A Target Dependent Colorspace for Robust Tracking

Francesc Moreno-Noguer and Alberto Sanfeliu
Institut de Robòtica i Informàtica, UPC-CSIC
Llorens Artigas 4-6, 08028, Barcelona, Spain
fmoreno, asanfeliu@iri.upc.edu

Dimitris Samaras
Computer Science Department
SUNY Stony Brook, NY 11794-4400, USA
samaras@cs.sunysb.edu

Abstract

The selection of the appropriate colorspace for tracking applications has not been an issue previously considered in the literature. Many color representations have been suggested, based on the invariance to illumination changes. Nevertheless, none of them is invariant enough to deal with general and unconstrained environments. In tracking tasks, we might prefer to represent image pixels into a colorspace where the distance between the target and background colorpoints were maximized, simplifying the task of the tracker. Based on this criterion, we propose an 'object dependent' colorspace, which is computed as a simple calibration procedure before tracking. Furthermore, this colorspace may be easily adapted at each frame. Synthetic and real experiments show how this colorspace allows for a better discrimination of the foreground and background, and permits to track in circumstances where the same tracking algorithm relying on other colorspaces would fail. [1]

1 Introduction

An important initial issue for any color-based tracking algorithm, concerns to the selection of the colorspace where the data is going to be represented. Many colorspaces have been proposed in the literature: for instance RGB, HSV, XYZ, YUV, YIQ, and its corresponding normalized versions. Unfortunately, there is not a clear consensus about which one to use, and the different alternatives have been indistinctly utilized: The *RGB* colorspace is used in [6,9]. The efforts described in [1,10] represent color by the normalized RGB model. Maybe the most extensively used colorspace is the *HSV* [8,11,13], and a two dimensional version considering only the *HS* components [3,7]. Some approaches [4,12] are based on the *YUV* as well.

All this variety indicates that there is not a criterion for the selection of the appropriate colorspace. In most of the previously cited approaches, the selection is based on a trial and error procedure among the various available colorspaces. In other circumstances, the selection is based on the invariance of the color representation to illumination changes, and normalized or colorspaces where chromaticity is separated from intensity (such as HSV, YUV or YIQ) are utilized. Nevertheless, none of the existing colorspaces is robust enough to deal with real and general environments.

In contrast to previous approaches, we propose to select the colorspace using the following criteria focused on visual tracking applications: **1)** Since the first goal in tracking is to discriminate the object of interest from the rest of the scene, an important function of the colorspace should be to maximize the separation between the target and background color points. **2)** Furthermore, in order to deal with dynamic environments, the colorspace should demonstrate a certain degree of invariance to illumination changes. Alternatively, the representation of the colorspace should permit an easy adaptation to these changing conditions.

In this work we suggest to use an 'object dependent' colorspace which satisfies both previous criteria. The colorspace, which we call *Fisher colorspace*, is computed as a simple calibration procedure before tracking, based on the *Linear Discriminant Analysis* (LDA) [2]. We prove, by a set of synthetic and real experiments, that the Fisher colorspace provides better representations of the data, in terms of fore/background separability, than other existent colorspaces. We also show that this colorspace is invariant to uniform scaling and shifting of the illumination, and that its simple parameterization may be easily updated throughout time, thus becoming an adaptable colorspace.

2 Fisher colorspace

The representation of the target and background through a color model maximizing the separation of both classes (criterion 1) may be analyzed as a standard classification problem based on *Discriminant Analysis*. We are interested in the linear techniques, since as will be shown in the following subsection, this offers certain robustness to illumination changes (criterion 2). Therefore, using LDA [2], the problem may be reduced to the search of the hyperplane (*Fisher plane*) that best separates both classes.

The Fisher colorspace may be computed from a single *RGB* training image, where the points belonging to the object and background are manually identified. Assume that

[1] This work was supported by CICYT project DPI2004-05414 from the Spanish Ministry of Science and Technology.

the set of n image pixels are arranged into a $n\times 3$ matrix $\mathbf{C}=[\mathbf{c}_1,\ldots,\mathbf{c}_n]^T$. $n_{\mathcal{O}}$ of these pixels belong to the object $\mathcal{O}$, represented by $\mathbf{C}_{\mathcal{O}}=[\mathbf{c}_{\mathcal{O},1},\ldots,\mathbf{c}_{\mathcal{O},n_{\mathcal{O}}}]^T$ and the rest of $n_{\mathcal{B}}$ pixels $\mathbf{C}_{\mathcal{B}}=[\mathbf{c}_{\mathcal{B},1},\ldots,\mathbf{c}_{\mathcal{B},n_{\mathcal{B}}}]^T$ belong to the background $\mathcal{B}$. We wish to determine which plane is the most effective in discriminating between these two subsets of points. Let us denote such a plane by $\mathbf{W}=[\mathbf{w}_1,\mathbf{w}_2]^T\in\mathbb{R}_{2\times 3}$, where $\mathbf{w}_1$ and $\mathbf{w}_2$ are vectors in the *RGB* space, spanning the points lying on the plane. The projection of $\mathbf{C}_{\mathcal{O}}$ and $\mathbf{C}_{\mathcal{B}}$ onto this plane, generates the sets $\mathbf{F}_{\mathcal{O}}=\mathbf{C}_{\mathcal{O}}\mathbf{W}^T\in\mathbb{R}_{n_{\mathcal{O}}\times 2}$ and $\mathbf{F}_{\mathcal{B}}=\mathbf{C}_{\mathcal{B}}\mathbf{W}^T\in\mathbb{R}_{n_{\mathcal{B}}\times 2}$, respectively.

The goal of the LDA is to find the best orientation of the plane $\mathbf{W}$, such that the separation of the projected subsets $\mathbf{F}_{\mathcal{O}}$ and $\mathbf{F}_{\mathcal{B}}$ is maximized. In order to determine such a plane, LDA considers the maximization of the objective function $J(\mathbf{W})=(\mathbf{W}\mathbf{S}_b\mathbf{W}^T)(\mathbf{W}\mathbf{S}_w\mathbf{W}^T)^{-1}$ where $\mathbf{S}_w$ is the *within class scatter matrix* and $\mathbf{S}_b$ is the *between class scatter matrix*:

$$\mathbf{S}_w = \sum_{\varepsilon=\{\mathcal{O},\mathcal{B}\}}\frac{n_\varepsilon}{n}\sum_{i=1}^{n_\varepsilon}(\mathbf{c}_{\varepsilon,i}-\bar{\mathbf{c}}_\varepsilon)(\mathbf{c}_{\varepsilon,i}-\bar{\mathbf{c}}_\varepsilon)^T \quad (1)$$

$$\mathbf{S}_b = \sum_{\varepsilon=\{\mathcal{O},\mathcal{B}\}} n_\varepsilon(\bar{\mathbf{c}}_\varepsilon-\bar{\mathbf{c}})(\bar{\mathbf{c}}_\varepsilon-\bar{\mathbf{c}})^T \quad (2)$$

with, $\bar{\mathbf{c}}_\varepsilon=\frac{1}{n_\varepsilon}\sum_{i=1}^{n_\varepsilon}\mathbf{c}_{\varepsilon,i}$ being the ε-class mean vector, and $\bar{\mathbf{c}}=\sum_{\varepsilon=\{\mathcal{O},\mathcal{B}\}}\frac{n_\varepsilon}{n}\bar{\mathbf{c}}_\varepsilon$ the total mean vector.

The classic LDA method maximizes the J objective function by constructing the rows of $\mathbf{W}$ with the eigenvectors of $\mathbf{S}_w^{-1}\mathbf{S}_b$ having the highest eigenvalues. Nevertheless, this approach has a limitation, since $\mathbf{S}_w^{-1}\mathbf{S}_b$ is not a full rank matrix, and in the two class problem discussed here, the hyperplane $\mathbf{W}$ would be a line. Although we are interested in projecting the data onto a linear space (to increase robustness to illumination changes), by projecting the data onto a line we might lose too much information. A better choice, consists of projecting the *RGB* data onto a plane. Therefore, we will use the nonparametric version of LDA [2]. The key point of this LDA extension is that it computes the between class scatter matrix $\mathbf{S}_b$ using local information and the *K-Nearest Neighbors* (KNN) rule, which allows to obtain a full rank $\mathbf{S}_w^{-1}\mathbf{S}_b$ matrix. For further details about the nonparametric LDA, the reader is referred to [2].

Fig. 1 depicts the main steps to compute the Fisher plane, for two different targets in the same image. Fig. 1a shows the scene and in Fig. 1b all the image pixels are represented in the *RGB* colorspace. Note that the Fisher colorspace is 'target dependent'. For instance, observe in Fig. 1c the difference of the Fisher planes obtained for different targets: the ladybird and the flower petals. For example, if we wish to track the ladybird (Fig. 1d), the points belonging to the object and background are initially provided (Fig. 1e), and the Fisher plane is computed based on the non-parametric LDA (Fig. 1f). Similar stages are depicted in Fig. 1g-i for tracking the flower petals.

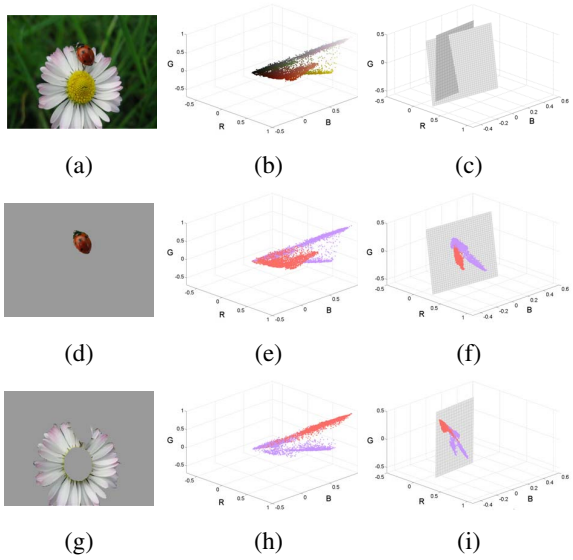


Figure 1. Computing the Fisher plane for two different targets.

2.1 Fisher colorspace in the presence of lighting changes

Next, we will see that the Fisher colorspace is invariant to certain lighting effects, such as uniform scaling and shifting of illumination. Without loss of generality, we will assume the parametric version of the LDA, described by the maximization of the objective function $J(\mathbf{W})=(\mathbf{W}\mathbf{S}_b\mathbf{W}^T)(\mathbf{W}\mathbf{S}_w\mathbf{W}^T)^{-1}$ where $\mathbf{S}_w$ and $\mathbf{S}_b$ are defined by Eqs. 1 and 2, respectively.

Lemma 1 *The Fisher plane is invariant to a uniform illumination scaling*

Proof: Given all image points $\mathbf{c}_i, i=1,\ldots,n$ represented in the *RGB* colorspace, a uniform illumination scaling is defined by the mapping $\mathcal{S}$: $\mathbf{c}_i\rightarrow\alpha\mathbf{c}_i$, where α is the scaling factor.

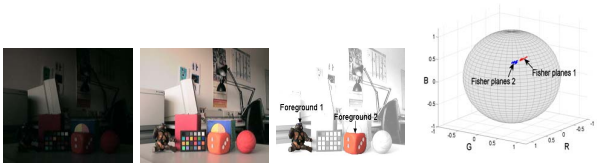
We assume that the classification of the image points into the object ($\mathcal{O}$) and background ($\mathcal{B}$) classes is available. Under these circumstances, the following statements about the total mean and the class means are satisfied:

$$\mathcal{S}(\bar{\mathbf{c}}_\varepsilon) = \mathcal{S}\left(\frac{1}{n_\varepsilon}\sum_{i=1}^{n_\varepsilon}\mathbf{c}_{\varepsilon,i}\right)=\frac{1}{n_\varepsilon}\sum_{i=1}^{n_\varepsilon}\alpha\mathbf{c}_{\varepsilon,i}=\alpha\bar{\mathbf{c}}_\varepsilon$$

$$\mathcal{S}(\bar{\mathbf{c}}) = \mathcal{S}\left(\sum_{\varepsilon=\{\mathcal{O},\mathcal{B}\}}\frac{n_\varepsilon}{n}\bar{\mathbf{c}}_\varepsilon\right)=\sum_{\varepsilon=\{\mathcal{O},\mathcal{B}\}}\frac{n_\varepsilon}{n}\mathcal{S}(\bar{\mathbf{c}}_\varepsilon)=\alpha\bar{\mathbf{c}}$$

As a consequence, the transformed within class scatter matrix and between class scatter matrix may be written as

$$\mathcal{S}(\mathbf{S}_w)=\alpha^2\mathbf{S}_w \qquad \mathcal{S}(\mathbf{S}_b)=\alpha^2\mathbf{S}_b$$



(a) (b) (c) (d)

Figure 2. Fisher colorspace & lighting changes.

Finally, the objective function to be maximized is:

$$\mathcal{S}(J(\mathbf{W})) = \mathcal{S}\left(\frac{\mathbf{W}\mathbf{S}_b\mathbf{W}^T}{\mathbf{W}\mathbf{S}_w\mathbf{W}^T}\right) = \frac{\mathbf{W}\mathcal{S}(\mathbf{S}_b)\mathbf{W}^T}{\mathbf{W}\mathcal{S}(\mathbf{S}_w)\mathbf{W}^T} = J(\mathbf{W})$$

Thus, the criteria used to compute the Fisher plane for two images related by a linear scaling are exactly the same. Therefore, we conclude that the Fisher plane is invariant to a uniform illumination scaling. □

Lemma 2 *The Fisher plane is invariant to uniform illumination shifting*

Proof: Given all image points $\mathbf{c}_i, i = 1, \ldots, n$ represented in the *RGB* colorspace, a uniform lighting shifting is defined by the mapping $\mathcal{T}$: $\mathbf{c}_i \rightarrow \mathbf{c}_i + \beta$, where β is the shifting factor.

Following a similar procedure than for Lemma 1, it can be shown that $\mathcal{T}(J(\mathbf{W})) = J(\mathbf{W})$, i.e, the Fisher plane is also invariant to a illumination shifting effect. □

3 Evaluation of the Fisher colorspace

3.1 Invariance to illumination

The invariance of the Fisher plane to illumination changes is demonstrated in the following experiment, where the images of a still scenario illuminated by natural lighting have been acquired during a whole day (Figs. 2(a,b) show two representative frames).

Several foreground objects have been selected, and for each of them, the Fisher plane has been computed throughout the whole sequence. The results show that the Fisher planes (represented by their normal vectors) form separate clusters for every individual target, and the variance in each cluster is relatively small, proving that the Fisher colorspace is quite invariant to illumination changes. In Fig. 2d we depict the distribution of the normals to the Fisher plane for two different targets (indicated in Fig. 2c). The unitary sphere represents the space of all possible configurations of the normal to the Fisher plane. Observe how for each target, the Fisher plane distributions just occupy a small region onto the configuration space.

3.2 Fisher *versus* other colorspaces

Next, we will compare the performance of the proposed Fisher colorspace, versus other commonly used colorspaces. In order to make a fair comparison we will consider only those colorspaces defined by two variables, such as the combination of two components of the normalized *RGB* (namely *rg*, *rb* and *gb*), two components of the normalized *XYZ* (*xy*, *xz*, and *yz*), and two components of the *HSV* colorspace.

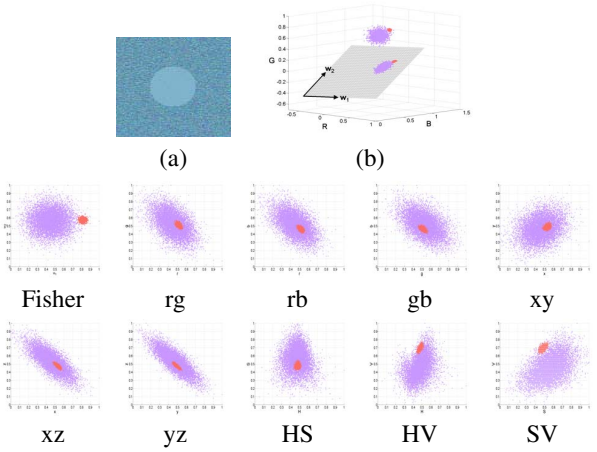


Figure 3. Performance of the Fisher colorspace for camouflaging targets.

The first experiment shows an example where the foreground is in some degree camouflaged with the background (Fig. 3a shows the test image, where the central circle is the target). In these circumstances, the Fisher colorspace clearly performs better than other colorspaces. Fig. 3b shows the color distributions of all image points into the RGB colorspace. Blue dots correspond to the background points, and red dots are the target color points. Note that both classes are in close contact. In spite of this, the projection of the colorpoints onto the Fisher plane does not overlap the target and background classes. On the other hand, when we represent the points using the other colorspaces, the two classes are greatly overlapping, which will cause difficulties in future tracking tasks.

A more precise comparison is obtained by computing the distance between the target and the background representations. Given the set of object color points $\mathbf{F}_\mathcal{O} = \{\mathbf{f}_{\mathcal{O},1}, \ldots, \mathbf{f}_{\mathcal{O},n_\mathcal{O}}\}$ and the set of background color points $\mathbf{F}_\mathcal{B} = \{\mathbf{f}_{\mathcal{B},1}, \ldots, \mathbf{f}_{\mathcal{B},n_\mathcal{B}}\}$ we compute the distance between both sets with the following metric (which is in accordance with the nonparametric LDA):

$$\mathrm{dist}(\mathbf{F}_\mathcal{O}, \mathbf{F}_\mathcal{B}) = \frac{\frac{1}{n_\mathcal{O}}\sum_{i=1}^{n_\mathcal{O}}\frac{1}{k}\sum_{j=1}^{k}\|\mathbf{f}_{\mathcal{O},i} - \mathcal{NN}^j_\mathcal{B}(\mathbf{f}_{\mathcal{O},i})\|^2}{|\det(\mathbf{S}_w)|}$$

where $\mathcal{NN}^j_\mathcal{B}(\mathbf{f}_{\mathcal{O},i})$ is the j-th nearest neighbor in the set $\mathbf{F}_\mathcal{B}$ to a point $\mathbf{f}_{\mathcal{O},i}$, $\|\cdot\|$ is the Euclidean norm, and $|\cdot|$ the absolute value function. Using this metric, we have computed the fore/background distance for six different targets. The results are shown in Table 1. Note that in most of the cases, the Fisher colorspace provides the largest separation.

	T1	T2	T3	T4	T5	T6
Fisher	429.1	2445.5	202.1	133.8	9.9	74.2
rg	464.3	214.8	8.2	63.9	2.7	4.8
rb	115.5	219.1	4.0	61.6	1.8	4.9
gb	117.5	183.7	4.5	64.5	2.9	4.8
xy	141.3	275.0	2.8	54.3	1.7	4.5
xz	115.1	303.7	3.5	70.9	1.9	6.8
yz	87.7	274.1	4.7	88.0	4.2	7.9
HS	5.8	5.6	147.7	4.5	3.1	5.1
HV	5.4	1.3	156.1	0.3	5.1	16.4
SV	1.0	16.2	110.3	1.4	5.1	32.8

Table 1. Foreground/Background distances.

3.3 Integrating the Fisher colorspace into a tracking framework

Although we have proved that the Fisher colorspace exhibits a certain robustness to illumination changes, in a real world environment we should consider much more complex artifacts which might modify significantly the configuration of the Fisher plane at each frame. As a consequence, another important property of the proposed colorspace, is the ease with which in can be updated and integrated into a tracking framework.

Based on the work presented in [5], we designed a tracking algorithm, where the target was represented simultaneously by multiple cues. Among these cues, we considered the Fisher plane, parameterized by its normal vector ($\mathbf{n} = \mathbf{w}_1 \times \mathbf{w}_2$), and whose state was continuously updated and estimated using a particle filter formulation.

The basic idea of this approach, consisted of approximating the state of $\mathbf{n}$ by a set of p weighted samples $\{\mathbf{n}_i, \pi_i\}_{i=1}^{p}$, with π_i being the weight associated with sample $\mathbf{n}_i$. The updating procedure, had two stages. At the beginning of a particular iteration, the most likely samples $\{\mathbf{n}_i\}_{i=1}^{p}$ were randomly propagated, to the set $\{\tilde{\mathbf{n}}_i\}_{i=1}^{p}$. Subsequently, these propagated samples were weighted according to $\tilde{\pi}_i \propto \text{dist}(\mathbf{F}_{\mathcal{W},i}, \mathbf{F}_{\bar{\mathcal{W}},i})$, where 'dist' is the metric function previously defined. $\mathbf{F}_{\mathcal{W},i}$ and $\mathbf{F}_{\bar{\mathcal{W}},i}$ are a set of RGB points inside and outside of an image region $\mathcal{W}$, which have been projected on the Fisher plane determined by $\tilde{\mathbf{n}}_i$. Region $\mathcal{W}$ represents an area in the image where the object is expected to be (just a coarse estimation). Finally, the set $\{\tilde{\mathbf{n}}_i, \tilde{\pi}_i\}_{i=1}^{p}$ approximates the state of the Fisher plane, at the end of the iteration.

Using this formulation, tracking was achieved in complex scenes (with abrupt lighting changes and cluttered backgrounds) such as those shown in Fig. 4.

4 Conclusions

In this paper, we have introduced the Fisher colorspace, a color representation of the image points which is appropriate for tracking tasks. The main features of this 'object-dependent' colorspace is that it offers a color representation where the distance between the target and background is maximized. This property is extremely important in order to simplify the task of any tracking algorithm. We have also proved that the Fisher plane offers certain robustness to illumination changes. Furthermore, its simple representation by only a 3D vector parameterizing the normal direction of the plane, permits to integrate the colorspace representation into a more general tracking framework, and conceive it as any other object feature which may be updated throughout time. Both synthetic and real experiments have proved that the Fisher colorspace is much more effective than other existent colorpaces.

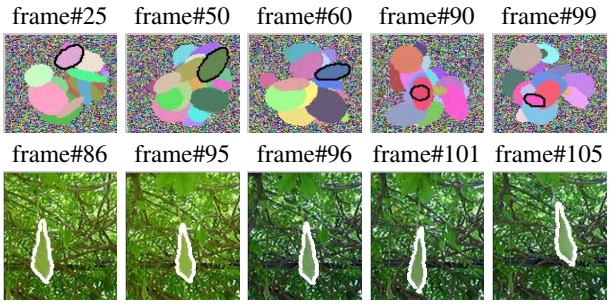


Figure 4. Tracking results of complex sequences adapting the Fisher colorspace.

References

[1] Crowley, J., Berard, F.: Multi-modal tracking of faces for video communications. CVPR, 1997
[2] Fukunaga, k.: Introduction to statistical pattern recognition. Academic Press, 1990
[3] Hayman, E., Eklundh, J.O.: Probabilistic and voting approaches to cue integration for figure-ground segmentation. ECCV, 2002
[4] Khan, S., Shah, M.: Object based segmentation of video using color, motion and spatial infomation. CVPR, 2001
[5] Moreno-Noguer, F., Sanfeliu, A., Samaras, D.: Integration of conditionally dependent object features for robust figure/background segmentation. ICCV, 2005
[6] Nummiaro, K., Koller-Meier, E., Van Gool, L.: An adaptive color-based particle filter. Im. and Vision Comput., 2003
[7] Perez, P., Hue, C., Vermaak, J., Gangnet, M.: Color-based probabilistic Tracking. ECCV, 2002
[8] Raja, Y., McKenna, S., Gong, S.: Color model selection and adaption in dynamic scenes. ECCV, 1998
[9] Rasmussen, C., Hager, G.D.: Probabilistic data association methods for tracking complex visual objects. PAMI, 2001
[10] Sherrah, J., Gong, S.: Continuous global evidence-based bayesian modality fusion for simultaneous tracking of multiple objects. ICCV, 2001
[11] Spengler, M., Schiele, B.: Towards robust multi-cue integration for visual tracking. Mach. Vision & Applications, 2003
[12] Wren, C., Azarbayejani, A., Darrel, T., Pentland, A.: Pfinder: Real-time tracking of the human body. PAMI, 1997
[13] Wu, Y., Huang, T.S.: Nonstationary color tracking for vision-based human-computer interaction. Trans. Neural Networks, 2002