

A Color Constancy Algorithm for the Robust Description of Images Collected from a Mobile Robot*

Jaume Vergés-Llahí and Alberto Sanfeliu

Institut de Robòtica i Informàtica Industrial
Technological Park of Barcelona, U Building
Llorens i Artigas 4-6, 08028 Barcelona, Catalonia
{jverges, asanfeliu}@iri.upc.es

Abstract. In mobile robotics, it is necessary to have a robust and efficient way of describing the visual stream provided by a vision system to be used afterwards in tasks such as object recognition. Color histograms are a useful tool to capture and represent color properties of sets of images taken from a certain position. Since those images were obtained at different time and light conditions, their appearance have greatly changed, reducing the performance of the color descriptor. In this work, we develop a color constancy algorithm that copes with the color variation among sets of images taken from nearly the same place. We show that the performance of the color histogram descriptor rises after color constancy, becoming a more robust and useful color descriptor. In the results section, we support that claim with several sets of images of scenes belonging to different positions.

Keywords: color histograms, color constancy, mobile robots

1 Introduction

Suppose a mobile robot equipped with a camera which is moving about an area in a way that is able to take images of the environment and associate them to a pose \mathbf{p} consisting in a position $(x, y) \in \mathbf{R}^2$ and an orientation $\theta \in [0, 2\pi)$, expressed in a chosen framework \mathcal{O} .

Consequently, after some time of moving, the robot will have provided a huge database of images of such area indexed through the pose \mathbf{p} . Any time a query is carried out to know what the robot is likely to see when located at a pose \mathbf{p} , given an uncertainty ϵ , the system answer will be a set of images $\mathcal{I}_{\mathbf{p}} = \{I_{\mathbf{p}_1}, \dots, I_{\mathbf{p}_M}$ such that $\|\mathbf{p} - \mathbf{p}_i\| < \epsilon\}$.

The set $\mathcal{I}_{\mathbf{p}}$ encompasses the whole visual knowledge of the system at a certain pose which can be useful in a number of tasks in mobile robotics, such as recognition of an object or person in the scene. Furthermore, if the set were stable and distinctive enough it could be used in the reversed way. Once in an

* Part. funded by the Gov. of Catalonia and the CICYT DPI2001-2223.

unknown position, identify where the robot is located matching what it sees with the descriptors of different locations in a database previously gathered.

Aforementioned tasks require that any set $\mathcal{I}_{\mathbf{p}}$ should be described in a simple, generic and computationally inexpensive way. Additionally, the images in $\mathcal{I}_{\mathbf{p}}$ must be mutually comparable, i.e., look similar at close poses, and so their descriptors. In other words, any descriptor accounting for the visual information of $\mathcal{I}_{\mathbf{p}}$ should be representative of the actual content of the scene at \mathbf{p} and be unaffected by the presence of variations.

There are three main sources of variation in $\mathcal{I}_{\mathbf{p}}$. First, $\|\mathbf{p}_i - \mathbf{p}_j\| < \epsilon$ implies the points of view where the images were taken from are not exactly the same, so are not the images $I_{\mathbf{p}_i}$ and $I_{\mathbf{p}_j}$. Secondly, some objects may have appeared, disappeared or just changed their relative position in images taken at different times due to independent movement or mutual occlusions. Third, objects can look differently due to changes in the position relative to light sources, but also to a variation through time of those lights.

This paper focuses on a framework to describe sets of images in a simple, generic and compact way coping with the previous sources of variation. First, we employ the α -trimmed average histogram to integrate the images in one descriptor by analogy to the problem of representing a group of frames in a video by their color histograms [1]. Since our images have greater color variation, we apply a color constancy technique to improve the descriptor stability. At the end, we show the performance of the descriptor and its improvement after using the color constancy step. This way, we get a robust representation of an area build up by different color histograms reduced to a canonic light conditions and combined in a way that outliers and light variation have been greatly reduced.

2 Descriptor of a Set of Images (SoI)

Our problem is close to that of finding a general framework for an efficient representation of video sequences, where the goal is usually to develop fast and robust algorithms for the identification of the video segment to which a query belongs [2]. In both cases, there is a set of images to be reduced to an efficient, robust and stable descriptor which can be used in a content-based query.

A generic mechanism to describe the content in a video sequence is the shot-based representation model. Once the shot is selected, it is customary to describe the visual and color content of shots using key frames and key frame histograms, respectively. Another framework is the groups of frames that are collections of frames selected according to a certain criterion. It is more flexible and general than the shot-based approach and well-suited for representing the sequential and hierarchical nature of video data.

Although the key frame histogram is a very simple descriptor of the color content of a shot, it is highly dependent on the selection criterion of the representative frame(s) and may lead to unreliable results. A more favourable approach is to consider the color content of *all* frames within a shot. Considering different strategies to get this *cumulative* color histogram, the simplest approach is the mean histogram, though potential problems with its sensitivity to outliers may make advisable the use of the median [1].

An alternative approach for computing the histogram of a set of images (SoI) is to define a family of α -trimmed average histograms [1], obtained by sorting the set of image histogram values for each bin in ascending order and averaging only the central $M - 2\lfloor\alpha M\rfloor$ elements of the ordered array, where M is the number of images in the SoI. Then, each bin j is computed as

$$\alpha TrimHist(j, \alpha) = \frac{1}{M - 2\lfloor\alpha M\rfloor} \sum_{i=\lfloor\alpha M\rfloor+1}^{M-\lfloor\alpha M\rfloor} \tilde{\mathcal{H}}_i(j) \quad (1)$$

where $\{\tilde{\mathcal{H}}_1(j), \dots, \tilde{\mathcal{H}}_M(j)\}$ is the sorted array of image histogram values for the j^{th} bin, i.e., $\{\mathcal{H}_1(j), \dots, \mathcal{H}_M(j)\}$. The trimming parameter $\alpha \in [0, \frac{1}{2}]$, controls the number of data points excluded from the average computation. While $\alpha = 0$ corresponds to the mean histogram, $\alpha = \frac{1}{2}$ computes the median one.

The interest of the above aggregation scheme is that of reducing the effect of outliers such as those belonging to objects that appear or disappear in the frames embodied in a SoI while easily encompassing either the mean or the median by tuning the α parameter. Additionally, the use of histograms minimizes the importance of relative positions of the objects in the scene. This way, two out of three main variation sources are reduced using those descriptors.

The third source of variability, illumination, needs special consideration. The kind of sequences in video data [1] have a relatively small light variation within a shot since a great effort in the visual uniformity of sequences has been put during the shooting and post-production. Nevertheless, in the case of images taken from a mobile robot, no control on the illumination can be done and the images belonging to the same SoI may have been taken at different times implying greater light variations which affect the color of the objects in the scene, as appreciated in Fig. (1).

3 Color Constancy in a Set of Images

Images in a SoI may present some degree of color variation due to changing light conditions. Our approach consists in reducing this variation before aggregating their color histograms using Eq. (1). On that purpose, we compute for every image in a SoI the set of all feasible color mappings rendering the image back to a canonical illumination – corresponding to a canonical image selected from the SoI – and selecting one mapping afterwards based on a measure of its likelihood.

More precisely, let I_c and I_a be the *canonic* and the *actual* images, respectively, picturing pretty much a similar scene since they are taken at nearly the same pose under two different illuminations. Our aim is to find the most likely color transformation T which maps the pixel colors of image I_a as close to those of image I_c as possible, hence reducing the color variation in the SoI. We now sketch the basis of our color constancy algorithm.

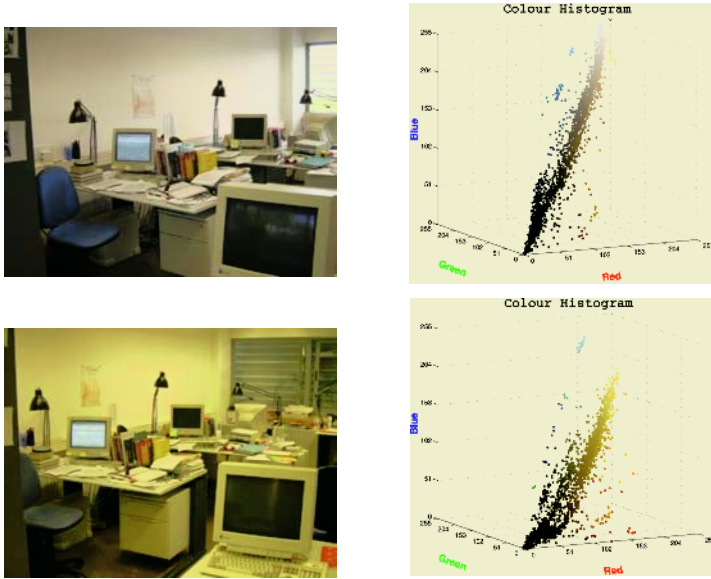


Fig. 1. Images taken nearly at the same a pose.

3.1 Likelihood Function

First, we introduce some notation and definitions. We note as $\mathcal{I} \subset \mathbf{R}^n$ a certain set of colors. Those colors can come from a specified color gamut or an image I . We also get the color histogram $\mathcal{H}(\mathcal{I})$ from \mathcal{I} . If the mapping $T \in \mathcal{T}$ is applied to every color in \mathcal{I} , the transformed set $T(\mathcal{I})$ is obtained. \mathcal{T} is the set of all feasible color mappings.

In general, given two color sets, \mathcal{I}_a and \mathcal{I}_c , a model of color change is a mapping $T \in \mathcal{T}$ so that $T(\mathbf{s}) = \mathbf{q}$, where $\mathbf{s} \in \mathcal{I}_a$ and $\mathbf{q} \in \mathcal{I}_c$ are corresponding colors. Finally, the set of feasible mappings can be defined as $\mathcal{T} = \{T = G(\mathbf{s}, \mathbf{q}) \mid \forall \mathbf{s} \in \mathcal{I}_a \text{ and } \forall \mathbf{q} \in \mathcal{I}_c\}$, where G is a *recovery scheme* (a function or an algorithm) computing one single mapping T out of two colors \mathbf{s} and \mathbf{q} .

The color constancy algorithm will select the most likely transformation \hat{T} from the set \mathcal{T} . Therefore, if $\mathcal{L}_G(T|\mathcal{I}_a, \mathcal{I}_c)$ is a function computing the likelihood for every mapping $T \in \mathcal{T}$, the algorithm finds the mapping that

$$\hat{T} = \operatorname{argmax}\{\mathcal{L}_G(T|\mathcal{I}_a, \mathcal{I}_c), T \in \mathcal{T}\} \tag{2}$$

A likelihood function \mathcal{L}_G can be related to a probability function \Pr in the way $\mathcal{L}_G(T|\mathcal{I}_a, \mathcal{I}_c) = \log(\Pr(T|\mathcal{I}_a, \mathcal{I}_c))$. So, first we must get a value for the probability \Pr of a certain mapping T . As an estimate of $\Pr(T|\mathcal{I}_a, \mathcal{I}_c)$ we use the histogram of the set of feasible mappings, $\mathcal{H}(\mathcal{T})$, computed from the sets \mathcal{I}_a and \mathcal{I}_c and the recovery scheme G . *The key idea is that the more likely a mapping is, the more frequent it should be in the histogram $\mathcal{H}(\mathcal{T})$.*

Let $G^{-1}(T) = \{(\mathbf{s}, \mathbf{q}) \in \mathcal{I}_a \times \mathcal{I}_c \mid G(\mathbf{s}, \mathbf{q}) = T\}$ be the set of all pairs (\mathbf{s}, \mathbf{q}) giving rise to a certain mapping T using the recovery scheme G . The set $G^{-1}(T)$ can

be taken instead of \mathbf{T} since $\mathbf{G}^{-1}(\mathbf{T}) = \mathbf{G}^{-1}(\mathbf{T}') \Leftrightarrow \mathbf{T} = \mathbf{T}'$. Hence, $\Pr(\mathbf{T}|\mathcal{I}_a, \mathcal{I}_c)$ is estimated as $\Pr(\mathbf{G}^{-1}(\mathbf{T})|\mathcal{I}_a, \mathcal{I}_c)$.

In addition, since our color sets are discrete, $\mathbf{G}^{-1}(\mathbf{T})$ can be thought as a finite disjoint union of singletons $\{(\mathbf{s}, \mathbf{q})\}$. Each singleton $\{(\mathbf{s}, \mathbf{q})\} \in \mathcal{I}_a \times \mathcal{I}_c$ can be further divided in two different pieces, namely, $\{\mathbf{s}\} \in \mathcal{I}_a$ and $\{\mathbf{q}\} \in \mathcal{I}_c$, which can be assumed to be independent. Therefore,

$$\Pr((\mathbf{s}, \mathbf{q})|\mathcal{I}_a, \mathcal{I}_c) = \Pr(\mathbf{s}|\mathcal{I}_a) \cdot \Pr(\mathbf{q}|\mathcal{I}_c) \quad (3)$$

where $\Pr(\mathbf{s}|\mathcal{I}_a)$ and $\Pr(\mathbf{q}|\mathcal{I}_c)$ estimates come from their corresponding bins in histograms $\mathcal{H}(\mathcal{I}_a)$ and $\mathcal{H}(\mathcal{I}_c)$, respectively. Then,

$$\Pr(\mathbf{T}|\mathcal{I}_a, \mathcal{I}_c) = \sum_{\forall(\mathbf{s}, \mathbf{q}) \in \mathbf{G}^{-1}(\mathbf{T})} \Pr(\mathbf{s}|\mathcal{I}_a) \cdot \Pr(\mathbf{q}|\mathcal{I}_c) \quad (4)$$

That is, the frequency of the bin corresponding to a mapping \mathbf{T} in the histogram of feasible mappings $\mathcal{H}(\mathcal{T})$ can be computed adding the product of frequencies of the two bins in $\mathcal{H}(\mathcal{I}_a)$ and $\mathcal{H}(\mathcal{I}_c)$ corresponding to all the color pairs giving rise to the mapping \mathbf{T} by means of the recovery scheme \mathbf{G} .

In order to improve the robustness of Eq. (4), a measure of similarity between the transformed set $\mathbf{T}(\mathcal{I}_a)$ and the canonical set \mathcal{I}_c is taken into account which evaluates the performance of a particular mapping. We use the Swain&Ballard intersection measure [3] defined as $\cap(\mathcal{H}, \mathcal{M}) = \sum_k \min\{H_k, M_k\} \in [0, 1]$ for its computational simplicity. This measure is helpful in practice to eliminate outlier mappings among the set of candidates.

Finally, we joint the probability and performance of a mapping in a single likelihood function as follows

$$\mathcal{L}_{\mathbf{G}}(\mathbf{T}|\mathcal{I}_a, \mathcal{I}_c) = \log(\cap(\mathbf{T}(\mathcal{H}_a), \mathcal{H}_c)) \cdot \Pr(\mathbf{T}|\mathcal{I}_a, \mathcal{I}_c) \quad (5)$$

where $\mathcal{H}_a = \mathcal{H}(\mathcal{I}_a)$, $\mathcal{H}_c = \mathcal{H}(\mathcal{I}_c)$ and $\Pr(\mathbf{T}|\mathcal{I}_a, \mathcal{I}_c)$ is as in Eq. (4). Furthermore, $\mathbf{T}(\mathcal{H})$ is the transformation of a histogram \mathcal{H} by \mathbf{T} .

3.2 Color Change Model and Recovery Scheme

To complete the previous scheme, the kind of color coordinates and the model of color change \mathbf{T} as well as the recovery function \mathbf{G} must be explicitly stated.

First, colors are vectors in \mathbf{R}^n , where $n = 3$ in a (R, G, B) color space or $n = 2$ in a *chromaticity space*. In our case, to alleviate problems found in images with specularities or shades, and to reduce at the same time the computational burden, we use the *perspective color coordinates* $(r, g) = (R/B, G/B)$ defined by Finlayson in [4] which discard intensity. Finlayson and Hordley proved in [5] that the set of feasible mappings computed in a 3D space and projected into a 2D space afterward is the same as the set computed directly in a 2D space.

A general lineal color change model can be mathematically describe as $\mathbf{T}(\mathbf{s}) = \mathbf{T} \cdot \mathbf{s} = \mathbf{q}$, where $\mathbf{T} \in \mathcal{M}_n(\mathbf{R})$ is a square matrix encompassing a particular color change between two different lights [6]. A reasonable tradeoff between simplicity and performance can be attained employing a diagonal model [6, 7, 4]. This model

assumes color sensors are completely uncorrelated and any change in the light arriving to them equates to independently scaling each channel value, that is, $\mathbf{T} = \text{diag}(t_1, \dots, t_n)$. Equivalently, \mathbf{T} can be also expressed as a vector $\mathbf{t} = (t_1, \dots, t_n) \in \mathbf{R}^n$. Therefore, for any pair $(\mathbf{s}, \mathbf{q}) \in \mathcal{I}_a \times \mathcal{I}_c$, the color change model we use is

$$\begin{aligned} \mathbf{T}: \mathcal{I}_a &\longrightarrow \mathcal{I}_c \\ \mathbf{s} &\longmapsto \mathbf{T}(\mathbf{s}) = \mathbf{T} \cdot \mathbf{s} = \mathbf{q} \end{aligned} \quad (6)$$

and the recovery scheme is

$$\begin{aligned} \mathbf{G}: \mathcal{I}_a \times \mathcal{I}_c &\longrightarrow \mathcal{T} \subset \mathbf{R}^n \\ (\mathbf{s}, \mathbf{q}) &\longmapsto \mathbf{G}(\mathbf{s}, \mathbf{q}) = \left(\frac{q_1}{s_1}, \dots, \frac{q_n}{s_n}\right) = \mathbf{t} \end{aligned} \quad (7)$$

where $\mathcal{T} = \left\{ \left(\frac{q_1}{s_1}, \dots, \frac{q_n}{s_n}\right) \mid \forall (\mathbf{s}, \mathbf{q}) \in \mathcal{I}_a \times \mathcal{I}_c \right\}$.

3.3 A Color Constant Set of Images

Once a *canonical* image $I_c \in \mathcal{I}$ is selected, we apply the previous color constancy to the set $\{I_1, \dots, I_M\}$ employing their color histograms $\{\mathcal{H}_1, \dots, \mathcal{H}_M\}$ and the canonical \mathcal{H}_c . Afterward, the set $\{\mathbf{T}_1(\mathcal{H}_1), \dots, \mathbf{T}_M(\mathcal{H}_M)\}$ of transformed histograms can be used to generate the α -trimmed average histogram defined in Eq. (1) representing the whole SoI, noted as \mathcal{H}_{SoI} .

4 Results

We want to show how the performance of the scene description \mathcal{H}_{SoI} improves after using our color constancy algorithm. A natural way to measure the fidelity of these descriptors is in terms of the average error $E_{\mathcal{H}_{SoI}}$ within a set of images defined in [1] as the average of the accumulated distance between each image histogram \mathcal{H}_i in the SoI and the SoI histogram \mathcal{H}_{SoI} ,

$$E_{\mathcal{H}_{SoI}} = \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^N |\mathcal{H}_i(j) - \mathcal{H}_{SoI}(j)| \quad (8)$$

where N and M are the number of bins and that of images, respectively. This error provides a consistent way to assess the performance of the proposed descriptor. The less representative of the color content of the SoI the descriptor is, the greater the error measure is.

Our database consist in a set of 21 images at five different poses and with very different light conditions picturing a regular office environment. The canonical image of these scenes can be seen in the leftmost column of Fig. (2). The first four scenes have 3 images each, while the last one has 9. Besides, images belonging to the same scene have appreciable variations in the point of view and light conditions, as depicted in Fig. (1). The group with a greater variation is the last one, where there are occlusions and some objects have appeared or disappeared.

The rightmost column in Fig. (2) shows the error $E_{\mathcal{H}_{SoI}}$ as a function of the trimming parameter α . The results obtained if no color constancy step – No CC

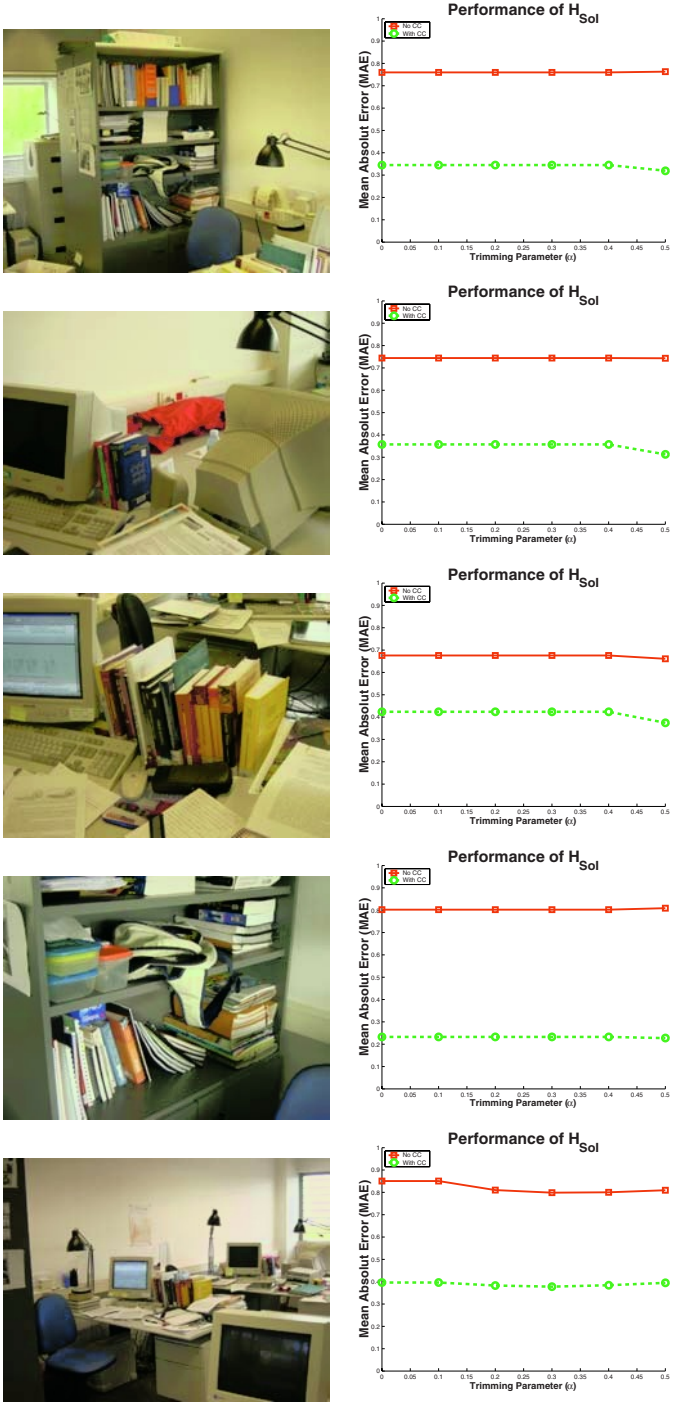


Fig. 2. Scenes and Performance plots.

Table 1. No Color Constancy Step.

Scene	Mean	Std	Max	Min
No.1	0.760	0.001	0.763	0.760
No.2	0.744	0.000	0.744	0.743
No.3	0.673	0.006	0.676	0.661
No.4	0.804	0.003	0.809	0.802
No.5	0.820	0.024	0.851	0.799

Table 2. With Color Constancy Step.

Scene	Mean	Std	Max	Min	% Red.
No.1	0.341	0.010	0.345	0.320	55.13
No.2	0.350	0.018	0.358	0.313	52.96
No.3	0.415	0.020	0.424	0.374	38.34
No.4	0.232	0.002	0.233	0.227	71.14
No.5	0.389	0.008	0.396	0.378	52.56

– is applied are plotted with a red solid line with squares, while those obtained after the color constancy step – With CC – are depicted with a green dashed line with circles. To clarify numerically those results, we have expressed the mean, standard deviation, maximum and minimum values of these plots in Table 1 and 2, as well as the error reduction in the last column in Table 2.

The results clearly show there has been a reduction (54% in average) of the error $E_{\mathcal{H}_{SoI}}$ between the image histograms \mathcal{H}_i and the scene descriptor \mathcal{H}_{SoI} after the color constancy step, which means a decrease of the internal image variation and an improvement for representativeness of SoI by the proposed scene descriptor. Furthermore, the obtained error reduction is far greater than that achieved by only tuning the trimming parameter, which means that a simpler descriptor could be used instead, for example, by only employing the mean ($\alpha = 0$) or the median ($\alpha = 1/2$) to compute \mathcal{H}_{SoI} .

5 Conclusions

This paper has exposed the problem of obtaining a global color-based description of a set of images captured from a mobile robot picturing a particular location. Our main concern has been to generate a kind of descriptors robust to a certain number of sources of variation, being color variation due to illumination changes the most important one to cope with among images taken from a mobile platform. As a result, we have shown that a reduction in the average error between the histograms of the set images and the histogram describing the whole set can be attained by applying a color constancy step. The color descriptor is then more robust and representative of the set since its color variation has been greatly reduced. A future work would be the extension of these descriptors to the task of robot localization only by means of the set of captured images.

References

1. Ferman, A.M., Tekalp, A.M., Mehrotra, R.: Robust color histogram descriptors for video segment retrieval and identification. *IEEE Trans. on Image Processing* **11** (2002) 497–508
2. Idris, F., Panchanathan, S.: Review of image and video indexing techniques. *Journal of Visual Communication and Image Representation* **8** (1997) 146–166
3. Swain, M., Ballard, D.: Indexing via color histograms. In: *Proc. Int. Conf. on Computer Vision.* (1990) 390–393

4. Finlayson, G.: Color in perspective. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **18** (1996) 1034–1038
5. Finlayson, G., Hordley, S.: Improving gamut mapping color constancy. *IEEE Trans. on Image Processing* **9** (2000) 1774–1783
6. Forsyth, D.: A novel algorithm for color constancy. *Int. Journal of Computer Vision* **5** (1990) 5–36
7. Finlayson, G., Hordley, S., Hubel, P.: Colour by correlation: A simple, unifying framework for colour constancy. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **23** (2001) 1209–1221