. This problem is made more difficult by sparse camera networks that cover large geographic areas with diverse imaging conditions. A primary challenge is that extracted appearance models, captured in one camera, are coupled with the specific spectral response of each sensor, characteristics of the local illumination conditions, viewing direction, and other confounding factors. The result is that appearance models acquired in one view are not generalizable to other cameras.

Fig. 1.1 depicts the problem for two cameras observing the same moving object in disjoint views and shows how the distribution of chromaticity measured for all pixels on the object throughout the video sequence are distorted from one view to the next. This color distortion function encompasses the factors that lead to inconsistent appearance between any two cameras. It is appropriate to ask if this color distortion can be recovered from samples such as the one in Figure 1.1 and then taken into account to improve object matching.

In this work we treat this appearance model distortion between two non-overlapping cameras as one in which some unknown function, h, warps a known appearance model, Θ from one view to another. This is a high-dimensional function (see [32] for an exploration of the dimensionality of image change under more constrained illumination effects) and cannot be represented in the simple 2D UV-plane. However, approximations to this function will allow it to be estimated and then corrected. In this work, we explore to what extent this is possible.

Our approach, then, is to estimate this unknown function for any pair of cameras for a particular choice of appearance model by observing transitions of objects as
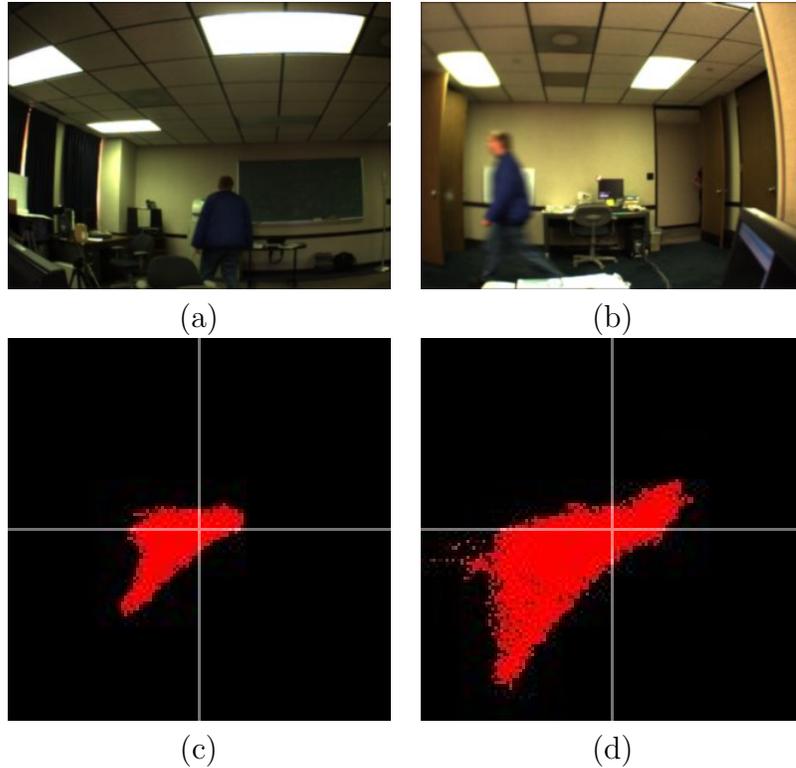
Figure 1.1: The same subject observed in two views (a-b), is illuminated differently with different camera intrinsics, and under different pose in a surveillance network. This leads to a change in chromaticity samples measured on the moving object. (c-d). This deformation should be taken into account to support persistent object tracking in multiple views.

they move between two views in uncontrolled conditions. It is assumed that the conditions under which this function is estimated remain fixed when the function is then applied to objects for recognition. For example, cameras remain stationary and the illumination conditions at each camera do not change.

Rather seeking an h that models the complex physical sources of color appearance between two views, we select a parameterization of h that is capable of capturing the class of distortions that a model will undergo. A tradeoff between complexity of this *color transfer function* and its ability to capture expected color variability exists. This tradeoff has impact on both learning the model and the computational costs in applying the model at run-time. Here we explore a model that can be learned and

then applied in real-time and is capable of capturing a large percentage of appearance variation for individuals moving through a surveillance network.

Given our emphasis on non-overlapping cameras, results in cooperative multi-camera tracking cannot be easily applied to this domain. Traditionally, different appearance models have been used to complement spatial tracking in more than one camera [16], or increase tracking robustness through occlusions [3]. However, these methods are focused on cameras with overlapping fields of view. Furthermore, work that studies how changes in illumination can be characterized in a low-dimensional space and perhaps compensated [32] require approximate knowledge of the three-dimensional surface under observation and is inappropriate for the tracking domain where no such a priori object model is available.

Our approach [13] is complementary of other work that is focused on consistent tracking across non-overlapping fields of view. For example Rahimi et al. [4] introduce the use of spatial-temporal constraints to consistently track objects between disjoint views by explicitly estimating the spatial relationship between several cameras. Alternatively, Kang et al. [14] use a pan-tilt-zoom camera to maintain persistent tracking of objects as they leave the field of view of stationary cameras to avoid assumptions about constant velocity. In fact, this work is motivated by the need to augment these results with the ability to predict appearance change in the different cameras.

In response to this need, several researchers [35, 34] have been proposed methods that deal with object matching between non-overlapping cameras using image edges as object features under an assumption that the edge maps of two vehicles are already aligned for use of vehicle matching. The results of this method are quite promising and are applicable to multi-camera traffic analysis and flow-prediction. However, in practice extracted edges of two images of the same vehicle may be quite different when the images are taken under different illumination conditions, different viewing angles, and other confounding factors. As a result, the feature-based approach may

not extend to general surveillance scenarios. Moreover, edge-based methods can be instable when applied to deformable objects such as people.

Other researchers have explored how to address color and intensity changes across disjoint views. Porikli et al. [10] attenuates color differences by utilizing a similarity metric that is somewhat invariant to changes in global illumination. Javed et al. [29] have used a low dimensional subspace method for a brightness transfer function via principal component analysis (PCA) on a set of known intensity mappings obtained from object observation samples, under some linearity assumptions of the transfer function and independence of the color channels. To increase accuracy, their matching scheme is then combined with additional space-time cues from known camera topologies. However, there are some drawbacks to the subspace based transfer function. One drawback would be that the choice of subspace dimensionality may affects the accuracy of matching. Moreover, since the intensity-based models are inherently sensitive to small changes in illumination [25], the linear subspace based transfer function which is a static representation have to be rebuilt to incorporate new intensity mappings as imaging conditions change.

We develop a method that operates on chromaticity samples to increase the temporal stability of the transfer function between any camera pair. Furthermore, the approaches proposed by [10, 29] are quite different than the work here that explicitly constructs a model of the expected color transfer between views. We emphasize the importance of learning a transfer function that is specific to the appearance model in question (rather than the entire space) to reduce the dimensionality of the model to be learned.

It should be pointed out that our work [13] is similar in spirit to Javed et al. [29] but there also exist several important distinctions in the technical components used by the two approaches. A comparison of the technical components are presented in Table 1.1.

Table 1.1: Comparison of the technical components used Javed et al. [29] and our method.

Method	Feature space	Representation model	Transfer function	Independence of the color channels
Javed et al. [29]	Intensity	1D histogram	Linear subspace	Yes
Our approach [13]	Chromaticity	2D GMM	Affine transformation	No

1.1 Problem Description and Methodology

Consider two stationary cameras c_a and c_b with non-overlapping fields of view and differing illumination conditions. A set of foreground color samples are acquired by tracking objects in motion. As the object moves through the scene (UV chromaticity samples are extracted from the YUV color space and each channel is represented by one byte in the range of $[0, 255]$.). Tracking can be accomplished using any number of methods [6, 2, 31]. In this work, we use the foreground tracking algorithm by [22, 31] that is capable of generating few false positives and eliminates moving shadows at real-time rates. Although the framework can be extended to support other parametric models, a Gaussian mixture model (GMM) that represents the probability distribution of the chromaticity of pixels on the object in all tracked frames is fit to the observed foreground samples. The choice of the GMM is motivated by its common use within the community for color-blob tracking and foreground region modeling [16, 7, 37, 28]. In addition, parametric models had important advantages over non-parametric representations such as a histogram of kernel (Parzen window) in that they are compact and require fewer learning parameters.

As an object leaves the view of the camera, the acquired GMM model is transmitted to a server along with a unique identifier for further processing. This unmatched object is then compared to other unmatched models previously observed by the network. If correspondence is available, the two disjoint appearance models are provided as input into the algorithm that computes the color transfer between the two cameras.

These training pairs can be generated via other sources of information such as RFID or by running a multi-view matching algorithm that, given enough time, will produce these matches with low probability for false positives.

We approximate the unknown transfer function, h, with a general affine transformation. The use of an affine model was inspired by work in colorimetic modeling for both CRTs and other color devices [21, 15, 17] and our results justify this choice (see Chapter 3). Estimation of this transfer function proceeds in three steps. First, given the two GMM models, their different components (means and covariances of each mode) must be assigned correspondence. Secondly, the two models, now in correspondence, are used to compute the affine deformation that will bring them into alignment. Finally, a camera-pair specific transfer function that can generalize to other objects at run time is derived. This leads to a set of pairwise functions that can be used to predict appearance changes as subjects move through the network.

1.2 Assumptions

The approach relies on several assumptions of varying complexity. The robustness of the system with respect to degradation of violation of these assumption is studied in the experimental results chapter. In fact, removing one or more of these assumptions can motivate future work in the area.

Perhaps the most important assumption is that the color distribution on an object is somewhat isotropic. This assumption allows us to operate in an uncalibrated manner between the cameras. For example, cameras can be deployed over a wide area (throughout an office environment in our studies) without need for geometric registration which is very difficult to recover when view frustums do not overlap [4]. This assumption typically holds but can be a problem for subjects who are wearing shirts with significant color markings on their back, for example.

In addition, we assume that subjects do not exhibit significant self-shadowing.

This is often not a problem in indoor environments but strong self-shadows can corrupt color measurements when the object is directly illuminated by strong sunlight. When color measurements fail due to self-shadowing or non-isotropic color distributions, our system detects the color match candidate as an outlier and discard it from learning processing. In this way, the approach is somewhat tolerant to these types of conditions. However, automatic detection and removal self-shadows from moving objects to build invariant color models is an important and interesting problem.

We assume the color model acquired for a moving subject is invariant with respect to its position in the camera frame. This location-independent assumptions implies that we seek to derive a color transfer function that general to an entire camera pair and not pixels (or subregions) within the pair. Extensions to the framework can potentially incorporate more accurate transfer models that relate different regions within camera pairs (for example, a shadowed region in camera A to an unshadowed region in camera B).

Finally, we do not seek to color calibrate each camera in the network independently. This is the focus of a significant research effort that has produced results we can exploit here [23, 24]. Instead, we simply assume that each camera's saturation parameters are hand adjusted beforehand to maximize the response among the three color channels while observing a color target.

1.3   Contributions

The contributions of this work [13] is primarily a study of a low computational color transfer function that can be applied to real-time persistent tracking of moving objects in the multi-camera surveillance networks that cover large geographic areas with diverse imaging conditions.

In particular the contributions of the thesis can be listed as follows:

1. A novel color transfer approach to robust object matching across disjoint cam-

eras.

2. An efficient estimation of the color transfer function

   (a) Bijective mapping scheme in the chromaticity space between distored appearance models, of which each models the foreground distribution of a particular camera.

   (b) A robust refining process of the function parameters.

3. A low computational color transfer function that can be applied to real-time persistent tracking in a network of non-overlapping cameras.

4. Extensive performance evaluations of the proposed method in the context of multi-camera surveillance network and comparisons to other color correction/calibration approaches under varying complexity.

Chapter 2

Technical Details

Given a set of chromaticity samples of a tracked object $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ and an initial model estimate Θ_{init}, we iteratively fit the parameters of a Gaussian mixture model Θ that maximizes the probability of $p(X|\Theta)$ using the EM algorithm [1], where $\Theta = \{\omega_i, \mu_i, \Sigma_i\}_{i=1}^{k}$ are the parameters of k, d-dimensional Gaussians. We denote the weight of component i as $\omega_i > 0$ so that $\sum_{i=1}^{k} \omega_i = 1$, mean vectors as $\mu_i \in \Re^d$ and covariance matrices as $\Sigma_i \in \Re^{d \times d}$.

The number of mixture components should accommodate the major color distributions of the object as well as potential appearance variation due to changes in lighting throughout the tracking sequence. To achieve this goal, we sort the k Gaussian components returned by the EM algorithm in a decreasing order by $\omega_i / \det(\Sigma_i)$, select only the top m components that represent the dominant colors of the object, and then normalize the weights so that $\sum_{i=1}^{m} \omega_i = 1$. This technique is quite well known and has been used to model appearance in a number of different domains [33, 6].

We now generalize the individual (and independent) color models to the case where a pair of corresponding models of the same object are derived from two different views. Given the models, $\Theta = \{\theta_i = (\omega_i, \mu_i, \Sigma_i) \mid 1 \leq i \leq m\}$ and $\tilde{\Theta} = \{\tilde{\theta}_i = (\tilde{\omega}_i, \tilde{\mu}_i, \tilde{\Sigma}_i) \mid 1 \leq i \leq m\}$, correspondence between the constituent model components must be established in order to recover the color transfer function between them. Of course, recovering correspondence between model components is difficult because the observed models have already been distorted due to illumination, viewing parameters, and sensor response differences. We formulate the problem of corresponding the model components as one of finding a bijective map γ that minimizes some similarity metric

$\mathcal{D}$:

$$\gamma = \arg\min_{\Gamma} \left[\mathcal{D}\left(\Theta, \tilde{\Theta}, \gamma^j\right) \right], \tag{2.1}$$

where $\Gamma = \{\gamma^j \mid 1 \le j \le m!\}$ is the set of all possible bijection functions (i.e. each $\gamma^j : \mathcal{S} \to \mathcal{S}$, where $\mathcal{S} = \{1, \ldots, m\}$). Although the specifics of model component similarity metrics are described in the next Sections, the general form of $\mathcal{D}$ is the log-sum inequality [38]. This form computes the model-wise dissimilarity according to bijection function γ for a specific definition of component similarity Ψ:

$$\mathcal{D}\left(\Theta, \tilde{\Theta}, \gamma\right) \overset{\text{def}}{=} \sum_{i=1}^{m} \omega_i \left[\Psi\left(\theta_i \| \tilde{\theta}_{\gamma(i)}\right) + \log \frac{\omega_i}{\tilde{\omega}_{\gamma(i)}} \right]. \tag{2.2}$$

Ψ, then, is a similarity metric for each model component correspondences (assigned during bijective mapping) that will be described in the next Section.

2.1 Component-wise similarity metric using bijective mapping in chromaticity space

In this section, we describe the component-wise similarity metric, Ψ, that forms the basis for both assigning correspondence and computing model similarity. Secondly, we will also detail an algorithm for computing a bijection function γ between any two models known to have arisen from the same object.

We denote δ as the dissimilarity metric between any two components θ and $\tilde{\theta}$ given by the relative entropy, also called known as the Kullback-Leibler divergence [36, 9]:

$$\delta\left(\theta \| \tilde{\theta}\right) = \frac{1}{2} \left[\log \frac{|\tilde{\Sigma}|}{|\Sigma|} - d + trace\left(\tilde{\Sigma}^{-1}\Sigma\right) + \left(\mu - \tilde{\mu}\right)^{\top} \tilde{\Sigma}^{-1} \left(\mu - \tilde{\mu}\right) \right] \tag{2.3}$$

An intuitive metric then, is a symmetrized entropy between the two model components:

$$\Psi_1\left(\theta_i \| \tilde{\theta}_{\gamma(i)}\right) = \delta\left(\theta_i \| \tilde{\theta}_{\gamma(i)}\right) + \delta\left(\tilde{\theta}_{\gamma(i)} \| \theta_i\right). \tag{2.4}$$

Given this metric, a bijective map, γ, can be discovered via brute-force search of all $m!$ possible permutations using the Ψ_1 metric. In the case that the number of

components is relatively low, the computational cost of this approach may not be an issue. However, it becomes inefficient as the number of components increases (the time complexity of a tighter upper bound is $o(m^m)$ given by Stirling's approximation). Moreover, this approach has significant potential for yielding a false bijection function. As depicted in Fig. 2.1, when the two components, θ_2 and $\tilde{\theta}_1$, are geometrically aligned, the model-wise distance (in terms of $\mathcal{D}$) with the false map, γ_a, can be smaller than or equal to that with the true map, γ_b.
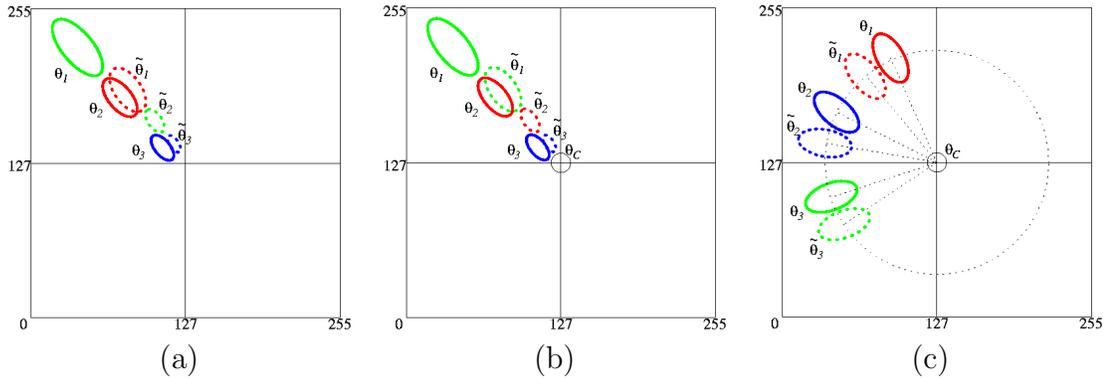
Figure 2.1: Assignment of model component correspondence is important to accurate transfer function estimation. Two models are plotted on the same chromaticity plane. (a) Straightforward pair-wise similarity metric using Ψ_1 can lead to a false bijective map ($\gamma_a(i) = (2, 1, 3)$) when $\mathcal{D}(\Theta, \tilde{\Theta}, \gamma_a) \leq \mathcal{D}(\Theta, \tilde{\Theta}, \gamma_b)$ where $\gamma_b(i) = (1, 2, 3)$ is considered to be true bijective map. (b) This is avoided by utilizing characteristics of the chromaticity plane where the achromatic locus acts as the polar origin. The relative entropy between each proposed correspondence and an achromatic distribution (θ_C) at this point is taken into account (see Text). (c) Relative orientation with respect to the achromatic locus resolves ambiguities arising from two models whose relative distances to θ_C are uniform.

A new metric that takes into account the importance that the achromatic locus plays in color space [21, 15] is needed. When color measurements are represented in polar coordinates, this point is the pole in the chromatic plane and distances between different color models must be made with respect to this polar coordinate system. The spectral composition of lights that appear achromatic has been of long-standing interest in colorimetry and color vision theory and our work is inspired by successful

approaches [21, 15, 17] using achromatic adjustment to discover the colorimetric relationship between color spaces. We therefore introduce an *intermediate component* θ_C whose parameters are $(\omega = \frac{1}{m}, \mu = (127, 127), \Sigma = \mathbf{I})$. This represents an impulse function, located at the achromatic locus against which all components can now be compared.

Given this intermediate component, an efficient algorithm to compute an appropriate bijection function γ between a pair of models that will be plugged into a new component similarity metric Ψ_2 can now be defined. Firstly, the similarity between the m components of an appearance model Θ with respect to the intermediate component is computed as $d_i = \Psi_1(\theta_i \| \theta_C)$ for $1 \leq i \leq m$. Components are then sorted in increasing order by d_i and let $A = (a_1, \ldots, a_m)$ be the sequence of the ordered indices. For the corresponding appearance model $\tilde{\Theta}$ in the second view, the sequence of the ordered indices, $B = (b_1, \ldots, b_m)$, can be similarly obtained by first computing $\Psi_1\left(\tilde{\theta}_j \| \theta_C\right)$ for $1 \leq j \leq m$ and sorting them in increasing order.

The sorting steps lead to a rank ordering of potential bijections. However, we need to take into account the case depicted in Fig. 2.1(c) that several components are equidistant in the polar plane. This can be disambiguated via their angular separation by computing the angular distance between the component means (written as a vector in the polar plane). Thus we subsort the components in the second view using an angular metric, $\cos^{-1}\left(\frac{\mu_{a_i} \tilde{\mu}_j}{|\mu_{a_i}||\tilde{\mu}_j|}\right)$ for all $1 \leq j \leq m$. In other words, $\Psi_1\left(\tilde{\theta}_j \| \theta_C\right)$ term is the primary sort field and the angular term is the secondary sort field that acts to disambiguate cases where color modes lie on the same radius from polar origin. Finally we obtain the subsorted sequence of the ordered indices of B, $B' = (b'_1, \ldots, b'_m)$.

As a result, the globally best bijection function γ is then constructed with the mappings of a_i and its corresponding $\gamma(a_i) = b'_j$ for all i. Intuitively, this algorithm gives priority to the color components closest to the achromatic locus and smallest angular distance, and assigns these components a match from the second model before

assigning matches to components that are more distant from the polar origin.

Given the computed bijection function γ above, a new component similarity metric, Ψ_2, written in terms of Ψ_1 can now be defined as:

$$\Psi_2\left(\theta_i\|\tilde{\theta}_{\gamma(i)}\right) = \Psi_1\left(\theta_i\|\theta_C\right) + \Psi_1\left(\tilde{\theta}_{\gamma(i)}\|\theta_C\right). \tag{2.5}$$

The first and second terms of Eq. (2.5) are the entropic distances between each component to the new intermediate component. Note that this is a similarity metric using the Kullback-Leibler divergence on the mixture of Gaussian model components and is not applied directly to the underlying intensity or color distributions.

This complete matching assignment γ and the component similarity metric Ψ_2 are then used with Eq. (2.2) to define $\mathcal{D}$ which will be used to check global model consistency as described in Eq. (2.12) as well as to compute distortions among models directly.

We now must discover a color transfer function that maximizes this color similarity metric. The color transfer function operates on the model parameters and is dependent on the camera used to derive the color models. This process is discussed in the following section.

2.2 Estimating the parameters of color transfer function

Given a pair of appearance models arising from the same object seen in two different cameras, $(\Theta, \tilde{\Theta})$ and the correspondence mapping between their components, γ, we write a functional form of the relationship between the components with respect to the function h with unknown parameters $\mathbf{T}$ as:

$$\tilde{\Theta} \approx h_{\mathbf{T}}(\Theta). \tag{2.6}$$

A global transform between two models is computing using only those components that are close to the polar origin. This region of color space has been shown to be

important in understanding the effect of illumination changes on color since the loci of achromatic points establishes equivalent chromatic appearances in the chromaticity space [21, 15, 17, 18].

Given the bijection function γ, and our definition of model similarity, $\mathcal{D}$, a model component pair that minimized $\mathcal{D}$ is selected:

$$a = \arg \min_i \left[\Psi_2 \left(\theta_i \| \tilde{\theta}_{\gamma(i)} \right) \right] \tag{2.7}$$

for all $i = 1, \ldots, m$, where m is the number of components that represent the dominant colors of the object. It should be noted that although all model components are used in the computation of γ, at this stage we only operate on these particular component pairs. This may seem counterintuitive but allows us to utilize all the information in the model to accurately assign the bijection function while retaining the robustness of using color modes that are closest to the achromatic points. These corresponding model components are far more reliable in a learning process that can be biased when using outlying color components.

The selected component pair $(\theta_a, \tilde{\theta}_{\gamma(a)})$ is used to extract the parameters of the color transfer function $\mathbf{T}$ by corresponding the component means,

$$\tilde{\mu}_{\gamma(a)} = \mathbf{A}\mu_a + \mathbf{t}, \tag{2.8}$$

and covariances,

$$\tilde{\Sigma}_{\gamma(a)} = \mathbf{A}\Sigma_a\mathbf{A}^\top, \tag{2.9}$$

where $\mathbf{A} \in \Re^{d \times d}$ is the affine matrix, and $\mathbf{t} \in \Re^d$ is the translation vector in the polar plane.

In order to reduce the complexity in solving the quadratic equation given in Eq. (2.9), we introduce whitening transformation matrices $\mathbf{W}$ and $\tilde{\mathbf{W}}$ that can be obtained by performing an eigenvalue decomposition on each covariance matrix Σ_a and $\tilde{\Sigma}_{\gamma(a)}$:

$$\mathbf{W} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^\top \text{and } \tilde{\mathbf{W}} = \tilde{\mathbf{E}}\tilde{\mathbf{D}}^{-1/2}\tilde{\mathbf{E}}^\top,$$

14

where $\mathbf{E}$ and $\tilde{\mathbf{E}}$ are the orthogonal matrices of eigenvectors, and $\mathbf{D}$ and $\tilde{\mathbf{D}}$ are the real diagonal matrices of its eigenvalues.

By multiplying $\tilde{\mathbf{W}}^{-1}$ and $\mathbf{W}$, we get a solution $\mathbf{A}$ for the affine parameters:

$$\mathbf{A} = \tilde{\mathbf{W}}^{-1}\mathbf{W} \tag{2.10}$$

such that $\tilde{\Sigma}_{\gamma(a)} = \mathbf{A}\Sigma_a \mathbf{A}^\top$. This can be simply proved by substituting the solution in Eq. (2.10) into Eq. (2.9):

$$\overbrace{\tilde{\mathbf{E}}\tilde{\mathbf{D}}^{1/2}\tilde{\mathbf{E}}^\top\mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^\top}^{\mathbf{A}}\overbrace{\mathbf{E}\mathbf{D}\mathbf{E}^\top}^{\Sigma_a}\overbrace{\mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^\top\tilde{\mathbf{E}}\tilde{\mathbf{D}}^{1/2}\tilde{\mathbf{E}}^\top}^{\mathbf{A}^\top}$$

$$= \tilde{\mathbf{E}}\tilde{\mathbf{D}}^{1/2}\tilde{\mathbf{D}}^{1/2}\tilde{\mathbf{E}}^\top = \tilde{\mathbf{E}}\tilde{\mathbf{D}}\tilde{\mathbf{E}}^\top = \tilde{\Sigma}_{\gamma(a)} \tag{2.11}$$

since $\tilde{\mathbf{E}}^\top\mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^\top\mathbf{E}\mathbf{D}\mathbf{E}^\top\mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^\top\tilde{\mathbf{E}} = \mathbf{I}$.

Note that unlike the general image alignment domain where the rotational ambiguity in determining affine parameters arises, our goal can simply be solved by exploiting the rotation matrix by the decomposition of the covariance matrix. That is, the direction of the covariance axes does not impact the similarity of the two covariance matrices. In this aligned space, one can still rotate one Gaussian with respect to another without problem because of the rotational symmetry in the aligned space.

This is true because the orthogonal matrices, $\mathbf{E}$ and $\tilde{\mathbf{E}}$ represent rotation angles with respect to the standard basis, the rotational component of the affine matrix $\mathbf{A}$ can be determined by multiplying the inverse of the second orthogonal matrix $\tilde{\mathbf{E}}$ by the first orthogonal matrix $\mathbf{E}$. Again, thinking geometrically, each ellipse is defined by the parameters of the covariance matrix can firstly be aligned with the chromaticity coordinate axes by the inverse of the second orthogonal matrix and then be rotated by the first orthogonal matrix.

The translation parameters, $\mathbf{t}$, are then computed simply by substituting $\mathbf{A}$ into Eq. (2.8). Finally, the parameters of the color transfer function between the two color models $\mathbf{T}$ are given by $\mathbf{T} = [\mathbf{A}|\mathbf{t}]$.

We now evaluate if the parameter set $\mathbf{T}$ obtained from a particular object correspondence should be retained and used to estimate a global, camera-specific affine transform. This step is important and can increase the robustness of the transformation estimate when foreground pixels are detected in error, local occlusions, or sparse samples lead to an estimate of $\mathbf{T}$ that is inconsistent. Since the affine transformation is a map $\mathbf{T} : \Re^n \rightarrow \Re^n$ that is the composition of an invertible linear map, we can write the global transformation error Φ as a symmetric metric:

$$\Phi(h_{\mathbf{T}}, \Theta, \tilde{\Theta}, \gamma) = \left[\frac{1}{2} \left(\mathcal{D}\left(h_{\mathbf{T}}(\Theta), \tilde{\Theta}, \gamma\right)^2 + \mathcal{D}\left(\Theta, h_{\mathbf{T}^{-1}}(\tilde{\Theta}), \gamma\right)^2 \right) \right]^{1/2}. \qquad (2.12)$$

If the global error is below some threshold, τ, (for the results shown here, τ was set to 1.2), $\mathbf{T}$ is then accepted as a member of the parameter collection $\mathcal{T}$ to be used to estimate a camera-pair-specific color transfer model. If the resulting error exceeds τ, then it is discarded. This procedure is to eliminate the model correspondence from the learned transitions, in cases where objects transition observations are degenerate (i.e. tracking fails significantly in one view.) or our assumptions (see Section 1.2) are violated. In practice, this is quite rare and generally a transition model can be acquired quickly from few observations.

2.3 Refining the color transfer function

Given a set of model-specific affine deformations corresponding to a single camera pair in the surveillance network, these are now combined into a single color transfer function that can then be used to map general appearance models in one camera to their expected values in the next. The robust Least Median of Squares technique [30] is used to derive a camera-pair specific transfer function, $\hat{\mathbf{T}}$, from a collection of appearance model correspondences: $\mathcal{G} = \{(\Theta_i, \tilde{\Theta}_i) \mid i = 1, \ldots, M\}$ taken from a pair of views (c_a, c_b) as well as the corresponding collection of the color transfer function parameters between the two color models $\mathcal{T} = \{\mathbf{T}_j \mid j = 1, \ldots, N\}$.

The initial estimate of $\hat{\mathbf{T}}$, $\mathbf{T}_0$, is the transfer function that minimizes the medians of the squared residuals over all i models and j transformations:

$$\mathbf{T}_0 = \arg\min_{\mathbf{T}_j} \left[\text{median}_i \left(\Phi \left(h_{\mathbf{T}_j}, \Theta_i, \tilde{\Theta}_i, \gamma_i \right) \right) \right] \tag{2.13}$$

where Φ is the error function defined in Eq. (2.12). Simultaneously, an inlier set, $\mathcal{G}_{inliers}$, with respect to this median is determined by computing the transformation error $\varepsilon_i = \Phi(h_{\mathbf{T}_0}, \Theta_i, \tilde{\Theta}_i, \gamma_i)$ and comparing it to a threshold τ, for all i.

The iteration step begins with the affine distortion corresponding to each model contained in the inlier set $\mathcal{G}_{inliers}$ indexed by i, written as, $\mathbf{T}'_i$. These different models are linearly combined with unknown weights to derive the final camera-to-camera color transfer function:

$$\hat{\mathbf{T}} = \sum_{i=1}^{N} \alpha_i \mathbf{T}'_i \tag{2.14}$$

where N is the cardinality of $\mathcal{G}_{inliers}$ and α_i are the weighting coefficients. This weighting is necessary to give more weight to transfer functions which yield smaller transformation error, and consequently more likely to be representative of the underlying camera-pair color transfer. The weighting coefficients are normalized by the sum of the transformation errors:

$$\alpha_i = \begin{cases} \frac{\hat{\alpha}_i}{\sum_{i=1}^{N} \hat{\alpha}_i} & \text{if } \sum_{i=1}^{N} \varepsilon_i \neq 0 \\ \frac{1}{N} & \text{if } \sum_{i=1}^{N} \varepsilon_i = 0 \end{cases} \tag{2.15}$$

and $\hat{\alpha}_i$ is defined as:

$$\hat{\alpha}_i = \begin{cases} \frac{1}{N} & \text{if } \sum_{i=1}^{N} \varepsilon_i = 0 \text{ or} \\ & \qquad \sum_{i=1}^{N} \varepsilon_i = \varepsilon_i \, (\neq 0) \\ N & \text{if } \sum_{i=1}^{N} \varepsilon_i = \varepsilon_j \, (\neq 0, i \neq j) \\ 1 - \frac{\varepsilon_i}{\sum_{i=1}^{N} \varepsilon_i} & \text{otherwise.} \end{cases} \tag{2.16}$$

At each iteration, the color transfer function, $\hat{\mathbf{T}}$, is re-estimated and used to re-compute the transformation error $\varepsilon_i = \Phi(h_{\hat{\mathbf{T}}}, \Theta_i, \tilde{\Theta}_i, \gamma_i)$, for all i. Processing halts when the total transformation error: $\sum_{i=1}^{N} \varepsilon_i$ stops decreasing. Typically, the process

17

requires fewer than ten iterations and is quite efficient and discovering the optimal transfer function with ε. The result is a color function estimate that is robust to the presence of outliers, since we ignore the magnitudes of the largest residuals in $\mathcal{G}$.

Although a global and optimal transfer function may be valuable, the iterative approach we utilize has advantages in that it is robust with respect to a large number of outliers and can be coupled with any robust technique. Furthermore, the method can incrementally learn an updated color transfer function as new observations are available. This is important in a video surveillance context where color transfer functions are likely to become obsolete over time (i.e. as lights are turned on and off) but new color model samples via tracked subjects are readily available.

Chapter 3

Experimental results

In this section we demonstrate how the acquired color transfer function can be utilized in a wide-area video surveillance network to improve baseline appearance matching techniques. In particular, we study how the accuracy of the system under real-world imaging conditions including uncontrolled lighting and subjects wearing normal clothing. The technique is compared to other color correction/calibration approaches of varying complexity.

3.1   Multi-camera dataset

In the experimental setup, nine non-overlapping cameras from the "Terrascope" dataset [5], were used. These cameras were deployed in an indoor office environment under dramatically different illumination conditions (i.e. strong light from windows, fluorescent lights, and dark rooms lit by desk lamps) and various viewing conditions to capture subjects as they move from room to room throughout the space.

The dataset, consisting of eight individuals moving through the camera network, was divided into training and validation sequences. Each sequence was approximately three minutes long and contains video captured from all nine cameras. Initially, color transfer functions for all $(9 \times 8)/2$ pairs were computed using the method described in this paper. The training sequence contained four different individuals whose ground-truth transitions between cameras are part of the dataset were provided as input to the algorithm.

In order to study the utility of these color transfer functions, a second video surveillance sequence from the same camera network was captured and used as validation. Here the goal is to correctly identify subjects as they transition from one

camera to the next both with and without the trained color transfer functions. Eight difference subjects were used for validation who were seen in all nine camera views over the span of 10 minutes. At the end of processing, the similarity between the stored models is measured as:

$$s = \arg \min_j \left[\Phi \left(h_{\mathbf{T}}, \Theta, \tilde{\Theta}_j, \gamma_j \right) \right], \tag{3.1}$$

and the nearest model, s, (in terms of Φ) is labeled as a match. This matching process is performed with application of the color transfer function and without for comparison.

Table 3.1 depicts the object confusion matrix for all eight subjects when applying the similarity measure without the benefit of the corresponding color transfer function. Element i, j where column and row are indexed by i and j respectively in each Table represents the number of times subject i was recognized as subject j in another camera. Each subject's true positive rate is computed by dividing the value at (i, i) by the total number of times that the subject i appeared and is reported in the last column of the Table.

Table 3.1: Confusion matrix for all eight subjects seen throughout the experiment. Table contains true positive rates without applying the appropriate color transfer functions.

From \ To	A	B	C	D	E	F	G	H	%
A	12			5			6		52
B	5	10	5				3		43
C		6	9		5		4		37
D	6	4		9		3	5	4	29
E	3	5		2	5		4	2	24
F		1		5	2	4	5	4	19
G	6	2		6			7	2	30
H				4		4	5	7	35
Mean									**34**

The color transfer functions were then used to improve recognition rates across the different cameras. Let Θ_a and $\hat{\Theta}_b$ be an incoming appearance model sent from

camera a and a model sent from camera b stored in the central server respectively. The corresponding transfer function, $\hat{\mathbf{T}}_{ab}$, is then be chosen accordingly to be applied to map model Θ_a to the model space of camera b and vice versa by using Eq. (3.1). Table 3.2 shows the confusion matrix for all eight subjects when the transfer functions are used.

Table 3.2: Confusion matrix for all eight subjects observed in the experiment when using the derived color transfer functions. Table shows true positive rates for each subject. The use of color transfer functions leads to an overall improvement of over 50%.

From \ To	A	B	C	D	E	F	G	H	%
A	19			1			3		83
B		22			1				96
C			24						100
D	4			24			1	2	77
E	2	1		1	16		1		76
F						17	2	2	81
G	3			3			17		74
H				1		1		21	91
Mean									**85**

The mean recognition rate when using the new color transfer functions grows from 34% to 85%. The results demonstrate that the color transfer models can support the recognition of subject transfer between disjoint views even when using a relatively simple color matching scheme.

Obviously more sophisticated object recognition schemes can improve these results somewhat, however, any color-based appearance models will be degraded by the diverse imaging conditions and the poor results here are unsurprising but can serve as baseline to study improvement using the color transfer functions.

3.2 Object pixel classification

The color transfer functions were then tested to show the pixel classification rates of objects across the different cameras. Fig. 3.2 shows a typical camera transition
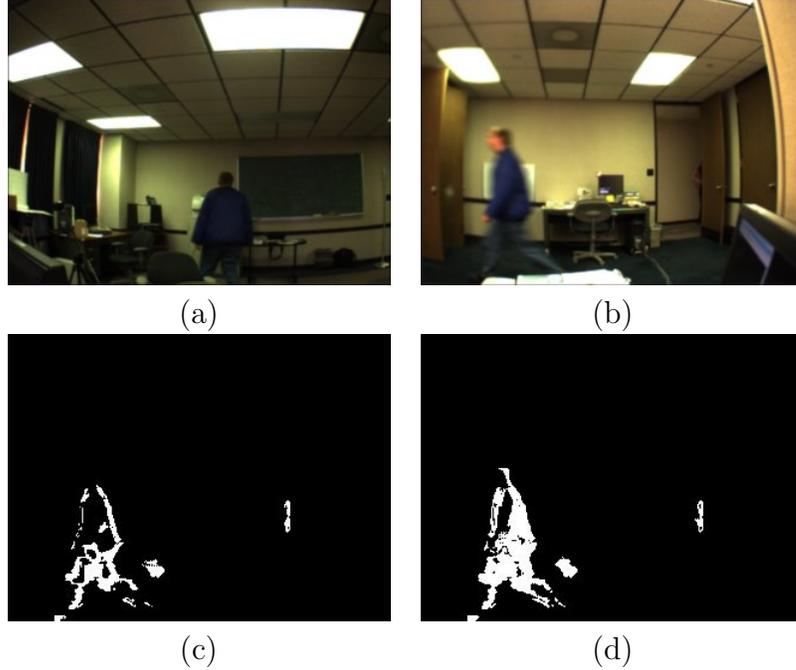
Figure 3.1: A typical camera transition from camera 3 to camera 5 shown in images (a) and (b) respectively. (c) Classification result when all pixels in motion from camera 5 are compared directly to appearance model derived in camera 3. White pixels correspond to chromaticity samples within a Mahalanobis distance of less than 1.5. (d) By first applying the color transfer function from camera 3 to camera 5, classification results improve dramatically.

event for subject B. The dramatically different illumination and pose leads to pixel classification results (Fig. 3.2(c)) of only 29% if the appearance model from Fig. 3.2(a) is directly applied to the pixels in Fig. 3.2(b). A significant improvement (Fig. 3.2(d)) to 84% true positives results when the appearance model is first transformed using the color transfer function.

Fig. 3.2 shows why recognition and classification rates improve significantly when using the color transfer functions. Two views of the same subject lead to dramatically different color models (of three components each), shown in Fig. 3.2(c) and (d). Using the recovered deformation parameters, model in Fig. 3.2(c) is aligned with the model from the view of Fig. 3.2(b) (shown in Fig. 3.2(d)). Geometrically the mean μ gives center of ellipse and the semi-axes are $\sqrt{\lambda_i}\,\mathbf{v}_i$ where $\mathbf{v}_i$ are orthogonal eigenvectors of

covariance Σ with eigenvalues λ_i, for all i columns. Classification results both without applying the color transfer model and applying the model are shown in Fig. 3.2(e) and Fig. 3.2(f) respectively.

(a)

(b)

(c)

(d)

(e)

(f)
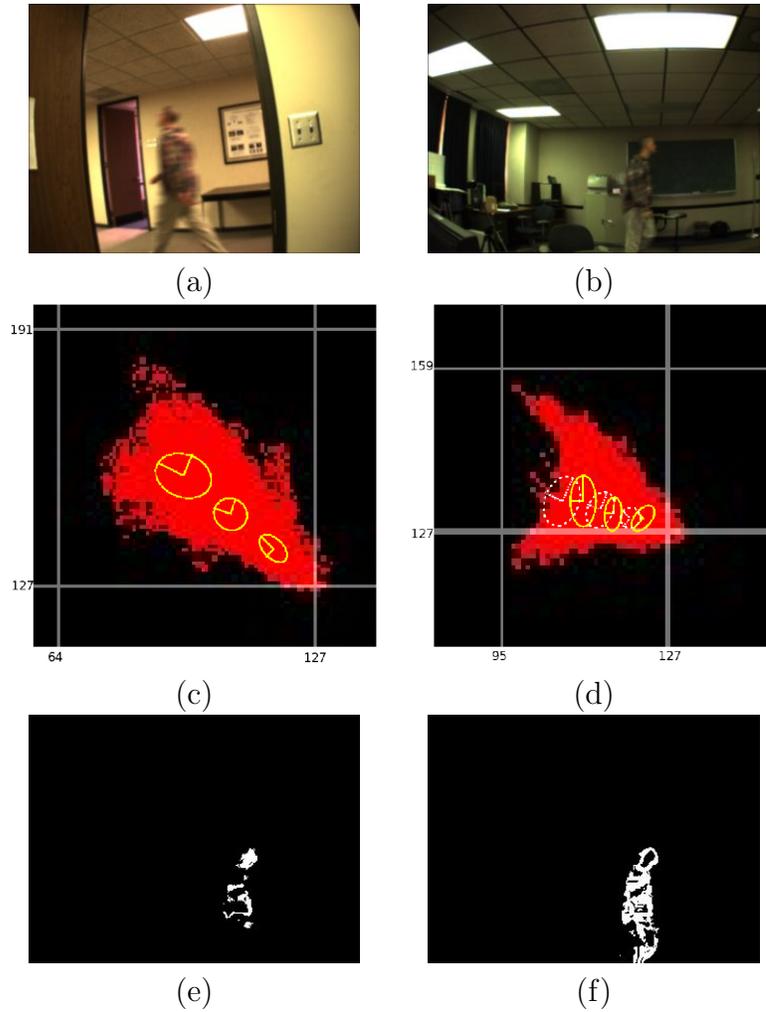
Figure 3.2:   Result of two views of subject F under dramatically different illumination conditions. Subject F's transition is shown in (a) and (b). (c) and (d) display enlarged sections of chromaticity samples of the subject shown in (a) and (b) respectively. Note that the different axis scales are used for clarity. Solid ellipses represent the components of the color model. In (d), dotted ellipses depict the recovered deformation parameters of the model shown in (c). (e) and (f) are the pixel classification result without/with transforming the model in view (a) respectively.

3.3 Direct comparison to other techniques

The technique was directly compared against color histogram-based techniques as well as object matching with and without applying the color transfer function. By varying these different approaches slightly, six different baseline object matching schemes that involve different levels of color matching complexity, were generated. These baseline techniques were then compared against the method proposed here. For these experiments, 190 object models were used.

For each baseline variation, we additionally apply an RGB diagonal color correction to the raw pixel values. In order to obtain each channel scale factor between a camera pair, we first compute the mean values of each channel of object's pixel values that were tracked in each camera view. With known ground-truth matching information of each object, we stack the mean values of all objects from one camera to the other into a measurement matrix $\mathbf{M}$ as follows:

$$\mathbf{M} = \begin{bmatrix} x_1 & -x_1' \\ \vdots & \vdots \\ x_n & -x_n' \end{bmatrix} \tag{3.2}$$

where the x and x' are the mean channel values of each object and n is the total number of objects observed for a given camera pair. Each channel ratio between the camera pair is the solution of a least squares problem, whose solution can be found using SVD, taking the last column vector of the right orthogonal matrix of $\mathbf{M}$ and normalizing the column vector by dividing it by its last component.

For the color histogram methods, we have tested two different color spaces. The first is a 3-dimensional histogram that is built from the RGB values corrected by the RGB diagonal color corrections model and the second is a 2-dimensional histogram corresponding to the UV chromaticity values transformed from the RGB values corrected by the RGB diagonal color correction model. From our experiments, the 2-dimensional UV color histogram performed better than the 3-dimensional RGB color histogram in terms of the space and time complexities as well as matching

performance. This may be because the brightness component (discarded in the UV-space) is more sensitive to lighting variation from shadows and the relative distance of the tracked object to the light source. Therefore, the results here compare the 2-dimensional UV (32×32) color histogram to the method proposed in this paper. For the comparison, the red chrominance (U) and blue chrominance (V) bins were divided into 32 and 32 sections respectively.

Fig. 3.3 shows the performance of the different object matching methods. The graph represents the true positive rates (y axis) of each method as the number of objects (x axis) were increased. The solid and dashed curves represent cases where the RGB diagonal color correction is applied and not applied respectively. The red (square) and green (diamond) curves are matching pipelines that use the Gaussian mixture model matching method with/without applying our color transfer function respectively. The blue (circle) curves correspond to the color histogram matching methods.

The effect of the RGB diagonal color correction rarely improved matching performance when using the GMM method in conjunction with our color transfer function. However, it can help improve results if a more sophisticated color transfer model (i.e. the one presented here) is unavailable.

In order to demonstrate the robustness of each method, the true positive rate was measured as the number of objects to be matched were increased. It should be noted that the matching rates were computed by counting correct matches based on the ground-truth information where a match for a given query object is defined as the nearest model in the stored models during runtime. For the GMM methods the matching scheme shown in Eq.(3.1) was used, while color histogram methods utilized the Bhattacharyya distance.

Since a match to be applied to the true positive rate is counted when a single nearest model retrieved from the stored models correctly matches to a queried object,
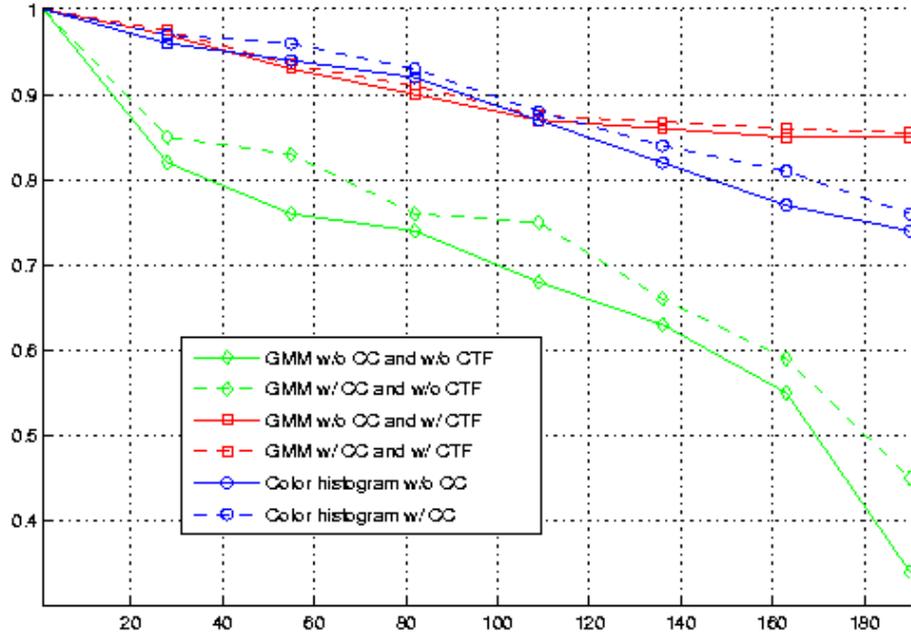
Figure 3.3: Performance comparison of object different matching methods as compared to the approach introduced here. Graph shows the true positive rates (y axis) versus the number of objects (x axis). The solid and dashed curves denote with/without the RGB diagonal color correction (CC) respectively. CTF stands for the color transfer function method.

it is not surprising that the true positive rates are decreasing relatively fast as the number of objects are added in the model database. However, one would hope that a recognition scheme that utilizes color transfer should still perform well as the number of objects in the database becomes large.

It can be seen that matching methods using the color histogram showed better performance when only the number of objects are small but as the number of increases in the object model database the methods started breaking down. On the other hand, the GMM methods using the color transfer function was resilient from the event of increasing the number of objects. Quantitatively, the true positive rates of our method was more than 11% higher than the more straightforward color histogram matching methods, while average performance degradation rate of our method and the color histogram were measured as 1.87% and 3.25% respectively. This means that the color

transfer method is almost half as affected from as a color histogram matching scheme as the number of objects are increased.

Finally, we studied the separability of the objects in the database using the color transfer measure and a the GMM matching scheme. Overall, the method should be capable of reducing the appearance differences measured by similarity distance. In order to measure this, the Φ distance between each object and itself across all cameras was measured using the ground-truth object transition information provided by the dataset. Fig. 3.4 plots the mean Φ distance computed for each object with and without the color transfer.
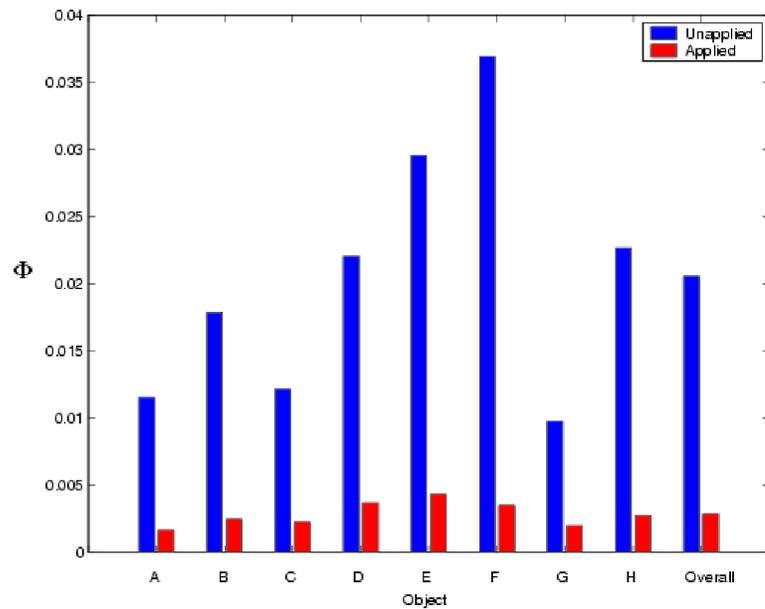

Figure 3.4: The Φ error measure between each object and itself (based on ground-truth information) across all cameras with and without the color transfer function.

Chapter 4

Discussion

Most existing computational color constancy algorithms attempt to recover chromaticity information about the illumination in an input scene for which the only known information is the sensor responses for each point (or color patch) in that scene. In contrast to the color constancy work, our goal was to show how well the color transfer method would do on an typical vision task, namely, consistent object tracking in non-overlapping multi-camera video surveillance. In the multi-camera video surveillance, the sensor response for point or patch is not generally known in advance, rather it must be recovered by extracting pixels on the object as it moves from one view to the next. Furthermore, change of object pose, size, and spectral reflectance are likely as the object is observed in different cameras. These problems make it nearly impossible to reliable correspond single color patches from the same point on the object across views. We have shown that, instead of utilizing a color calibration method that requires these correspondences, a more general model of color transfer that relates the entire extracted color model in each view is feasible.

Due to the complexity of the problem and domain constraints, it is difficult to directly compare the new approach to existing techniques that may require direct pixel color correspondence. However, a few comments about similar methods and our results can be made. In important and recent work, Javed et al. [29] have proposed a brightness transfer function that is acquired across multiple non-overlapping views, and run detailed experiments in three different multi-camera scenarios. The validity of their subspace based method was demonstrated by using appearance combined with supplements of known camera topology where the space-time information of entry and exit locations, exit velocity and inter-camera time interval are available in prior. We

summarize their matching results of the three video scenarios for schemes they used. The space-time scheme achieved a matching rate of 83%, 93% and 88%. When only the brightness transfer function subspace was used, 88%, 90% and 81% were obtained. As expected, the joint scheme showed the best results of 94%, 100% and 96%. These performance results are similar to the ones achieved by the method proposed here. However, the tracking schemes are quite different in that known camera topology information was exploited in the maximum a posteriori estimation framework. This can be a considerable advantage in order to increase the accuracies in tracking but is quite constraining in a typical large-scale video surveillance network. However, it is clear from the results that both the appearance and space-time models are important sources of information as the tracking results improve significantly when both the models are used jointly.

Other work [11] presents a novel solution to the inter-camera color calibration problem, which is very important for multi-camera systems. This work considers radiometric properties using a non-parametric function to model color distortion for pair-wise camera combinations. A correlation matrix is computed from three 1-D color histograms, and the model function is obtained from a minimum cost path traced within the matrix. The first experiments was conducted with several synthesized image pairs. Each pair consists of a reference image and a distorted version of its illumination histogram. The histogram distortions were random, non-linear and non-parametric. After the cross-correlation matrix and the model function are computed, the histogram of the distorted image accordingly is transformed to obtained the illumination corrected image. The improvement was substantial even though histogram operations are invariant to spatial transformations but unfortunately the results were not reported quantitatively. In a second experiment, images were acquired under different lighting conditions. Since each image was taken at a different time, there are appearance mismatches in addition to the lighting and the camera difference. They

computed the aggregated cross-correlation matrices for each color channel from 150 pairs. It would be interesting to apply this technique to the challenging domain of video surveillance and directly compare recognition rates (and degradation) to the proposed approach. In particular, correlation between models (or histograms) must to yield separability between objects in the database even as models are degraded.

There is a significant amount of work related to understanding the physics of light change and determining how to predict and correct these changes. Much of this work cannot be directly compared to our method because it requires point-wise correspondence. However, some techniques utilize a diagonal transfer model[26, 12] of illumination change that can be acquired under the same conditions as the work here. The diagonal model simply describes the effect of moving from one scene illuminant to another by scaling the R, G, and B channels via independent scale factors. These scale factors can be written as the elements of a diagonal matrix. Previous work [20, 19] has shown that the diagonal model works with the particular type of sensors which have relatively narrow band and non-overlapping sensitivity functions under typical scene illumination. One thus might claim that a simpler approach that relates RGB space (as opposed to our UV approach) is linear and simple to compute. This is simply not the case in the multi-camera video surveillance. As addressed factors above, it is non-trivial to find corresponding pixel locations of the same object that will be the cue to estimate the change in illumination. Moreover it is well known from the illumination cone work of Belhumeur and Kriegman [27] to except for the case of a planar object seen by a fixed camera and illuminated by point light sources at infinity, RGB values of the object seen under different illuminations cannot be modeled by a simple diagonal scaling.

The approach using a log chromaticity domain by Berwick and Lee [8] has the advantage that it implies invariance under pure rotation. While we acknowledge if a single cluster is centered at the origin this approach is useful, this work has shown

that object appearance in the UV space under different imaging conditions cannot be modeled by a simple rotation. Rather, the appearance model of the object includes multiple clusters as well as scaling. In this case, transformation to the log-polar domain does not yield a rotational matching scheme.

Chapter 5

Conclusions and Future Work

In this work, a technique that estimates the unknown color transfer function between pairs of disjoint cameras was introduced. Object appearance model correspondences, generated as a subject moves throughout a surveillance network are provided as input to a robust estimation procedure. Transfer functions are then able to predict appearance change from one camera to the next and have been shown to increase recognition capabilities of even simple object matching schemes to above 80% in an indoor multi-camera network.

The technique was compared to a variety of baseline approaches of varying complexity and was shown to outperform them. In particular, the technique degrades more slowly than competing techniques as the number of objects to be matched across views is increased. Although the method cannot be directly compared to several state-of-the-art approaches in color calibration, either because these methods require different operational conditions, or data is simply not available. Many of the more related techniques were discussed in the context of our performance studies and advantages and disadvantages of each technique was discussed.

5.1   Future Work

We have demonstrated the validity of the technique using color (rather than intensity) to achieve matching stability of the recovered models. However, the method still relies on several assumptions of varying complexity. Violation of these assumptions can lead to degrade the robustness of the method. Future work as extensions to the current method can be motivated by removing one or more of these assumptions.

Automatic detection and removal algorithms of self-shadows and non-isotropically

colored objects would be important in order to obtain more consistent appearance models across difference camera views

Although we have emphasized the importance of using color to achieve stability of the recovered models over time, changes in illuminate color, relative pose of the cameras to the scene, or intrinsics, will require re-estimation of color transfer functions online. (for example, sudden illumination changes caused by turning on and off lights, adding lights, or clouds occluding sunlight). As future work, we are exploring online incremental learning methods of the functions that will make this feasible.

Bibliography

[1] A.Dempster, N.Laird, and D.Rubin. Maximum-likelihood from incomplete data via the em algorithm. *J. Royal Statist. Soc. Series B 39*, 1977.

[2] A.Elgammal, R.Duraiswami, D.Harwood, and L.Davis. Background and foreground modeling using nonparametric kernel density for visual surveillance. In *Proc. the IEEE*, 2002.

[3] A.Mittal and L.Davis. Unified multi-camera detection and tracking using region-matching. In *Proc. IEEE Workshop on Multi-Object Tracking*, 2001.

[4] A.Rahimi, B.Dunagan, and T.Darrell. Simultaneous calibration and tracking with a network of non-overlapping sensors. In *Proc. Computer Vision and Pattern Recognition*, volume 1, pages I–187– I–194, Jun 2004.

[5] C.Jaynes, A.Kale, N.Sanders, and E.Grossmann. The terrascope dataset: A scripted multi-camera indoor video surveillance dataset with ground-truth. In *Proc. the IEEE Workshop on VS PETS*, October 2005.

[6] C.Stauffer, W.Eric, and L.Grimson. Learning patterns of activity using real-time tracking. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2000.

[7] C.Wren, A.Azarbayejani, T.Darrell, and A.Pentland. Pfinder: Real-time tracking of the human body. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1997.

[8] D.Berwick and S.Lee. A chromaticity space for specularity-, illumination color- and illumination pose-invariant 3-d object recognition. In *ICCV*, pages 165–170, 1998.

[9] D.H.Johnson and G.Orsak. Relation of signal set choice to the performance of optimal non-gaussian detectors. *IEEE Trans. Comm.*, 41(9):1319–1328, Sep 1993.

[10] F.M.Porikli and A.Divakaran. Multi-camera calibration, object tracking and query generation. In *Proc. IEEE International Conference on Multimedia and Expo (ICME)*, volume 1, pages 653–656, Jul 2003.

[11] F.Porikli. Inter-camera color calibration using cross-correlation model function. In *Proc. IEEE Int. Conf. on Image Processing*, 2003.

[12] G.D.Finlayson, B.Schiele, and J.L.Crowley. Comprehensive colour image normalization. In *Proc. European Conference on Computer Vision*, Jul 1998.

[13] Kideog Jeong and C.Jaynes. Object matching in disjoint cameras using a color transfer approach. *To Appear in Machine Vision and Applications Journal*, 2006.

[14] J.Kang, I.Cohen, and G.Medioni. Persistent objects tracking across multiple non overlapping cameras. In *Proc. IEEE wacv-motion 2005*, volume 2 No.2, 2005.

[15] J.M.Speigle and D.H.Brainard. Predicting color from gray: The relationship between achromatic adjustment and asymmetric matching. *J. Opt. Soc. Am. A*, 16, No.10:2370–2376, Oct 1999.

[16] J.Orwell, P.Remagnino, and G.A.Jones. Multi-camera color tracking. In *Proc. the Second IEEE Workshop on Visual Surveillance*, 1999.

[17] J.S.Werner and B.E.Schefrin. Loci of achromatic points throughtout the life span. *J. Opt. Soc. Am. A*, 10, No.7:1509–1516, Jul 1993.

[18] J.Walraven and J.S.Werner. The invariance of unique white; a possible implication for normalizing cone action spectra. *Vision Res.*, 31:2185–2193, 1991.

[19] K.Barnard, L.Martin, A.Coath, and B.Funt. A comparison of color constancy algorithms. part one: Methodology and experiments with synthesized data. *IEEE Transactions in Image Processing*, 11(9):972–984, 2002.

[20] K.Barnard, V.Cardei, and B.Funt. A comparison of color constancy algorithms. part two: Experiments with image data. *IEEE Transactions in Image Processing*, 11(9):985–996, 2002.

[21] K.Bauml. Color appearance: Effects of illuminant changes under different surface collections. *J. Opt. Soc. Am. A*, 11, No.2:531–542, Feb 1994.

[22] K.Jeong and C.Jaynes. Moving shadow detection using a combined geometric and color classification approach. In *IEEE MOTION*, 2005.

[23] M.D.Grossberg and S.K.Nayar. Determining the camera response from images: What is knowable? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(11):1455–1467, Nov 2003.

[24] M.Grossberg and S.K.Nayar. Modeling the space of camera response functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(10):1272–1282, Oct 2004.

[25] M.J.Swain and D.H.Ballard. Color indexing. *International Journal of Computer Vision*, 7:11–32, 1991.

[26] M.S.Drew, J.Wei, and Z.Li. Illumination-invariant color object recognition via compressed chromaticity histograms of color-channel-normalized images. In *Computer Vision, 1998. Sixth International Conference*, pages 533–540, Jan 1998.

[27] N.Belhumeur and D.Kriegman. What is the set of images of an object under all possible illumination conditions? *Int. Journal of Computer Vision*, 28(3):245–260, 1998.

[28] N.Sanders and C.Jaynes. Class-specific color camera calibration with application to object recognition. In *IEEE WACV*, 2005.

[29] O.Javed, K.Shafique, and M.Shah. Appearance modeling for tracking in multiple non-overlapping cameras. In *Proc. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20–26, Jun 2005.

[30] P.J.Rousseeuw and A.M.Leroy. *Robust Regression and Outlier Detection*. John Wiley and Sons, New York, 1987.

[31] Q.Xiong and C.Jaynes. Multi-resolution background modeling of dynamic scenes using weighted match filters. In *Proc. the ACM 2nd international workshop on Video surveillance & sensor networks (VSSN '04)*, pages 88–96, New York, NY, USA, 2004. ACM Press.

[32] R.Basri and D.W.Jacobs. Lambertian reflectance and linear subspaces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25 no. 2:218–233, Feb 2003.

[33] R.Gross, J.Yang, and A.Waibel. Growing gaussian mixture models for pose invariant face recognition. In *Proc. the 15th International Conference on Pattern Recognition*, volume 1, pages 1088 – 1091, Sep 2000.

[34] Ying Shan, Harpreet S. Sawhney, and Rakesh Kumar. Vehicle identification between non-overlapping cameras without direct feature matching. *iccv*, 1:378–385, 2005.

[35] Ying Shan, Harpreet S. Sawhney, and Rakesh (Teddy) Kumar. Unsupervised learning of discriminative edge measures for vehicle matching between non-overlapping cameras. *cvpr*, 1:894–901, 2005.

[36] S.Kullback and R.A.Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, March 1951.

[37] S.McKenna, Y.Raja, and S.Gong. Tracking colour objects using adaptive mixture models. *Image and Vision Computing*, pages 225–231, 1999.

[38] T.M.Cover and J.A.Thomas. *Elements of Information Theory*. Wiley, New York, 1991.

Background

- Date of Birth: February 28th, 1971

- Place of Birth: Busan (or Pusan), South Korea

Education

- Bachelor of Science in Computer Science, Chongju University, Chongju, South Korea, 1994

Professional Experience

- September 2004 - Present. Research Assistant, Metaverse Lab, Center for Visualization and Virtual Environments.

- January 2003 - May 2004. Teaching Assistant, Department of Computer Science, University of Kentucky.

- January 1997 - June 2002. SAP R/3 Basis Engineer, TriGem Information Consulting, Inc.

- January 1994 - January 1997. Oracle Database Application Developer, TriGem Computer, Inc. (www.trigem.co.kr)

Publications

- K. Jeong and C. Jaynes, "Object Matching in Disjoint Cameras using a Color Transfer Approach," To Appear in Machine Vision and Applications Journal, 2006, Springer Berlin Heidelberg, ISSN: 0932-8092

- K. Jeong and C. Jaynes, "Moving Shadow Detection using a Combined Geometric and Color Classification Approach," IEEE Workshop on Motion and Video Computing (WACV/MOTION'05), Breckenridge, CO, January 5-7 2005.