NOTE

Motion from Color

P. Golland* and A. M. Bruckstein†

Center for Intelligent Systems, Computer Science Department, Technion, I.I.T., Haifa 32000, Israel

The use of color images for motion estimation is investigated in this work. Beyond the straightforward approach of using the color components as separate images of the same scene, a new method, based on exploiting color invariance under motion, is discussed. Two different sets of color-related, locally computable motion "invariants" are analyzed and tested in this paper, and the results of motion estimation based on them are compared to the direct use of the RGB brightness functions. © 1997 Academic Press

1. INTRODUCTION

Optical or image flow estimation is considered by many researchers to be an important low-level stage of spatial motion recovery from a sequence of images. It is supposed to yield an estimate of the 2D projection of the velocity field on the image plane, which is submitted to further analysis aimed at inferring high-level, 3D motion descriptions. It is well known that the image flow cannot be completely determined from a single sequence of black-and-white images without introducing additional assumptions about the nature of motions present in the image. Color images are a natural source for additional information that should greatly facilitate the solution of this problem.

The motion of objects in a time-varying scene is completely defined by the so-called *motion field*, which is a 3D flow field defining object velocities at each point in space. One of the aims of the motion recovery process is to reconstruct the motion field of the scene. The motion field can, and usually does, change in time. If we look at the scene via a camera, we can only see its time-varying 2D projection. Hence as an intermediate stage in the 3D motion estimation process we need to find the 2D projection of a motion field onto the image plane. This 2D field of velocities is called *image flow*. Then we could attempt to recover the 3D motion field from its 2D projection. To do this, however, seems impossible without additional information about the motion field. Instead, we can extract the so-called *optical flow*, which is a 2D field of velocities associated with the variation of brightness patterns of the image [7, 17].

The following two examples are often given to help understand the difference between an image flow and an optical flow. A uniformly painted ball is rotating around its center in some way. In this case the image flow in non-zero for every point of the ball's projection on the image plane, while the optical flow is zero, since the image brightness will not change at all. The second example is a stationary scene with a moving light source. Now the situation is exactly the opposite: the optical flow is non-zero due to intensity changes in the image, whereas absence of any true motion causes zero image flow.

To summarize, the motion recovery problem can be formulated as follows: given a sequence of images of a dynamic scene, recognize moving objects and find their velocities (trajectories). The solution to this problem, like many others in the field of computer vision, is usually divided into two main stages:

• Low-level processing. During this stage a 2D field of velocities (the *image flow* or the *optical flow*) associating a velocity vector to each point of the image plane is determined.

• High-level motion analysis. At this stage the 3D velocity field (the true *motion field*) is estimated from the 2D field determined at the previous stage and is analyzed in order to find the motion of objects in space.

* To whom correspondence should be addressed at MIT AI Lab, 545 Technology Square, Room 737, Cambridge, MA 01239.

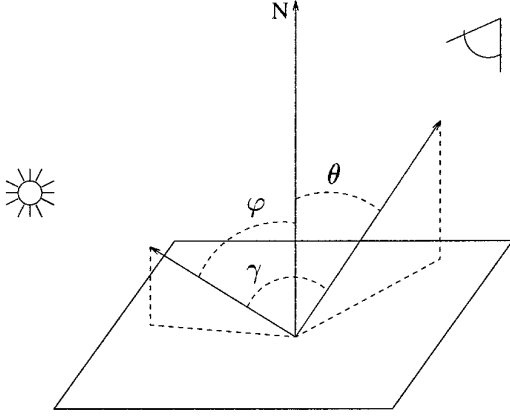† On sabbatical at AT&T Bell Laboratories, Murray Hill, NJ 07974.

FIG. 1. Light reflection by a surface.

δt the image of the point X will move to a new position $\mathbf{x} + \mathbf{u}\,\delta t$. The brightness conservation assumption implies

$$E(\mathbf{x}, t) = E(\mathbf{x} + \mathbf{u}\,\delta t, t + \delta t), \qquad (1)$$

where E is brightness in the image plane. Using Taylor series expansion we can write (1) as

$$\frac{\partial E}{\partial x} u\,\delta t + \frac{\partial E}{\partial y} v\,\delta t + \frac{\partial E}{\partial t}\,\delta t + O(\delta t^2) = 0, \qquad (2)$$

and assuming an infinitesimal time interval we obtain the equation

$$\frac{\partial E}{\partial x} u + \frac{\partial E}{\partial y} v + \frac{\partial E}{\partial t} = 0, \qquad (3)$$

which will be referred to in what follows as the major optical flow constraint. It involves two unknowns at each point of the image plane: the optical flow components u and v. If we rewrite this equation using the gradient notation

$$(\nabla E)^T \mathbf{u} + E_t = 0, \qquad (4)$$

it can be easily seen that only the velocity component parallel to the brightness gradient vector can be determined from the major optical flow constraint. This is the celebrated *aperture problem.* It is obvious that the optical flow field cannot be found from this equation only, but some further assumptions have to be made, or additional constraints must be found in order to complete the system of equations and determine both optical flow components. Gradient-based methods address this problem in various ways.

Horn and Schunck [7] were the first to use a smoothness constraint on the optical field. They used the magnitude of the velocity field gradient as a measure of smoothness (or, more precisely, unsmoothness), determining the flow that minimized a weighted sum of the field unsmoothness measure and the errors in satisfying the major optical flow constraint (3). Their method provided quite satisfactory results, but failed on edges. This drawback was subsequently corrected by Nagel [11] via the introduction of a weighted optimization process, in which smoothness of the velocity field was required only for the optical flow component normal to the intensity gradient in places where its magnitude was large (edges).

Haralick and Lee [5] then proposed a method using higher-order derivatives of the image intensity. They assumed that not only the intensity function itself but also its first-order derivatives remain unchanged under motion, obtaining three additional equations similar to (3). A slightly different method also using higher-order deriva-

While the high-level stage of motion recovery assumes that it receives the image flow as its input, the low-level stage can usually produce only the optical flow defined by the image sequence. It is immediately recognized that a problem arises due to the difference between these two fields. A number of authors [17, 20] investigated the connection between them, introducing constraints on motions and surface properties in order to either satisfy assumptions that make the image flow and the optical flow identical or propose methods to obtain the true image flow. For spatial motion recovery it is usually assumed that the optical flow and the image flow are close enough to be used interchangeably. In some situations this approximation is quite reasonable, but one should not forget the two examples shown above, where the optical flow and the image flow differ significantly.

A number of approaches have been developed for optical flow estimation: gradient-based methods, region-matching (or correlation) methods, energy based methods, etc. For each approach, several variations have been proposed by different authors. Since the new methods proposed in this work belong to the gradient-based group, our discussion will concentrate on methods from this group and the general philosophy underlying them.

2. GRADIENT-BASED APPROACH FOR OPTICAL FLOW ESTIMATION

As defined above, the optical flow is a velocity field associated with brightness changes in the image. This suggests an assumption often made in methods for optical flow estimation, the brightness conservation assumption, which states that brightness of the image of a point on the object is invariant under motion.

Let $\mathbf{x} = (x, y)$ be a position of an image of some point X of the scene at time t and $\mathbf{u} = (x, v)$ be a projection of this points velocity onto the image plane. Then after time

tives was developed by Tretiak and Pastor [18]. In [12] Nagel showed that these two methods were particular cases of a more general approach and developed a generalized scheme for optical flow estimation using higher-order derivatives.

Another approach, called "the neighborhood-sampling approach" assumes a certain local behavior for the optical flow, implying that the velocity components at all points of a certain small enough neighborhood can be well approximated by the velocity components at one (usually central) point of this neighborhood and the assumed spatial behavior of the field. This approach was used by Lucas and Kanade [9] (constant model), Campani and Verri [3] (linear in space and constant in time), and Otte and Nagel [15] (linear both in space and in time). The velocity components at a certain point of the chosen neighborhood were expanded into Taylor series relative to the central point of the neighborhood and then substituted into the major optical flow constraint (3). This provided additional equations on the flow components and their first-order derivatives (in a linear case); the number of equations was equal to the number of points in the chosen neighborhood.

The last group of methods proposed in the literature is based on using multiple constraints. Each of these methods tries to find and extract more than one function invariant under motion and thereby determine the optical flow components using a number of constraints based on those invariants. This approach was used by Wohn *et al.* [21], Mitiche *et al.* [10], and Ohta [13], each of them using different invariants such as brightness averages, medians, curvature, and the RGB color channels.

The methods discussed in this work use the multiple constraint approach for image flow estimation. Local properties based on the color of object surfaces are extracted from raw RGB color images and are used to get an overdetermined system of linear equations for the two unknown components of the optical flow. This is done under the basic assumption that under changing illumination conditions the color of objects is more intimately related to surface properties than their brightness.

3. IMAGE FLOW FROM COLOR IMAGES

In this section two methods using color images for image flow estimation are introduced and discussed. The first one assumes brightness conservation under motion. It uses the (R, G, B) quantities, provided by color images, in a straightforward way. The second method uses quantities which represent the intrinsic color properties of an object more precisely than the (R, G, B) values and uses them for image flow estimation. The assumption underlying this method is a color conservation assumption, which is clearly weaker and more realistic than the brightness conservation assumption.

3.1. *The Brightness Conservation Approach*

Let us first briefly consider the loop perception process and discuss differences between color and black-and-white vision.

We can think about any vision system as an array of light detectors of different types. Each type is characterized by its sensitivity to light in the visible range of wavelengths. The number of detector types present in a system can vary. A system with detectors of one type is associated with black-and-white vision. According to biological research, the human vision system contains light detectors of three different types, as do most color cameras.

Detectors of different types produce different images in response to the same input spectrum $S(\lambda)$ due to differences in their sensitivity functions, denoted by $D_r(\lambda)$, $D_g(\lambda)$, and $D_b(\lambda)$. If all detectors are exposed to the same input spectrum $S(\lambda)$, the color perceived by the vision system is determined by three non-negative numbers, denoted by (R, G, B), obtained from the formulae [2, 6, 19]

$$R = \int_\Omega S(\lambda) D_r(\lambda) \, d\lambda,$$
$$G = \int_\Omega S(\lambda) D_g(\lambda) \, d\lambda, \qquad (5)$$
$$B = \int_\Omega S(\lambda) D_b(\lambda) \, d\lambda,$$

the integration being over the visible range of wavelength Ω ([400 nm, 700 nm] for the human vision system).

The only difference between a black-and-white vision system and a color one is that a BW system has detectors that produce at each point in the image a single number, which is usually called the brightness, or the intensity function of the image, whereas a color system produces three values for each image point. This implies that any color image can be straightforwardly considered as three different black-and-white images produced by three types of light detectors with different sensitivity functions in response to the same input image. Each one of the color image components should therefore satisfy reasonable assumptions on the brightness function of black-and-white images.

The brightness conservation assumption implies that the (R, G, B) values of a certain point in the image should remain unchanged under motion of this point within a small temporal neighborhood. Therefore, the images R, G, and B can be used in a similar way to the usual BW brightness function to constrain the velocity flow components at each point of the image. If the major optical flow constraint (3) is applied to each of the (R, G, B) quantities, the following overdetermined system of linear equations is obtained:

$$\frac{\partial R}{\partial x} u + \frac{\partial R}{\partial y} v + \frac{\partial R}{\partial t} = 0,$$

$$\frac{\partial G}{\partial x} u + \frac{\partial G}{\partial y} v + \frac{\partial G}{\partial t} = 0, \qquad (6)$$

$$\frac{\partial B}{\partial x} u + \frac{\partial B}{\partial y} v + \frac{\partial B}{\partial t} = 0.$$

The image flow components can be found from this system using a pseudo-inverse solution. If we define

$$A = \begin{pmatrix} R_x & R_y \\ G_x & G_y \\ B_x & B_y \end{pmatrix}, \quad b = \begin{pmatrix} -R_t \\ -G_t \\ -B_t \end{pmatrix}, \qquad (7)$$

then the pseudo-inverse solution of the system (6) is equivalent to the solution of the well-defined system of two linear equations

$$(A^T A)\mathbf{u} = A^T b, \qquad (8)$$

given by

$$\mathbf{u} = (A^T A)^{-1} A^T b. \qquad (9)$$

This assumes, of course, that the matrix $A^T A$ has rank 2, i.e., it is non-singular.

In addition to the estimates of the image flow components at a certain point of the image, we would like to get some measure of confidence in the result at this point, which would tell us to what extent we could trust our estimates. It is common to use the so-called *condition number* of the coefficient matrix of a system (in our case $A^T A$) as a measure of confidence in the solution of this system.

The formal definition of the condition number k of a matrix B is [14, 8]

$$k(B) \triangleq \begin{cases} \|B\| \|B^{-1}\|, & \text{if } B \text{ is non-singular} \\ +\infty, & \text{if } B \text{ is singular} \end{cases} \qquad (10)$$

where $\|B\|$ is the norm of the matrix B equal to the maximal eigenvalue of the matrix $B^T B$. The condition number of the matrix B measures "numerical stability" of a system of the form

$$Bx = b, \qquad (11)$$

where b is a vector of free coefficients and x is a vector of unknowns. It provides an estimate of the relative errors induced in the result due to the presence of some errors in data (i.e., in the vector b). To summarize, both the

velocity vector and the condition number of the matrix $A^T A$ are computed at each point of the image and then analyzed in order to get a final result.

A similar method was first proposed by Ohta [13]. He suggested using pairs of color channels in order to get two equations for the two unknown components of the image flow. However, he did not test this method experimentally; neither did he propose ways to combine the results obtained from different pairs of equations.

The experimental testing of the pseudo-inverse method demonstrated that it is quite stable to noise and produces rather accurate results for images of objects translating in a plane parallel to the image plane. If rotation or motion normal to the image plane were present, this method sometimes provided erroneous estimates. The reason for this failure is explained in the next section, and a new method, capable of dealing with all kinds of motion, is proposed.

3.2. The Color Conservation Approach

Let us consider a surface illuminated by light, which is characterized by some power distribution $I(\lambda, \mathbf{r})$ over the effective range of wavelength λ, at any point $\mathbf{r}$ at the object surface. The reflected light has its own power distribution $\hat{I}(\lambda, \mathbf{r})$, not necessarily equal to that of the incident light. A common assumption is that reflectivity of a certain point $\mathbf{r}$ depends only on the material the object is made of and on the surface geometry at the point $\mathbf{r}$. Hence

$$\hat{I}(\lambda, \mathbf{r}) = R(\lambda, \varphi, \theta, \gamma, \mathbf{r}) I(\lambda, \mathbf{r}), \qquad (12)$$

where $(\varphi, \theta, \gamma)$ are the angles of incidence, observation, and phase, respectively (Fig. 1). The factor $R(\lambda, \varphi, \theta, \gamma, \mathbf{r})$ defines completely the reflectivity properties of the surface and is called the *reflectivity function* of the surface.

It has been proven experimentally that the reflectivity function can be separated into two factors, one only depending on the properties of the surface and the second depending on the geometry of the reflection process only:

$$\hat{I}(\lambda, \mathbf{r}) = \rho(\lambda, \mathbf{r}) c(\varphi, \theta, \gamma, \mathbf{r}) I(\lambda, \mathbf{r}). \qquad (13)$$

Here $\rho(\lambda, \mathbf{r})$ is the spectral factor, representing color properties of the object surface, and $c(\varphi, \theta, \gamma, \mathbf{r})$ is the geometric factor. This very simple model conforms to reality with considerable accuracy in many cases.

If Eqs. (5) and (13) are combined together, we obtain

$$R = \int_\Omega c(\varphi, \theta, \gamma) \rho(\lambda) I(\lambda) D_r(\lambda) \, d\lambda,$$

$$G = \int_\Omega c(\varphi, \theta, \gamma) \rho(\lambda) I(\lambda) D_g(\lambda) \, d\lambda, \qquad (14)$$

$$B = \int_\Omega c(\varphi, \theta, \gamma) \rho(\lambda) I(\lambda) D_b(\lambda) \, d\lambda,$$

Since the geometry component $c(\varphi, \theta, \gamma)$ of the reflectivity function does not depend on the light wavelength, it can be moved out of the integral to obtain

$$R = c(\varphi, \theta, \gamma) \int_\Omega \rho(\lambda) I(\lambda) D_r(\lambda) \, d\lambda = c(\varphi, \theta, \gamma) C_r,$$

$$G = c(\varphi, \theta, \gamma) \int_\Omega \rho(\lambda) I(\lambda) D_g(\lambda) \, d\lambda = c(\varphi, \theta, \gamma) C_g, \quad (15)$$

$$B = c(\varphi, \theta, \gamma) \int_\Omega \rho(\lambda) I(\lambda) D_b(\lambda) \, d\lambda = c(\varphi, \theta, \gamma) C_b.$$

These equations imply that each of the brightness functions (R, G, B) can be separated into the geometry component $c(\varphi, \theta, \gamma)$, which depends entirely on the relative position and orientation of the object, the light source, and the camera, and the spectral component C_i defined by

$$C_i = \int_\Omega \rho(\lambda) I(\lambda) D_i(\lambda) \, d\lambda, \quad i \in \{r, g, b\}. \quad (16)$$

Let us examine all the factors that influence the C_i quantities. The spectral component $\rho(\lambda)$ of the reflectivity function depends neither on the geometry of the scene nor on the illumination properties. It represents color properties of the object and remains invariant under any changes of the object position and orientation. The sensitivity function $D_i(\lambda)$ of the light detector does not change under any variations of the scene either. If the illumination spectrum $I(\lambda)$ is assumed to change slowly enough, so that it can be considered constant over a small temporal neighborhood,[1] the C_i quantities defined by (16) satisfy the invariance assumption: they do not change as the object or the camera move.

We can think of the C_i quantities as representing the object color under a certain illumination $I(\lambda)$. Since the object color is invariant under motion (unless we are talking about a chameleon!), the quantities representing color under constant illumination are also invariant and can be used for motion estimation. Note that the flow estimates obtained using the color conservation assumption should no longer be called the optical flow, because, according to the definition, the optical flow is associated with the brightness changes in the image. However, the new estimates, associated with the color changes in the image, are closer to the image flow since they use properties intimately connected to surface properties of objects. We shall therefore use the term "image flow estimation" to denote the flow estimates obtained using the color functions C_i of the image.

Now it is obvious that any method using the brightness conservation assumption produces good results, i.e., close to the image flow, if also the reflection geometry (and therefore the geometry component $c(\varphi, \theta, \gamma)$) does not change while the object moves. If a parallel light source or uniform illumination is used, then translation does not change the geometry of the reflection process. A similar observation was made by Singh [17], who showed theoretically that the optical flow is equal to the translation component of the image flow for objects with lambertian surface reflectivity. If the geometry of the reflection process does change significantly with object motion (e.g., for rotation, motion toward the camera, etc.), the brightness function does not satisfy the conservation assumption any more. The new functions C_i, however, remain invariant under any kind of motion, because they are not influenced by the geometry of the reflection process.

The next question is how to extract the C_i quantities from the (R, G, B) functions provided by a color image. Although it is impossible to extract explicitly the C_i values from the (R, G, B) values, it is immediately seen that ratios of any two linear combinations of the (R, G, B) values are equal to ratio of the corresponding C_i value combinations, since the geometry component $c(\varphi, \theta, \gamma)$ is the same for each of the (R, G, B) functions. If the quantities C_i are invariant under motion, then their ratios are also invariant, and therefore we can use ratios of C_i values to estimate motion. The interesting fact is that a number of different color representations were defined using ratios of the (R, G, B) quantities. They usually separate color properties of a spectrum from its brightness, or intensity. Two such representations, the normalized *rgb* system and the HSV system, were chosen in this work for experimental testing. Their definitions and a brief discussion on their properties can be found in the Appendix.

Each of these systems introduces two independent quantities to represent color properties of a spectrum: the *rgb* system uses *r* and *g* (or any other pair of the *r*, *g* and *b* quantities) and the HSV system defines *Hue* and *Saturation* for this purpose. In both systems, the two independent quantities representing color properties of a spectrum are defined as different ratios of the (R, G, B) quantities.

If the quantities used by the method are denoted F_1 and F_2 (where $F_1 = r$, $F_2 = g$ for the *rgb* system, and $F_1 = Hue$, $F_2 = Saturation$ for the HSV system), the color conservation assumption implies

$$\frac{\partial F_1}{\partial x} u + \frac{\partial F_1}{\partial y} v + \frac{\partial F_1}{\partial t} = 0,$$

$$\frac{\partial F_2}{\partial x} u + \frac{\partial F_2}{\partial y} v + \frac{\partial F_2}{\partial t} = 0. \quad (17)$$

---

[1] Such an assumption on the illumination behavior is quite a realistic one if stationary light sources are used or if the outdoor scene is photographed while weather conditions do not change sharply between successive frames.

If F_1 and F_2, or more precisely, their gradients, are linearly independent, this system is well determined and its solution provides an estimate of the image flow.

Similarly to the method proposed in the previous section, we would like to compute not only the velocity estimates, but also a confidence measure in these estimates, which is defined as the condition number k of the coefficients matrix

$$A = \begin{pmatrix} \dfrac{\partial F_1}{\partial x} & \dfrac{\partial F_1}{\partial y} \\[2ex] \dfrac{\partial F_2}{\partial x} & \dfrac{\partial F_2}{\partial y} \end{pmatrix}. \tag{18}$$

At the final stage of the estimation process only the results with the condition numbers less than a certain threshold would be taken into account.

When a new method is proposed, its drawbacks should also be discussed. Like all gradient-based methods, the new method proposed for image flow estimation requires the presence of significant gradients of the functions it is based on. If the magnitude of the brightness gradient in the image is small, any gradient-based method using brightness would fail to produce reliable results. In a similar way, if there is no gradient of color in the scene, our method will fail to produce good estimates. This implies that the method should not be used when a scene contains objects with uniform color. In such cases, if motion estimation is necessary a color gradient could be artificially created by painting certain areas of the object surfaces with good color gradients. Another possibility is to use brightness information in regions of the insufficient color gradient. Even though the brightness conservation assumption might lead to higher errors than the color-based method, it should be used if the color information is not rich enough to yield reliable estimates of the velocity field. Since the main concern of this paper is the possibility for using color properties of the object surface for motion estimation, we used only one set of functions in every experiment, but one can readily extend this method to use the brightness function and its local variations in regions where the confidence in the velocity estimates obtained using the color functions is low.

Another interesting issue that should be further addressed is how close the color functions used in this paper are to the truly invariant color characteristics of the object surface. As was mentioned above, we use a fairly simple model for light reflectance that might fail if the surface reflectance is a mixture of both diffuse and specular components. In such cases a different reflectance model should be used to describe the reflectance process.

Ideally, we should have used a color constancy algorithm to extract the "true" color characteristics from the image and to use them as invariants for motion estimation. There has been some research done on exact separation of color brightness [2, 16, 4], but no definite method, that would incorporate the most general reflectance model, has yet been proposed. The normalized RGB and the HSV representations proved to be reasonable approximations of the color characteristics and have been used extensively in computer vision to separate the color properties of the surface from the brightness component.

4. EXPERIMENTAL RESULTS

All the methods proposed in this chapter have been tested using equipment at the Center for the Intelligent Systems in the Computer Science Department, Technion.

Both synthetic and real image sequences were used for the testing. The results obtained by methods using color functions were compared with the results provided by using brightness functions for motion estimation.

Barron *et al.* [1] implemented a number of existing techniques for optical flow estimation in order to compare their performance. This work provides a lot of useful information about the practical aspects of the optical flow estimation problem. We based our implementation on this work. All the images were first smoothed using Gaussian kernel with the different values of the σ parameter for synthetic ($\sigma = 1.5$) and real ($\sigma = 2.5$) images. The smoothing operation was performed separately in space (x and y directions) and in time (t direction). The first-order derivatives of the image functions (intensity or color functions) were estimated using the $1/12(1, -8, 0, 8, -1)$ kernel to compute each derivative. The spatial derivatives were computed using a 5×5 neighborhood, where each line was a copy of the estimation kernel defined above, and for the estimation of the temporal derivative a $3 \times 3 \times 5$ spatio-temporal neighborhood was used. The flow field obtained as a result of estimation was post-processed using a 5×5 median filter and the threshold operator on confidence measure with the threshold value varying between 20 and 50.

Synthetic Images. The synthetic image sequences were generated in the following way: a reflectivity function was defined for a ball, then a light source was defined by its longitude and tilt and the sensitivity functions of the three light detectors were defined. Images obtained by a simulation of the ball moving in space together with light reflection and image generation processes were then artificially corrupted by 3% noise and quantized.

Figures 2–4 demonstrate some statistics on errors for various displacements and different kinds of motion. As expected, all three methods produced similar results for translation, but if a more complex motion was considered, the method using the brightness conservation assumption failed to produce accurate estimates. Both the normalized *rgb* and the HSV systems provided good results for all
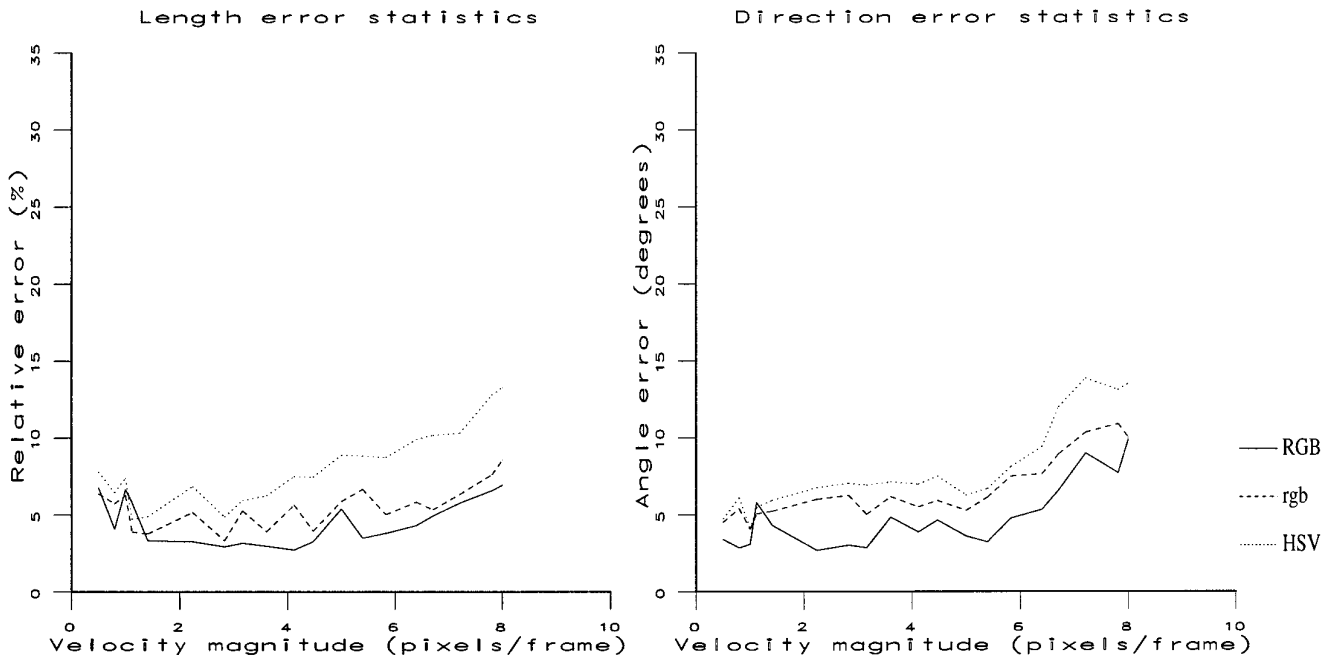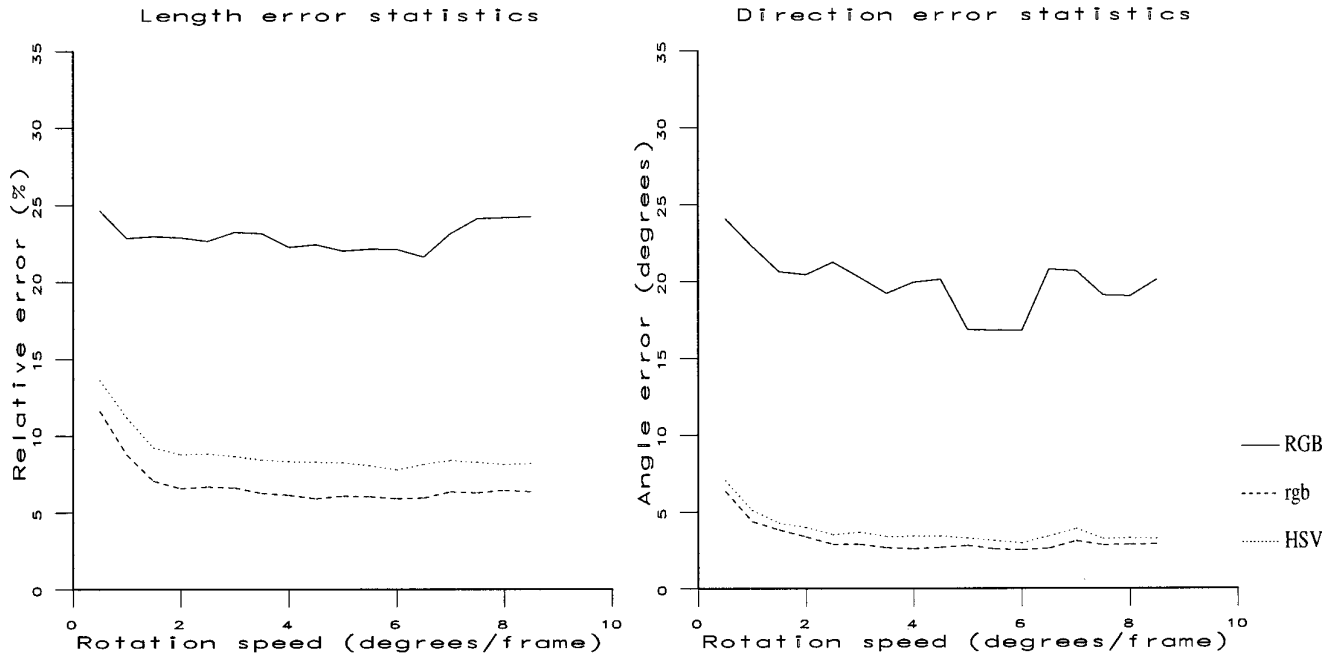
FIG. 2. Error statistics for the ball moving in the plane parallel to the image plane. Results for three different methods are reported, using RGB, normalized *rgb*, ahd HSV quantities. All the methods demonstrate similar performance.

kinds of motion, but the error rate of the method using the HSV functions was slightly higher than that of the normalized *rgb*. For any particular application both sets of color functions are to be tested in order to select the most suitable one for the particular sequences.

Figures 5–8 illustrate some of the experiments. Two images from the sequence and expected image flow are shown, and then the estimates are reported for three different methods, using RGB, normalized *rgb*, and HSV functions. For each method the estimated flow, the difference

FIG. 3. Error statistics for the ball rotating around the axis normal to the image plane. Results for three different methods are reported, using RGB, normalized *rgb*, and HSV quantities. The error rate of the method using brightness functions is much higher than those using color functions.
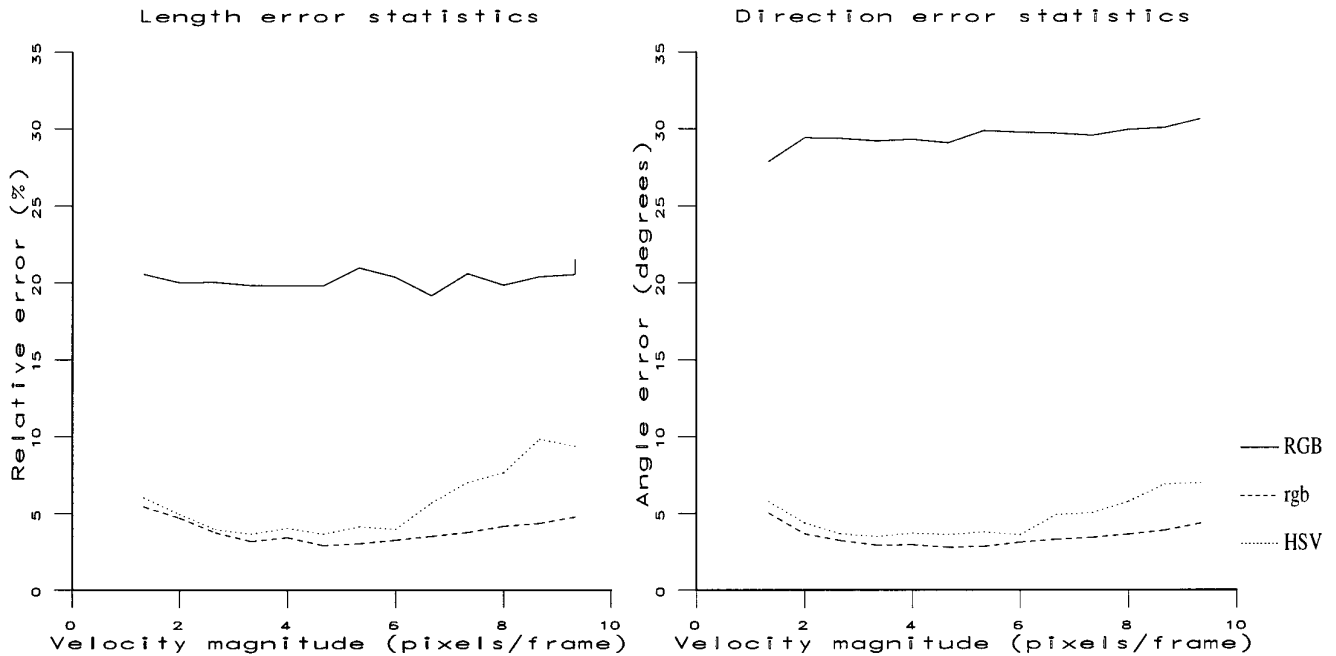
Length error statistics

Direction error statistics

FIG. 4. Error statistics for the ball moving toward the camera. Results for three different methods are reported, using RGB, normalized *rgb*, and HSV quantities. The error rate of the method using brightness is much higher than those using color functions.

between the expected and the estimated fields, and the confidence measure are shown. All the flow fields are displayed using needle diagrams; the length of a needle at a certain point in the image is proportional to the velocity magnitude at this point and the needle direction is identical to that of the velocity vector. The confidence measure is shown using black-and-white images, where the gray level at a certain point of the image is equal to the condition number of the coefficient matrix at this point. Hence, the darker a certain region is in the confidence measure image, the lower condition number it represents, the higher confidence we have in the velocity estimates in this region. It can be seen that in most of the image area, except edges (almost white in the confidence measure image), the confidence was considerably high (dark regions in the confidence measure image).

Edges present a problem for the color-based algorithm. At and near edges, both color functions have a very sharp transition in the direction normal to the edge direction; therefore the gradient vectors of the color functions are almost parallel, which results in high values for the condition number of the coefficient matrix. This is the case where color does not add sufficient information to disambiguate the aperture problem; therefore only the normal flow, the velocity component parallel to the gradient vector (normal to the edge), can be reliably determined.

Real Images. Real images were obtained using a color camera. An object with some color pattern on it was attached to a robot hand, which was moving in various ways while the image sequences were taken. Since it is very difficult to determine the true image flow for real images [15], usually only qualitative testing is performed [1].

Figures 9–11 illustrate some of the experiments carried out in order to test the proposed methods. The results are reported similarly to the experiments using synthetic images: flow fields are shown using needle diagrams, and the confidence measure of the estimates is displayed on the gray-level images with dark areas corresponding to high confidence and light areas corresponding to the confidence.

Note that while in the synthetic sequences a color gradient was generated also on the background and hence good estimates were obtained for the whole image, in the real images the background was uniformly painted and the confidence measure in the background region was quite poor. The estimation obtained in this region should not be taken into account and a zero motion field was assumed there. It is interesting to note that in this case the brightness-based method would also fail to produce reliable results. The reason for this failure is that a uniformly painted plane is "impossible to track," i.e., both its color and brightness functions are constant and hence cannot be used for motion estimation. This case is similar to the example of a uniformly painted ball discussed in the first section of the paper.
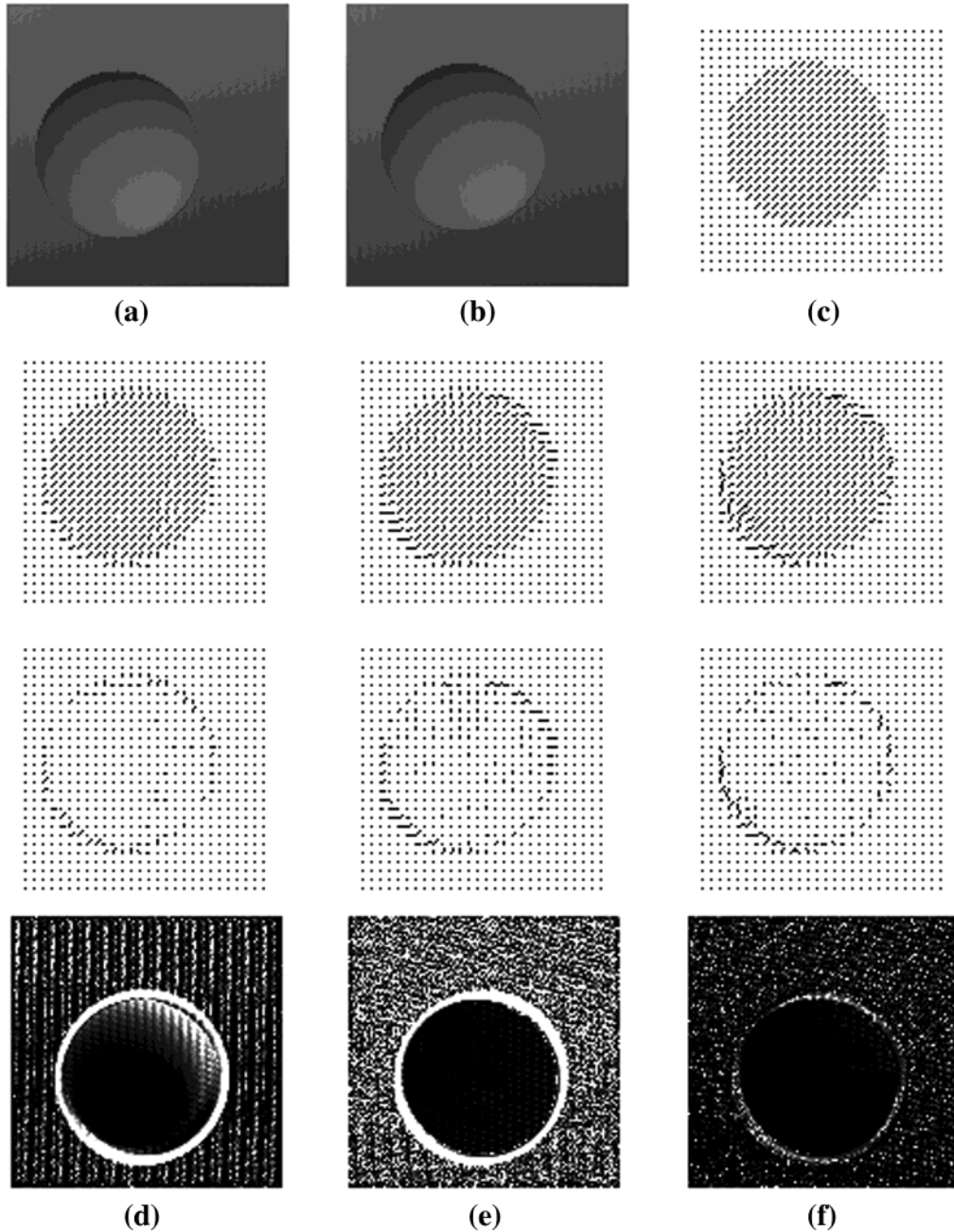
FIG. 5. Synthetic images. The ball velocity is (1, 1) pixel/frame in the plane parallel to the image plane. In the first line (a) the first frame of the sequence, (b) the eighth frame of the sequence, and (c) the true image flow are shown. For each method the estimated field, the error, and the confidence measure are reported: (d) the brightness (R, G, B), (e) the normalized RGB, and (f) the HSV quantities.

5. DISCUSSION AND CONCLUSIONS

It has been demonstrated in this work that color images, as opposed to black-and-white ones, provide reliable information for motion estimation. Two different approaches for motion estimation using color images have been discussed and tested.

The first approach proposed in this work considers a color image as a set of three different black-and-white images. The brightness conservation assumption can be applied to each one of the color image components, which leads to an overdetermined system of three linear equations for the velocity vector components at each point of the image. This approach provides sufficient quality for
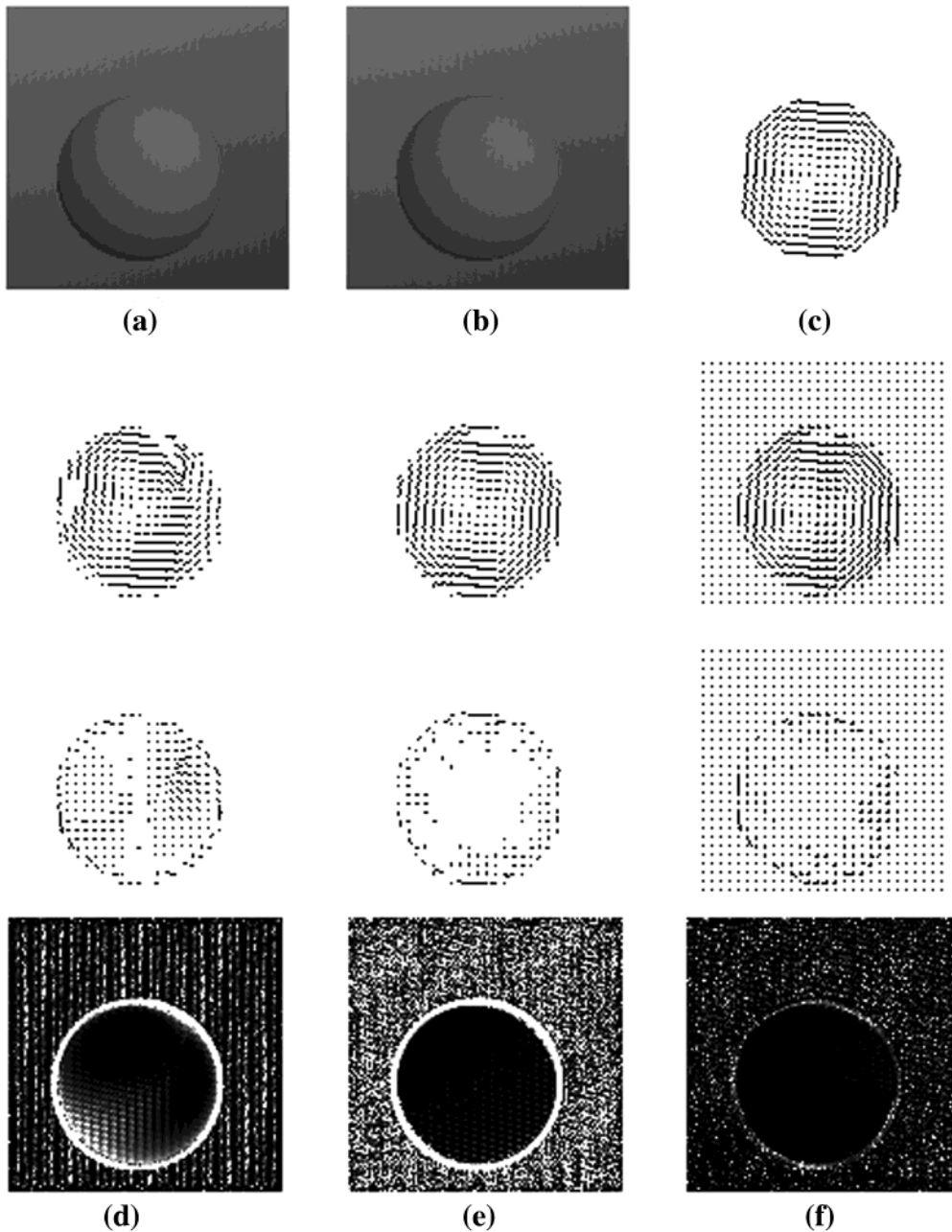
FIG. 6. Synthetic images. The ball is rotating at the angular speed of 2°/frame in the plane parallel to the image plane. In the first line (a) the first frame of the sequence, (b) the eighth frame of the sequence, and (c) the true image flow are shown. For each method the estimated field, the error, and the confidence measure are reported: (d) the brightness (R, G, B), (e) the normalized RGB, and (f) the HSV quantities.

velocity estimates when the object undergoes translations in a plane parallel to the image plane, but if more complex kinds of motion are involved, this method, like any other using the brightness conservation assumption, produces estimates with significant errors. Such a behavior is to be expected due to the inherent assumption underlying this method, since the brightness function is not truly invariant under many types of complex motions.

In order to improve quality of motion estimation, a new approach has been proposed. It uses color functions for motion estimation. These functions are extracted from the brightness functions of the image and represent color properties of an object surface under a certain illumination. Assuming that the illumination spectrum is locally constant, these quantities are invariant under any kind of motion and therefore allow us to obtain much better estimates
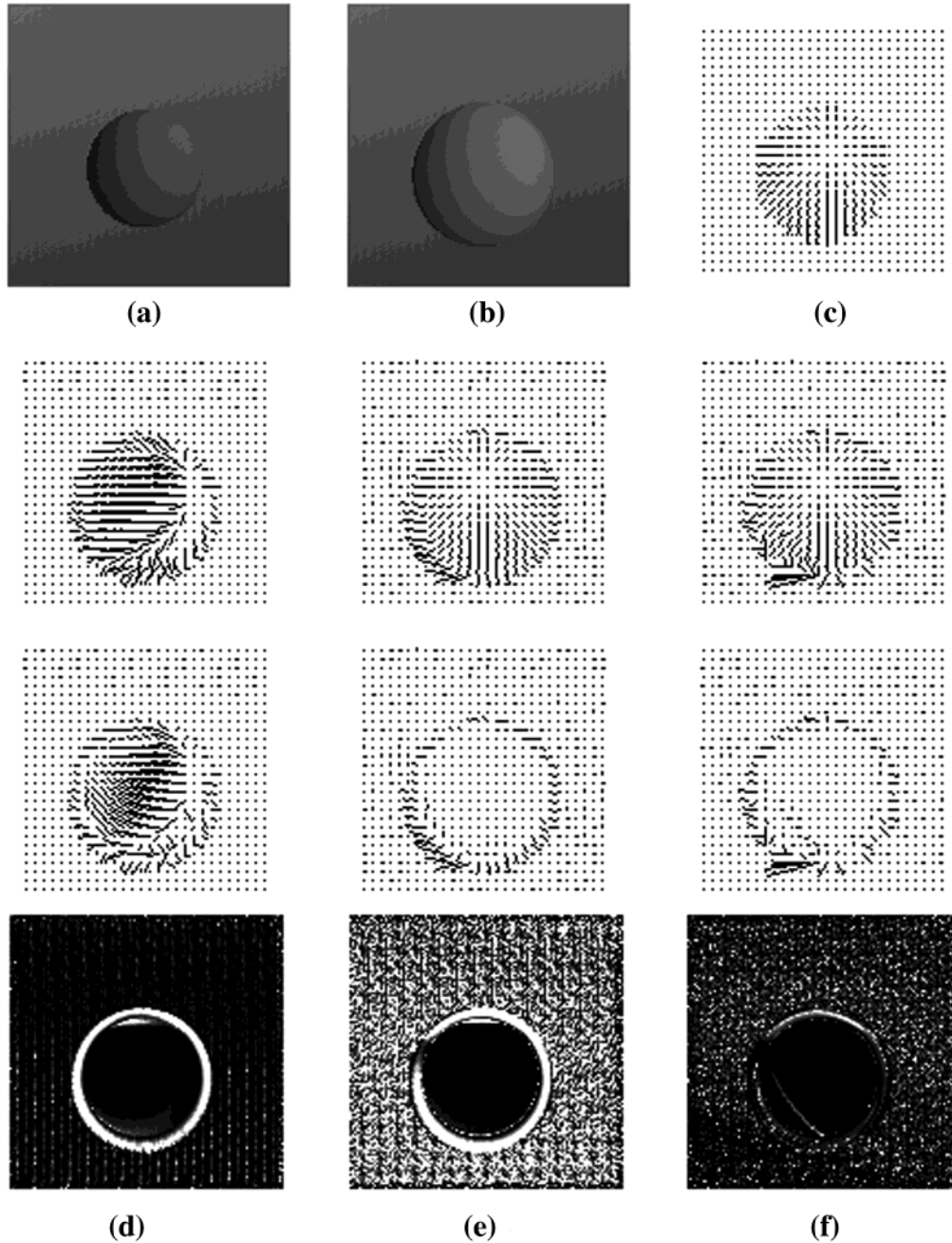
FIG. 7. Synthetic images. The ball is moving toward the camera at the speed of 10 pixels/frame. In the first line (a) the first frame of the sequence, (b) the eighth frame of the sequence, and (c) the true image flow are shown. For each method the estimated field, the error, and the confidence measure are reported: (d) the brightness (R, G, B), (e) the normalized RGB, and (f) the HSV quantities.

than the brightness functions. Two different sets of the color functions, invariant under motion, were used: the normalized RGB and the HSV color representations. Experimental results obtained by using these sets of the color functions were compared to the estimates obtained by using the brightness functions. These experiments confirmed our expectations that color functions, being directly related

to surface properties of objects, provide much more precise information about the object motion.

Both representations use certain assumptions on illumination and the reflectivity properties of the object surface. If these assumptions fail (e.g., for specular surfaces), a more general model should be used. If a general color constancy model were available, it would be used to obtain
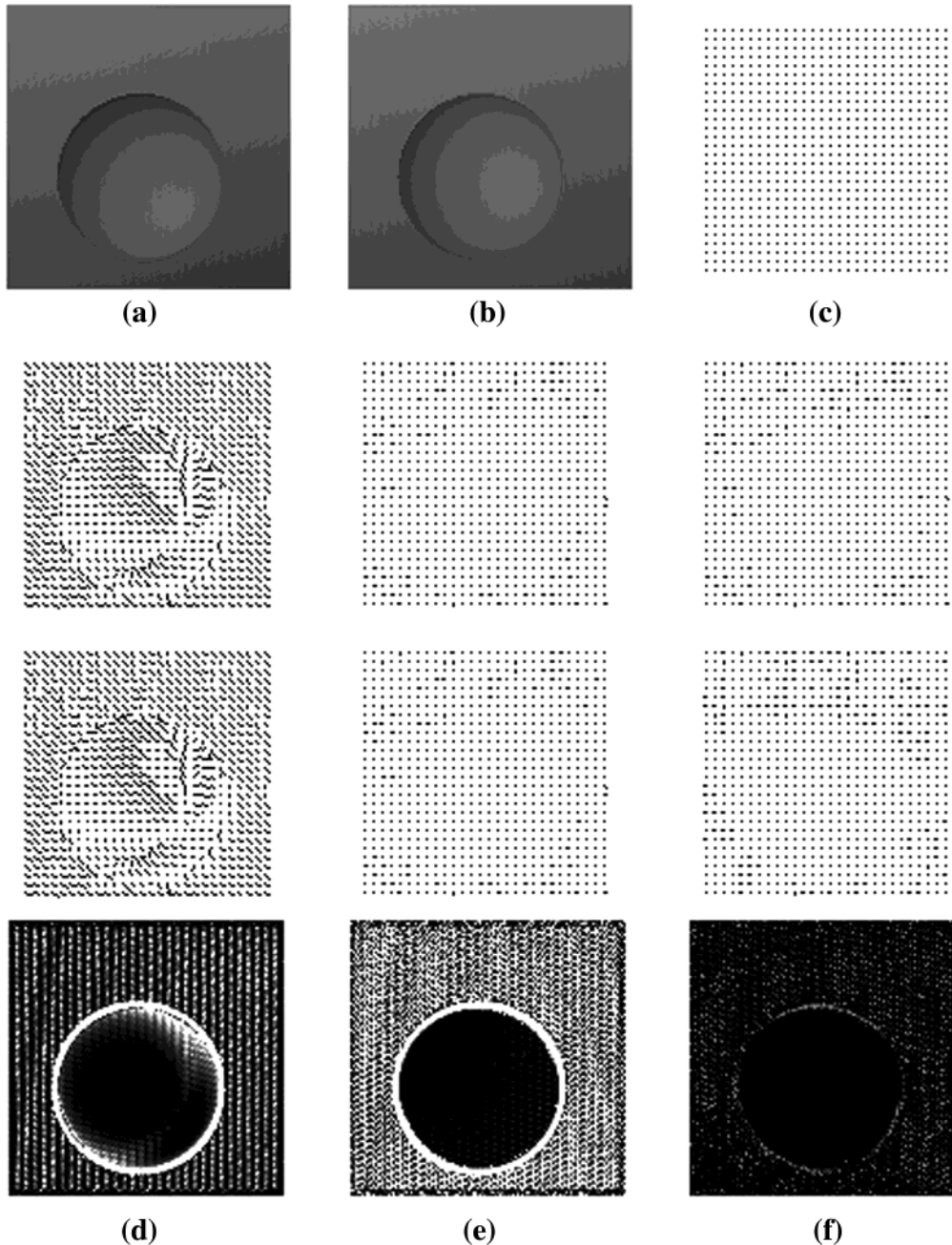
FIG. 8. Synthetic images. The light source is moving so that its longitude is changing with the speed of 3°/frame. In the first line (a) the first frame of the sequence, (b) the eighth frame of the sequence, and (c) the true (zero) image flow are shown. For each method the estimated field, the error, and the confidence measure are reported: (d) the brightness (R, G, B), (e) the normalized RGB, and (f) the HSV quantities.

invariant color descriptors. The normalized RGB and the HSV color representations, which were used in this paper, are good enough approximations of truly invariant color characteristics. They allow one to obtain velocity estimates with sufficient accuracy in most cases.

In this work the multiple constraint method was used to estimate image flow. But we note that the color functions

can be used also in conjunction with any other gradient-based approach for motion estimation. In the second section of this paper the existing methods using gradient based approach were discussed. Each of these assumes brightness conservation and then uses the brightness function to estimate optical flow. Given the color functions, each of these methods can be implemented using the color invariance
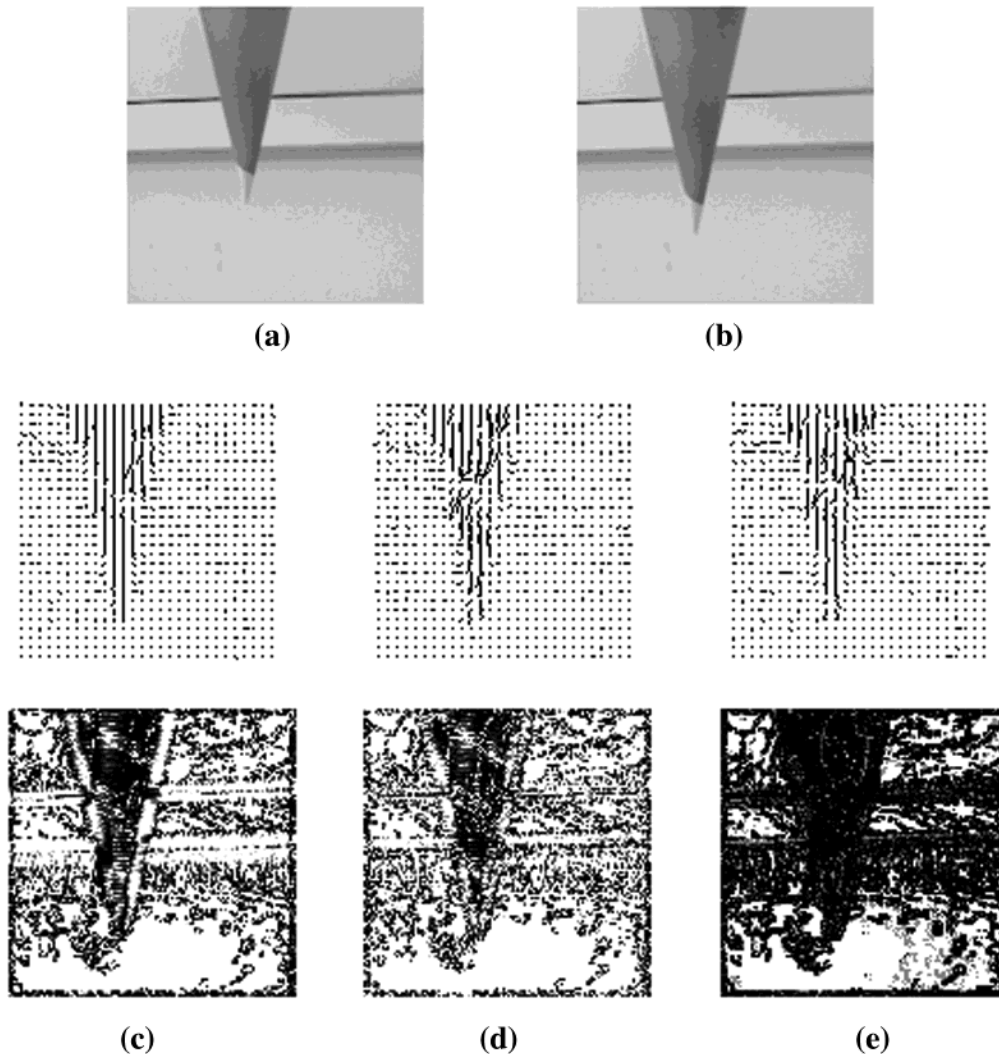
FIG. 9. Real images. The object is moving in the plane approximately parallel to the image plane. In the first line (a) the first frame of the sequence and (b) the eighth frame of the sequence are shown. For each method the estimated field and the confidence measure are reported: (c) the brightness (R, G, B), (d) the normalized RGB, and (e) the HSV quantities.

under motion, which will result in better estimation of complex types of motion. User requirements should be carefully considered when a certain gradient-based method is selected for implementation. For example, a multiple constraint approach for motion estimation uses entirely local information, and therefore the estimates are less stable than the ones produced using some neighborhood information. If black-and-white images are considered, the neighborhood-sampling approach provides a more stable solution than the multiple constraint approach, and hence it will also produce better results if the color functions are used instead of brightness. Therefore, if one is interested in higher accuracy and stability, the neighborhood-sampling approach should be considered, but using the color functions and not the brightness functions.

In certain regions of the image (especially for man-made objects), the color gradient is not strong enough to provide reliable information for estimation. In this case brightness information can and should be used in addition to color to obtain velocity estimates based on the stronger assumption of brightness conservation. Even though more precise estimates can be obtained using color in presence of the sufficient color gradient, brightness information definitely should not be ignored, but used as an additional cue for estimation when the color based method fails to produce reliable results.

Experimental testing confirmed that the new proposed methods provide good estimates of the image flow in regions with considerable gradient of color, whereas in regions of uniform color these methods failed to produce
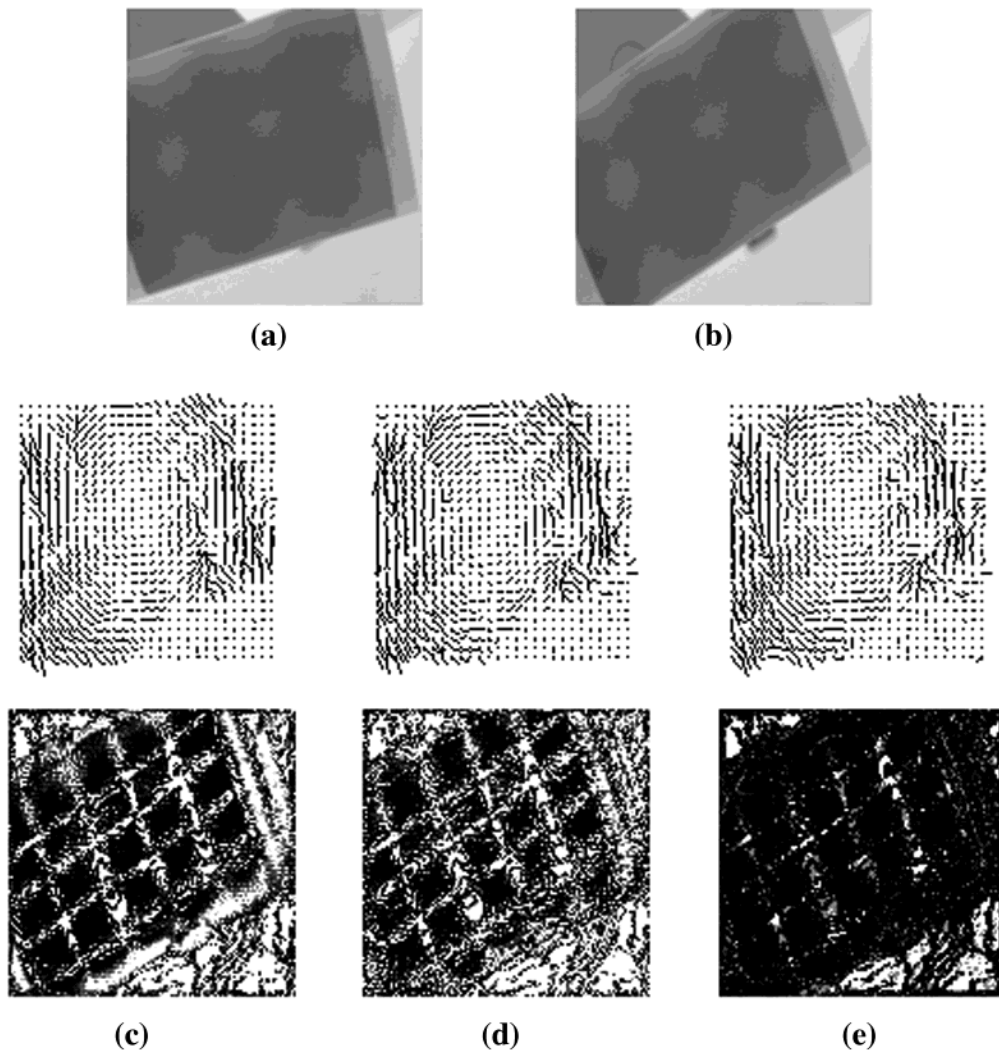
FIG. 10. Real images. The object is rotating in the plane approximately parallel to the image plane. In the first line (a) the first frame of the sequence and (b) the eighth frame of the sequence are shown. For each method the estimated field and the confidence measure are reported: (c) the brightness (R, G, B), (d) the normalized RGB, and (e) the HSV quantities.

reliable results. This defines the domain of possible applications for the methods using color for motion estimation.

APPENDIX: COLOR REPRESENTATION

The most natural way to represent the color perceived by a vision system is by the (R, G, B) triplet, which is the output of the three types of light detectors. This approach, however, does not represent intrinsic color in the best way, since the chromatic properties of the spectrum are not separated from the non-chromatic ones, such as brightness. A number of representations have been developed to separate the intensity and the color characteristics. Two such systems are presented below and used in this work for motion estimation.

5.1. Normalized RGB Representation

The normalized RGB representation uses the (R, G, B) values divided by their sum:

$$r = \frac{R}{R + G + B}, \quad g = \frac{G}{R + G + B}, \quad b = \frac{B}{R + G + B}. \quad (19)$$

These are clearly not independent quantities—given two of them the third one can be determined; therefore two independent parameters are enough to describe uniquely a point in the space of $[r, g, b]^T$ vectors (which is actually a two-dimensional space). This representation does not consider the intensity of light having a given spectrum, all possible spectra being normalized. A graphic illustration of
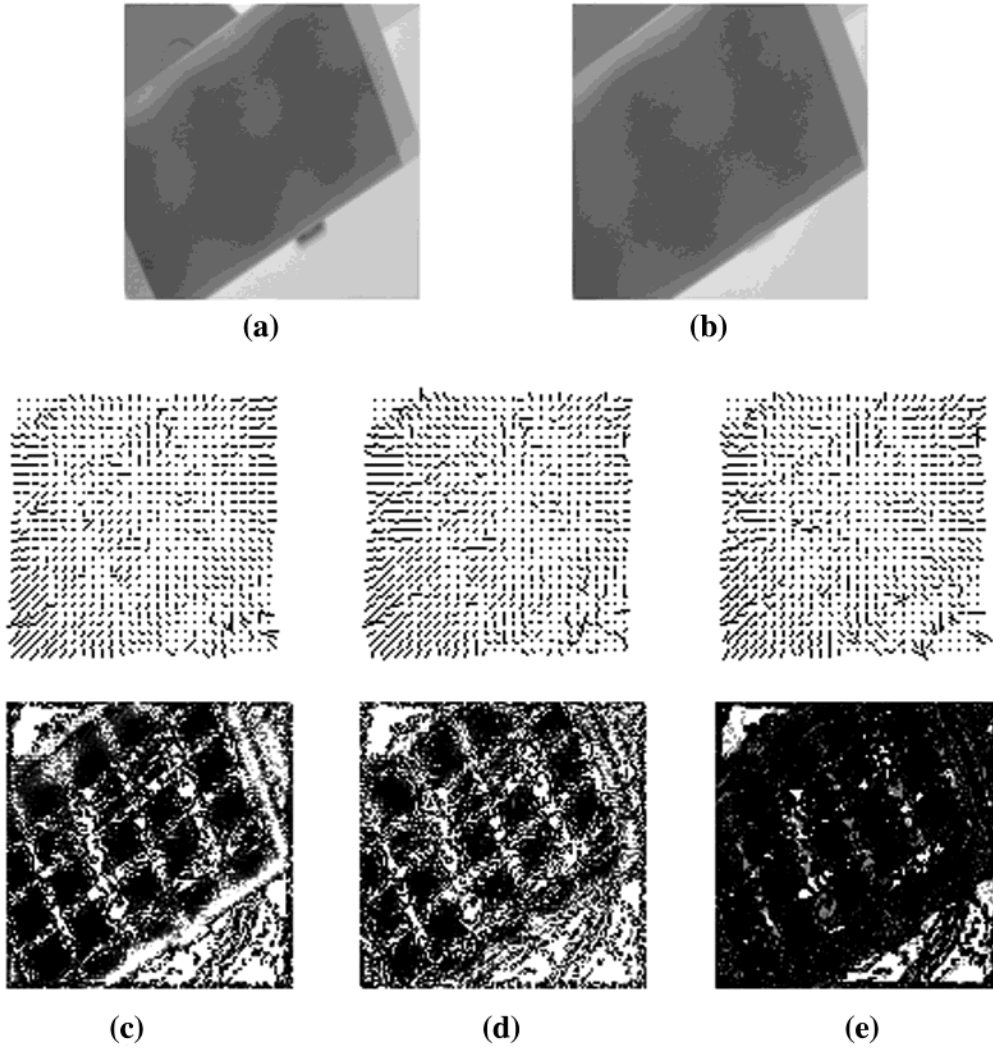
FIG. 11. Real images. The object is moving toward the camera. In the first line (a) the first frame of the sequence and (b) the eighth frame of the sequence are shown. For each method the estimated field and the confidence measure are reported: (c) the brightness (R, G, B), (d) the normalized RGB, and (e) the HSV quantities.

the normalized RGB system is the so-called "chromaticity diagram," which is a graph of (r, g) pairs corresponding to different colors (Fig. 12a). All feasible colors are represented by points inside the convex region defined by the "pure color curve"—the set of the points representing pure colors. A color represented by a point N on the line passing through white and any pure color M can be generated by mixing this pure color with white in quantities proportional to the ratio of lengths of the line segments NM and ON.

5.2. HSV Representation

The HSV representation uses three other values to define color—*hue*, *saturation*, and *value*. While *value* is an intensity measure and corresponds to non-chromatic light characteristics, *hue* and *saturation* are chromaticity parameters, encoding the color information. *Saturation* is a measure of pure color in a certain spectrum (ratio between pure color and white light), and *hue* encodes the color of wavelength information.

The mathematical formulae relating RGB and HSV systems are as follows:
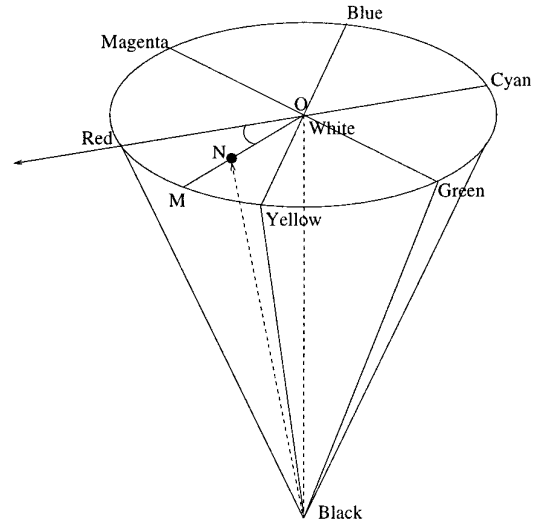
$$Value = \text{Max}(R, G, B),$$

$$Saturation = \frac{\text{Max}(R, G, B) - \text{Min}(R, G, B)}{\text{Max}(R, G, B)},$$

$$Hue = \text{if } R = \text{Max}(R, G, B)$$
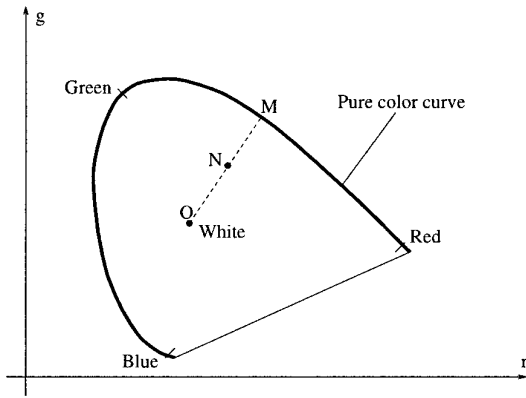
(a) Chromaticity diagram.   (b) HSV cone.

FIG. 12. Graphical representation of the normalized RGB and the HSV systems.

then $\dfrac{G - B}{\text{Max}(R, G, B) - \text{Min}(R, G, B)}$ (20)

if $G = \text{Max}(R, G, B)$

then $2 + \dfrac{B - R}{\text{Max}(R, G, B) - \text{Min}(R, G, B)}$

if $B = \text{Max}(R, G, B)$

then $4 + \dfrac{R - G}{\text{Max}(R, G, B) - \text{Min}(R, G, B)}$.

The graphical representation of Fig. 12b is useful in understanding this system. All colors are represented by vectors with endpoints inside the base of the cone and origins on the axis of the cone. Pure colors have their endpoints on the perimeter of the cone base. The closer a certain vector endpoint N is to the perimeter of the cone base, the more saturated the color it represents. Hue is defined by the angle between the radius passing through the endpoint of the color vector ON and the radius of red.

REFERENCES

1. J. Barron, D. Fleet, and S. Beauchemin, Performance of optical flow techniques, *Int. J. Comput. Vis.* **12**(2), 1994, 43–77.

2. G. Buchsbaum, A spatial processor model for object color perception, *J. Franklin Inst.* **310**, 1980.

3. M. Campani and A. Verri, Computing optical flow from an overconstrained system of linear algebraic equations, in *Proceedings of the Third International Conference on Computer Vision ICCV'90, Osaka, Japan, 1990,* pp. 22–26.

4. G. Finlayson, B. Funt, and K. Barnard, Color constancy under varying illumination, in *Proceedings of The 5th International Conference on Computer Vision, Cambridge, MA, 1995,* pp. 720–725.

5. R. Haralick and J. Lee, The facet approach to optic flow, in *Proceedings, Image Understanding Workshop, Arlington, VA, 1983.* (L. S. Baumann, Ed.), pp. 84–93.

6. B. Horn, Exact reproduction of colored images, *Comput. Vision Graphics Image Process.* **26,** 1984, 135–167.

7. B. Horn and B. Schunck, Determining optical flow, *Artificial Intelligence* **17,** 1981, 185–203.

8. A. Jennings, *Matrix Computation for Engineers and Scientists.* Wiley, New York, 1980.

9. B. Lucas and T. Kanade, An iterative image registration technique with an application to stereo vision, in *Proceedings of DARPA Image Understanding Workshop, 1981,* pp. 121–130.

10. A. Mitiche, Y. Wang, and J. Aggarwal, Experiments in computing optical flow with the gradient-based, multiconstraint method, *Pattern Recognition* **20**(2), 1987, 173–179.

11. H.-H. Nagel, Constraints for the estimation of displacement vector fields from image sequences, in *Proceedings IJCAI-83, Karlsruhe, Germany, 1983,* pp. 945–951.

12. H.-H. Nagel, On the estimation of optical flow: Relations between different approaches and some new results, *Artificial Intelligence* **33,** 1987, 299–324.

13. N. Ohta, Optical flow detection by color images, in *Proceedings of IEEE International Conference on Image Processing, Pan Pacific, Singapore, 1989,* pp. 801–805.

14. J. Ortega, *Numerical Analysis: A Second Course,* SIAM Press, Philadelphia, 1990.

15. M. Otte and H.-H. Nagel, Optical flow estimation: Advances and comparisons, in *Lecture Notes in Computer Science,* Vol. 800, pp. 51–60, Springer-Verlag, New York/Berlin, 1994.

16. S. Shafer, Using color to separate reflection components, *Color: Research and Application* **10**(4), 1985, 210–218.

17. A. Singh, *Optic Flow Computation: A Unified Perspective,* IEEE Computer Society Press, Los Alamitos, CA, 1991.

18. O. Tretiak and L. Pastor, Velocity estimations from image sequences with second order differential operators, in *Proceedings, International Conference on Pattern Recognition, Montreal, Canada, 1984,* pp. 16–19.

19. H. Trussel, DSP solutions run the gamut for color systems, *IEEE Signal Process. Mag.,* April 1993, pp. 8–23.

20. A. Verri and T. Poggio, Against quantitative optical flow, in *Proceedings of First International Conference on Computer Vision, London, 1987,* pp. 171–180.

21. K. Wohn, L. Davis, and P. Thrift, Motion estimation based on multiple local constraints and nonlinear smoothing, *Pattern Recognition* **16**(6), 1983, 563–570.