Robust Segmentation and Tracking of Colored Objects in Video

Theo Gevers, *member, IEEE*

Abstract— **Segmenting and tracking of objects in video is of great importance for video-based encoding, surveillance and retrieval. However, the inherent difficulty of object segmentation and tracking is to distinguish changes in the displacement of objects from disturbing effects such as noise and illumination changes.**

Therefore, in this paper, we formulate a colour-based deformable model which is robust against noisy data and changing illumination. Computational methods are presented to measure colour constant gradients. Further, a model is given to estimate the amount of sensor noise through these color constant gradients. The obtained uncertainty is subsequently used as a weighting term in the deformation process.

Experiments are conducted on image sequences recorded from 3D scenes. From the experimental results it is shown that the proposed colour constant deformable method successfully finds object contours robust against illumination, and noisy, but homogeneous regions.

Keywords- **object segmentation, object tracking, video, deformable models, colour, colour constancy, noise models, multi-valued gradients.**

I. Introduction

The segmentation and tracking of real-world 3D objects in video sequences is of great importance for video encoding, surveillance and retrieval, and significant progress has been made [1], [2], [5], [6], [12], [16], [21], [23]. The inherent difficulty of video-based object tracking is that as a rigid object moves in a 3D scene then its shape becomes a perspective projection on the image frames. In general, to achieve robust object tracking, geometric models are used for this shape transformation. In the simplest case, the shape of the object undergoes a translational transformation from one frame to another in which a simple cross-correlation technique will suffice. The translational model can be extended to include rotation, scaling, shearing, up to affine transformation [3]. When sufficient feature points (i.e. edges and corners) are available, objects can be tracked reliably in this manner. However, for non-rigid objects, such as humans, object motion is more complex than affine transformations. Therefore, Kalman filtering has been proposed for tracking of non-rigid objects. The performance of Kalman-based tracking systems is severely hampered in the presence of false observations. To reduce the effect of noisy data, the search area could be restricted [6] similar to active contours [10], [11], [13]. Active contours use edge detection to compute internal/external energies. Assuming that the object displacement between frames is small, object tracking by deformable models achieve high tracking performance. Therefore, in this paper, we focus on deformable models for object tracking. However, in all of the above object tracking approaches it is difficult to distinguish changes in the displacement of the objects which are due to real object movement from disturbing displacement effects such as noisy data and illumination changes. As a consequence, the tracking process can be distracted from the target object [23] decreasing the accuracy of the tracking process.

In this paper, we aim at formulating colour-based deformable models to segment and track objects in video robust against noisy data and varying illumination. To achieve this, computational methods are presented to measure colour constant gradients. Further, a model is proposed for the estimation of sensor noise through these color constant gradients. As a result, the associated uncertainty is known for each color constant gradient value. The associated uncertainty is subsequently used to weight the color constant gradient during the deformation process. As a result, noisy and unstable gradient information will contribute less to the deformation process than reliable gradient information yielding robust object segmentation and tracking.

The paper is organized as follows. In Section II, definitions are given on deformable contours. In Section III, computational methods are presented to integrate color and noise-robustness into these deformable contours. Finally, experiments are conducted in Section IV.

II. Definitions

Deformable models are used in the process of object segmentation and tracking by providing high-level information in the form of continuity constraints and low-level information in terms of minimum energy constraints related to image characteristics [4], [10], [15], [22], for example. In general, deformable models use low-level image characteristics based on intensity gradient information. However, intensity gradients are sensitive to illumination conditions. Therefore, our attention is focussed on the use of color information. Sapiro introduces the concept of color snakes [18], [19] using snakes (via level-sets) with gradients computed from multi-valued images. However, profound illumination effects may still introduce accidental edges such as shadow and shading edges. Also severe changes in spectral composition of the illumination may introduce artifacts. Therefore, we aim at computing colour constant gradients in a principled way to steer the deformable model to converge to object contours instead of boundaries produced by illumination changes. Further, the aim is to obtain robustness against noisy data. To this end, the associated

Theo Gevers is with the Computer Science Institute, University of Amsterdam, Kruislaan 403, 1098 SJ, Amsterdam, The Netherlands, gevers@science.uva.nl.

uncertainty is computed for the color constant gradients and integrated in the deformation scheme.

To be precise, consider a deformable contour [10]:

$$\mathbf{v}(t) = [x(t), y(t)], t \in [0, 1], \tag{1}$$

moving through the spatial domain of an image I to minimize an cost functional E associated with the curve. In fact, E is a weighted sum of internal and external energies:

$$E = \alpha E_{int} + \beta E_{ext}, \tag{2}$$

where α and β are appropriate weights.

To obtain smooth and physically feasible deformations, the internal cost is defined by an elasticity constraint as follows:

$$E_{int} = (\oint_t (||\mathbf{v}(t)'||^2 + ||\mathbf{v}(t)''||^2 dt)(\oint_t ||\mathbf{v}(t)'||dt), \tag{3}$$

where $\mathbf{v}(t)'$ and $\mathbf{v}(t)''$ denote the first and second derivatives of the curve with respect to t measuring respectively the elasticity of the curve.

The external cost is derived from the image to enable the curve to attract to salient image features (i.e. edges and corners). In most deformable contours, the intensity gradient is used giving the following external term:

$$E_{ext} = \oint_t -\nabla I(x, y)dt, \tag{4}$$

where the gradient image $\nabla I(x, y)$ is usually based on Gaussian derivatives. However, as stated above, intensity-based gradient images are dependent on the illumination conditions. Consequently, intensity gradients do not necessarily correspond to object boundaries.

Let the colour gradient be denoted by $\nabla \mathcal{C}$, then the colour-based external cost term is as follows:

$$E_{ext} = \oint_t -\nabla \mathcal{C}(x, y)dt. \tag{5}$$

Our aim is to measure colour gradient ∇C discounting illumination and which is robust against noisy data.

III. Colour-based Deformable Models

First, in section III-A, computational methods are presented to measure colour constant gradients. Then, in section III-B, a model is proposed for the estimation the uncertainty of these color constant gradients. The obtained uncertainty is used as a weighting term in the deformation process.

A. Illumination Invariant Derivatives

Consider the reflection model with narrow-band filters [17]:

$$C_k(\vec{x}) = G_B(\vec{x}, \vec{n}, \vec{s})E(\vec{x}, \lambda_k)B(\vec{x}, \lambda_k), \tag{6}$$

where $G_B(\vec{x}, \vec{n}, \vec{s})$ is the geometric function dependent on the surface orientation $\vec{n}$ and illumination direction $\vec{s}$ at position $\vec{x}$. Further, $E(\vec{x}, \lambda_k)$ is the illumination and $B(\vec{x}, \lambda_k)$ is the surface albedo at wavelength λ_k. Various illumination-independent color ratios have been proposed [8], [14]. These color ratios are derived from neighboring points. A drawback, however, is that these color ratios might be negatively affected by the geometry and pose of the object.

Therefore, we focus on the following color ratio [9]:

$$M(C^1_{\vec{x}_1}, C^1_{\vec{x}_2}, C^2_{\vec{x}_1}, C^2_{\vec{x}_2}) = \frac{C^1_{\vec{x}_1} C^2_{\vec{x}_2}}{C^1_{\vec{x}_2} C^2_{\vec{x}_1}}, C^1 \neq C^2, \tag{7}$$

expressing the color ratio between two neighboring image locations, for $C^1, C^2 \in \{C^1, C^2, ..., C^N\}$ giving the measured sensor pulse response at different wavelengths, where $\vec{x}_1$ and $\vec{x}_2$ denote the image locations of the two neighboring pixels.

For a standard RGB color camera, we have:

$$m_1(R_{\vec{x}_1}, R_{\vec{x}_2}, G_{\vec{x}_1}, G_{\vec{x}_2}) = \frac{R_{\vec{x}_1} G_{\vec{x}_2}}{R_{\vec{x}_2} G_{\vec{x}_1}}, \tag{8}$$

$$m_2(R_{\vec{x}_1}, R_{\vec{x}_2}, B_{\vec{x}_1}, B_{\vec{x}_2}) = \frac{R_{\vec{x}_1} B_{\vec{x}_2}}{R_{\vec{x}_2} B_{\vec{x}_1}}, \tag{9}$$

$$m_3(G_{\vec{x}_1}, G_{\vec{x}_2}, B_{\vec{x}_1}, B_{\vec{x}_2}) = \frac{G_{\vec{x}_1} B_{\vec{x}_2}}{G_{\vec{x}_2} B_{\vec{x}_1}}. \tag{10}$$

The color ratio is independent of the illumination, a change in viewpoint, object geometry as shown by substituting eq. (6) in eq. (7):

$$M(C^1_{\vec{x}_1}, C^1_{\vec{x}_2}, C^2_{\vec{x}_1}, C^2_{\vec{x}_2}) = \frac{C^1_{\vec{x}_1} C^2_{\vec{x}_2}}{C^1_{\vec{x}_2} C^2_{\vec{x}_1}} = \frac{B(\vec{x_1}, \lambda_{C_1})B(\vec{x_2}, \lambda_{C_2})}{B(\vec{x_2}, \lambda_{C_1})B(\vec{x_1}, \lambda_{C_2})}. \tag{11}$$

For the ease of exposition, we concentrate on m_1 based on the RG-color bands in the following discussion. Without loss of generality, all results derived for m_1 will also hold for m_2 and m_3.

Taking the natural logarithm of both sides of eq. (8) results for m_1 in:

$$\ln m_1(R_{\vec{x}_1}, R_{\vec{x}_2}, G_{\vec{x}_1}, G_{\vec{x}_2}) = \ln(\frac{R_{\vec{x}_1} G_{\vec{x}_2}}{R_{\vec{x}_2} G_{\vec{x}_1}}) =$$

$$\ln R_{\vec{x}_1} + \ln G_{\vec{x}_2} - \ln R_{\vec{x}_2} - \ln G_{\vec{x}_1} = \ln(\frac{R_{\vec{x}_1}}{G_{\vec{x}_1}}) - \ln(\frac{R_{\vec{x}_2}}{G_{\vec{x}_2}}) \tag{12}$$

Hence, the color ratios can be seen as differences at two neighboring locations $\vec{x}_1$ and $\vec{x}_2$ in the image domain of the logarithm of R/G:

$$\nabla \mathcal{C}_{m_1}(\vec{x}_1, \vec{x}_2) = (\ln(\frac{R}{G}))_{\vec{x}_1} - (\ln(\frac{R}{G}))_{\vec{x}_2} \tag{13}$$

By taking these differences between neighboring pixels, the derivative is obtained of the logarithm of image R/G which is independent of the illumination color, and also a change in viewpoint, the object geometry, and illumination intensity. We have taken the gradient magnitude by applying Canny's edge detector (derivative of the Gaussian) on image $\ln(R/G)$ with non-maximum suppression in a standard

way to obtain gradient magnitudes at local edge maxima denoted by $\nabla\mathcal{C}_{m_1}(\vec{x})$. The results obtained so far for m_1 hold also for m_2 and m_3, yielding (leaving out the spatial coordinates for illustration simplicity):

$$\nabla\mathcal{C}_{m_1m_2m_3} = (\nabla\mathcal{C}_{m_1}, \nabla\mathcal{C}_{m_2}, \nabla\mathcal{C}_{m_3}) \tag{14}$$

For pixels on a uniformly colored region (i.e. with fixed surface albedo), in theory, all three components will be zero whereas at least one the three components will be non-zero for pixels on locations where two regions of distinct surface albedo meet.

B. Noise Robustness of Illumination Invariant Derivatives

The above defined color ratios become unstable when intensity is low. In fact, these color ratios are undefined at the black point ($R = G = B = 0$) and they become very unstable near this singularity, where a small perturbation in the RGB-values (e.g. due to noise) will cause a large jump in the transformed values. For example, consider neighboring pixels having the values $R_{\vec{x}_1} = 1, R_{\vec{x}_2} = 1, G_{\vec{x}_1} = 2, G_{\vec{x}_2} = 2$ (i.e. low intensity) and another neighboring pixel-combination having $R_{\vec{x}_1} = 201, R_{\vec{x}_2} = 201, G_{\vec{x}_1} = 202, G_{\vec{x}_2} = 202$ (i.e. high intensity) on the range $[0, ..., 255]$. Then these pixels have the same color ratios, for example $m_1(1, 1, 2, 2) = 1$ is equal to $m_1(201, 201, 202, 202) = 1$. However, if we consider a minimal value change in RGB due to noise, e.g. $R_{\vec{x}_2} = 2$ (instead of $R_{\vec{x}_2} = 1$), then this value change will causes a large jump in the corresponding color ratios $m_1(1, 2, 2, 2) = 0.5$ which is different from $m_1(201, 202, 202, 202) = 0.99$.

As a consequence, false color constant gradients derived from the color ratios, are introduced due to sensor noise. We aim at providing a framework to determine the uncertainty for the color constant gradients which is subsequently used as a weighting term in the deformation process as follows.

Additive Gaussian noise is widely used to model thermal noise and is the limiting behavior of photon counting noise and film grain noise. Therefore, in this paper, we assume that sensor noise is normally distributed.

Then, for an indirect measurement, the true value of a measurand u is related to its N arguments, denoted by u_j, as follows

$$u = q(u_1, u_2, \cdots, u_N) \tag{15}$$

Assume that the estimate $\hat{u}$ of the measurand u can be obtained by substitution of $\hat{u}_j$ for u_j. Then, when $\hat{u}_1, \cdots, \hat{u}_N$ are measured with corresponding standard deviations $\sigma_{\hat{u}_1}, \cdots, \sigma_{\hat{u}_N}$, we obtain [20]

$$\hat{u} = q(\hat{u}_1, \cdots, \hat{u}_N). \tag{16}$$

Then, if the uncertainties in $\hat{u}_1, \cdots, \hat{u}_N$ are independent, random and relatively small, the predicted uncertainty in q is given by [20]

$$\sigma_q = \sqrt{\sum_{j=1}^{N} (\frac{\partial q}{\partial \hat{u}_i}\sigma_{\hat{u}_i})^2} \tag{17}$$

the so-called squares-root sum method. Although (17) is deduced for random errors, it is used as an universal formula for various kinds of errors.

Focusing on the first derivative, the substitution of (12) in (17) gives the uncertainty for the illumination invariant coordinates

$$\sigma_{\nabla_{m_1}}(\vec{x}_1, \vec{x}_2) = \sqrt{\frac{\sigma^2_{R_{\vec{x}_1}}}{R^2_{\vec{x}_1}R^4_{\vec{x}_2}G^4_{\vec{x}_1}} + \frac{\sigma^2_{G_{\vec{x}_1}}}{G^2_{\vec{x}_1}R^4_{\vec{x}_2}G^4_{\vec{x}_1}} + \frac{\sigma^2_{R_{\vec{x}_2}}}{G^2_{\vec{x}_2}R^4_{\vec{x}_2}G^4_{\vec{x}_1}} + \frac{\sigma^2_{G_{\vec{x}_2}}}{G^2_{\vec{x}_2}R^4_{\vec{x}_2}G^4_{\vec{x}_1}}} \tag{18}$$

Assuming normally distributed random quantities, the standard way to calculate the standard deviations σ_R, σ_G, and σ_B is to compute the mean and variance estimates derived from a homogeneously colored surface patches in an image under controlled imaging conditions.

From the analytical study of (18), it can be derived that color ratio becomes unstable around the black point $R = G = B = 0$.

Further, to propagate the uncertainties from these color components through the Gaussian gradient modulus, the uncertainty in the gradient modulus is determined by convolving the confidence map with the Gaussian coefficients. This results from the uncertainty in sums and differences as follows [20]. If several quantities

$$\hat{u}_1, \cdots, \hat{u}_N \tag{19}$$

are measured with uncertainties

$$\sigma_{\hat{u}_1}, \cdots, \sigma_{\hat{u}_N} \tag{20}$$

to compute

$$q = \hat{u}_1 + \hat{u}_2 \cdots + (\hat{u}_{N-1} + \hat{u}_N) \tag{21}$$

then the uncertainty in the computed value of q is the sum

$$\sigma_q = \sqrt{\sigma^2_{\hat{u}_1} + \sigma^2_{\hat{u}_2 \cdots} + \sigma^2_{\hat{u}_{N-1}} + \sigma^2_{\hat{u}_N}} \tag{22}$$

As a consequence, we obtain:

$$\sigma_{\nabla\mathcal{C}_{m_1m_2m_3}} \leq \frac{\sum_i \left[(\partial c_i/\partial x)\cdot\sigma_{\partial c_i/\partial x} + (\partial c_i/\partial y)\cdot\sigma_{\partial c_i/\partial y}\right]}{\sqrt{\sum_i \left[(\partial c_i/\partial x) + (\partial c_i/\partial y)\right]}}, \tag{23}$$

where i is the dimensionality of the color space and c_i is the notation for particular color channels. In this way, the effect of measurement uncertainty due to noise is propagated through the color constant ratio gradient.

For a Gaussian distribution 99% of the values fall within a 3σ margin. If a gradient modulus is detected which exceeds $3\sigma_\nabla$, we assume that there is 1% chance that this gradient modulus corresponds to no color transition:

$$\nabla\mathcal{C}_{m_1m_2m_3} = \begin{cases} 1 & \text{if } \nabla\mathcal{C}_{m_1m_2m_3} > 3\sigma_{\nabla\mathcal{C}_{m_1m_2m_3}} \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

deriving a local threshold value (leaving out the spatial coordinates).

C. Color Invariance

Color ratio gradient $\nabla\mathcal{C}_{m_1m_2m_3}$ requires narrow-band filters to achieve full color constancy. However, general purpose color CCD cameras do not contain narrow-band filters. To this end, spectral sharpening could be applied [7] to achieve this to a large extent. However, an alternative way is to assume that the illumination has a smooth or equally distributed spectral power over the wavelengths (e.g. white light). We propose to parameterize the color invariant model by polar coordinates $\theta_1\theta_2$ derived from RGB given by [9]:

$$\theta_1 = \arctan(\frac{R}{B}), \tag{25}$$

$$\theta_2 = \arctan(\frac{G}{B}), \tag{26}$$

which are insensitive to surface orientation, illumination direction and illumination intensity [9].

Substitution of eqs. (25) - (26) in eq. (17) gives the uncertainty for the $\theta_1\theta_2$ coordinates

$$\sigma_{\theta_1} = \sqrt{\frac{R^2\sigma_B^2 + B\sigma_R^2}{(R^2+B^2)^2}} \tag{27}$$

$$\sigma_{\theta_2} = \sqrt{\frac{G^2\sigma_B^2 + B\sigma_G^2}{(G^2+B^2)^2}} \tag{28}$$

where σ_R^2, σ_G^2 and σ_B^2 denote the sensor noise variance, and σ_{θ_1} and σ_{θ_2} represent the uncertainty (standard deviation) in the normalized red and green color components, respectively. From the analytical study of eqs. (27) and (28), it can be derived that normalized color becomes unstable around the black point $R = G = B = 0$.

As $\theta_1\theta_2$ is computed from the same position they do not contain any local (spatial) information. Therefore, the gradients are computed in the $\theta_1\theta_2$ domain by applying the Canny's edge detector. To propagate the uncertainties from the color components through the Gaussian gradient modulus, the method proposed in Section III-B is used. Then, the uncertainty in the gradient modulus is determined using (23) yielding for the $\theta_1\theta_2$ color model the following color invariant gradient

$$\nabla C_{\theta_1\theta_2} = \begin{cases} \nabla\mathcal{C}_{\theta_1\theta_2} & \text{if } \nabla\mathcal{C}_{\theta_1\theta_2} > 3\sigma_{\nabla\mathcal{C}_{\theta_1\theta_2}} \\ 0 & \text{otherwise} \end{cases} \tag{29}$$

and for the standard RGB color space we obtain

$$\nabla C_{RGB} = \begin{cases} \nabla\mathcal{C}_{RGB} & \text{if } \nabla\mathcal{C}_{RGB} > 3\sigma_{\nabla\mathcal{C}_{RGB}} \\ 0 & \text{otherwise} \end{cases} \tag{30}$$

IV. Experiments

Experiments are conducted on images from video sequences recorded from 3D scenes. To this end, in Section IV-A, we focus on the segmentation of colored objects. In Section IV-B, experiments on object tracking in video is considered.

A. Object Segmentation

In this section, the deformable model for object segmentation is experimentally verified with respect to varying imaging conditions and noise. The objects considered during the experiments were recorded in 3 RGB-colour with the aid of the SONY XC-003P CCD colour camera (3 chips) and the Matrox Magic Colour frame grabber. The digitization was done in 8 bits per colour. Two light sources of average day-light colour were used to illuminate the objects in the scene. The size of the images are 128x128. The software has been implemented in C under UNIX operating system running on a SPARC-station (300 Mhz). In the experiments, the same weights of eq. (2) have been used for the shape and image feature constraints. Further, the partial derivatives are computed through Gaussian smoothed derivatives with $\sigma = 1.0$ which is arrived at through experimentation. It has proved to be effective on our test images.

Figure 1.a shows the image of a matte cube against a homogeneous background. The initial contour is shown as specified by the user (the white contour) on input. The image is clearly contaminated by illumination effects and noise. Note that the cube is painted homogeneously. As one can see, the segmentation results based on ∇I and $\nabla\mathcal{C}_{RGB}$ are negatively affected by shadows and shading. In fact, for these gradient fields it is not clear to which boundaries the deformable contour should be pulled to. As a consequence, the final contour is biased and poorly defined. In contrast, the final contours obtained by the deformable method based on $\nabla\mathcal{C}_{\theta_1\theta_2}$ gradient information is nicely pulled towards the true boundary and hence correspond properly to the material transition. Note that in Figure 1.d the initial contour has been missed partially. This an inherent problem of snakes in general where there is always a tradeoff between shape (elasticity) and image feature constraints.

So far, the quality of segmentation results for the various colour models is judged qualitatively by visual inspection. First, the ground-truth has been obtained by a human operator by carefully selecting the outline of the objects, see Figure 2.

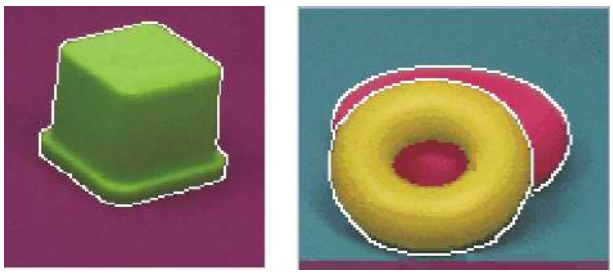

Fig. 2. The ground-truth has been obtained by a human operator by carefully selecting the outline of the objects.

To evaluate and compare the quality of object segmentation results more objectively, the mean error distance between the segmentation results and the ground-truth is taken. To be precise, let X be the image raster and a a

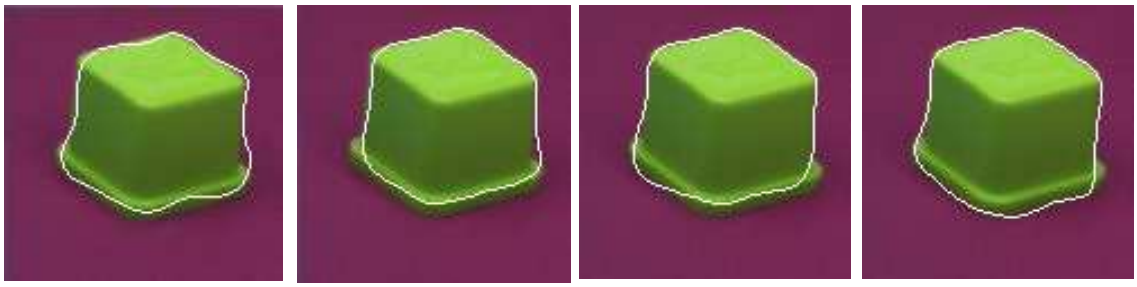

Fig. 1. From left to right a. Colour image with the initial contour as specified by the user (the white contour). b. Segmentation result based on intensity gradient field ∇I. c. Segmentation result based on RGB gradient field $\nabla \mathcal{C}_{RGB}$. d. Segmentation result based on $\theta_1\theta_2$ gradient field $\nabla \mathcal{C}_{\theta_1\theta_2}$.

Color Gradient	$\mathcal{E}$ of Fig. 1	$\mathcal{E}$ of Fig. 3
∇I	14.4	12.1
$\nabla \mathcal{C}_{RGB}$	10.4	9.3
$\nabla \mathcal{C}_{\theta_1\theta_2}$	2.5	2.4

TABLE I

Comparison of performance of snake-based segmentation differentiated for the various color models. The mean error distance between the segmentation results and the ground-truth is taken as a measure of correspondence.

binary image containing the "true" contour A defined by $A = \{\vec{x} \in X : a(\vec{x}) = 1\}$. Further, let b be a binary image, called the segmented image, containing the segmentation result $B = \{\vec{x} \in X : b(\vec{x}) = 1\}$. Let $d(\vec{x}, A)$ denote the shortest distance from pixel $\vec{x} \in X$ to A, then the mean error distance is given by:

$$\mathcal{E}(A, B) = \frac{1}{\eta(B)} \sqrt{\sum_{\vec{x} \in B} d(\vec{x}, A)^2} \tag{31}$$

The mean error distance between the segmentation results and the ground-truth yields a total average error of 14.4 pixels for ∇I, 10.4 pixels for $\nabla \mathcal{C}_{RGB}$, and 2.5 pixels for $\nabla \mathcal{C}_{\theta_1\theta_2}$ yielding promising results for $\nabla \mathcal{C}_{\theta_1\theta_2}$, see Table I. The time to compute the segmentation result was on average 49 seconds on a Ultra 10 Sparc station.

In Figure 3, an image is shown containing two plastic donuts on top of each other. Again the images are affected by shadows, shading, and inter-reflections. The segmentation results based on intensity I and colour RGB gradient are poorly defined due to the disturbing influences of the imaging conditions (mostly due to the shadows around the objects). The final contours obtained by the deformable method based on the $\nabla \mathcal{C}_{\theta_1\theta_2}$ gradient information is again nicely pulled towards the true edge. The mean error between the segmentation results of figure 3 and the ground-truth yielded a total average error of 12.1 pixels for ∇I, 9.3 pixels for $\nabla \mathcal{C}_{RGB}$, and 2.4 pixels for $\nabla \mathcal{C}_{\theta_1\theta_2}$, see Table I. The time to compute the segmentation result was on average 52 seconds on a Ultra 10 Sparc station.

B. Object Tracking

In this section, the tracking system is experimentally verified on a standard video, see figure 4. Note that, in this section, it is assumed that the object displacement between frames is small. Further, object occlusion is not tolerated. The initial location of the object contour in the first frame (in which the object appears) has been interactively selected by a human operator. In figures 4 and 5, six frames are shown of a person in front of a textured background playing ping-pong. The size of the image is 260x135. The frames are clearly contaminated by shadows, shading and inter-reflections. Note again that each individual object-part (i.e. T-shirt, short, wall and table) is painted homogeneously with a distinct colour. Further, the wall contains texture. The results of the tracking system are shown in figure 4 tracking the T-shirt, and in figure 5 tracking the body of a person. The tracking system is based on $\nabla \mathcal{C}_{m_1m_2m_3}$. As one can see, all objects are well tracked ignoring radiometrical effects. From the observed results, the tracking technique successfully segment and track the objects.

V. Conclusion

We have formulated a colour-based deformable model. Computational methods have been presented to measure colour constant gradients. Further, a model has been given to estimate the amount of sensor noise through these color constant gradients. The obtained uncertainty is subsequently used as a weighting term in the deformation process.

From the theoretical and experimental results, we conclude that the proposed tracking system successfully find material contours discounting illumination. Furthermore, the method is robust against noisy, but homogeneous regions.

Acknowledgments

The author is grateful to Sennay Ghebreab for partially implementing the algorithm, and the anonymous reviewers for their valuable comments.

References

[1] Y. Bar-Shalom and T. Fortmann, Tracking and Data Association, Academic Press London, 1988.

Fig. 3. From left to right a. Colour image with the initial contour as specified by the user (the white contour). b. Segmentation result based on intensity gradient field ∇I. c. Segmentation result based on RGB gradient field $\nabla \mathcal{C}_{RGB}$. d. Segmentation result based on $\theta_1\theta_2$ gradient field $\nabla \mathcal{C}_{\theta_1\theta_2}$.

Fig. 4. *Frames from a video showing a person against a textured background playing ping-pong. From top-down and left-right frames are showing by tracking the T-shirt of the person.*

[2] A. Blake and M. Isard, Active Contours, Springer, 1998.

[3] C. M. Brown and D. Terzopoulos, Introduction, in Real-Time Computer Vision, C.M. Brown and D. Terzopoulos (eds), pp. 3-33, Cambridge University Press, 1994.

[4] C. Chesnaud, P. Refregier, V. Boulet, Statistical Region Snake-Based Segmentation Adapted to Different Physical Noise Models, PAMI(21), No. 11, November 1999, pp. 1145-1157.

[5] D. Comaniciu and P. Meer, Mean Shift Analysis and Applications, IEEE ICCV, Greece, pp. 1197-1203, 1999.

[6] J. Evans, Image-enhanced Multiple Model Tracking, Automatica, vol. 35, pp. 1769-1786, 1999.

[7] Finlayson, G.D., Drew, M.S., and Funt, B.V., Spectral Sharpening: Sensor Transformation for Improved Color Constancy, JOSA, 11, pp. 1553-1563, May, 1994.

[8] Funt, B. V. and Finlayson, G. D., Color Constant Color Indexing, IEEE PAMI, 17(5), pp. 522-529, 1995.

[9] Th. Gevers and Arnold W.M. Smeulders, Color Based Object Recognition, Pattern Recognition, 32, pp. 453-464, March, 1999.

[10] M. Kass, A. Witkin, D. Terzopoulos, Snakes: Active Contour Models, International Journal of Computer Vision, 1(4), pp. 321-331, 1988.

[11] K.F. Lai and R.T. Chin, Deformable Contours: Modeling and Extraction, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 17(2), 1995.

[12] A.J. Lipton, H. Fujiyoski, R.S. Patil, Moving Target Classification of and Tracking from Real-Time Video, IEEE Workshop on Applications and Computer Vision, Princeton, pp. 8-14, 1998.

[13] R. Malladi, J.A. Sethian, B.C. Vemuri, Shape Modeling with Front Propagation, IEEE Trans. on Pattern Analysis and Machine Intelligence, 1994.

[14] S. K. Nayar, and R. M. Bolle, Reflectance Based Object Recognition, International Journal of Computer Vision, Vol. 17, No. 3, pp. 219-240, 1996

[15] K.P. Ngoi, J.C. Jia, An active contour model for colour region extraction in natural scenes, IVC(17), No. 13, 1999, pp. 955-966.

[16] H.T. Nguyen, M. Worring and A. Dev. Detection of Moving Objects in Video using a Robust Motion Similarity Measure, IEEE Trans. on Image Processing, Vol. 9, No. 1, pp. 137-141, Jan. 2000.

[17] S.A. Shafer, Using Color to Separate Reflection Components, COLOR Res. Appl., 10(4), pp 210-218, 1985.

[18] G. Sapiro, D. L. Ringach, Anisotropic Diffusion of Multi-valued Images with Applications, to Color Filtering, IEEE PAMI, (5)11, 1582-1586, 1996.

[19] G. Sapiro, Color Snakes, CVIU(68), No. 2, November 1997, pp. 247-253.

[20] J. R. Taylor, *An Introduction to Error Analysis*, University Science Books, 1982.

[21] D. Terzopoulos and R. Szeliski, Tracking with Kalman Snakes, in Active Vision, Blake and Yuille (eds), pp. 3-20, MIT Press, 1992.

[22] S.C. Zhu and A. Yuille, Region Competition: Unifying Snakes,

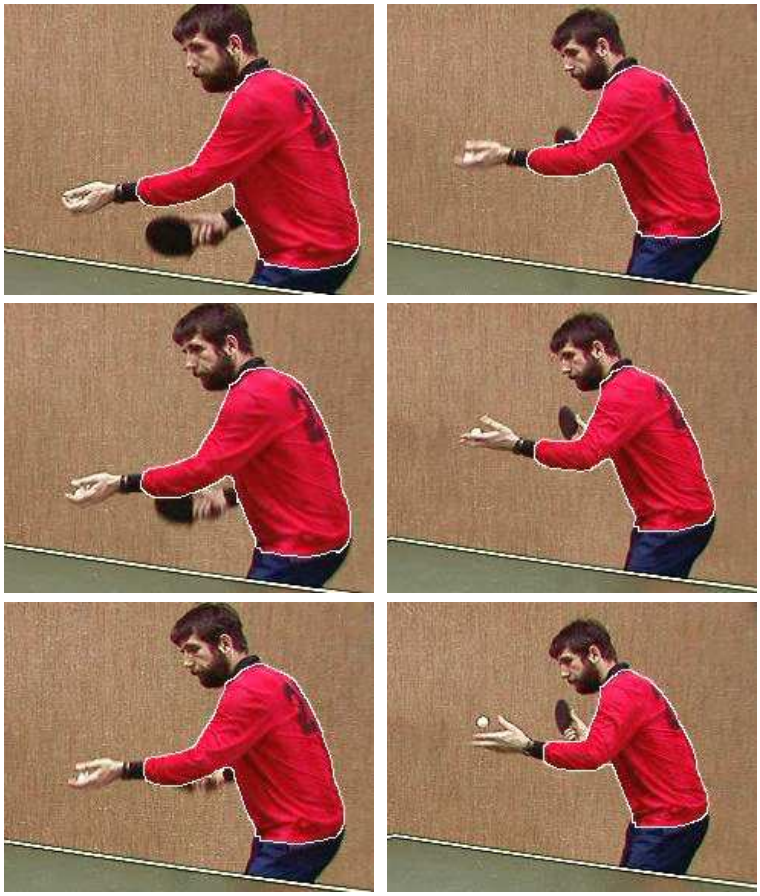

Fig. 5. *Frames from a video showing a man against a textured background playing ping-pong. From top-down and left-right frames are showing by tracking the body of the person.*

Region Growing, and Bayes/MDL for Multiband Image Segmentation, PAMI(18), No. 9, September 1996, pp. 884-900.

[23] A. Zisserman, A. Blake, R. Curwen,. Zisserman, A. Blake, R. Curwen, A Framework for Spatio-Temporal Control in the Tracking of Visual Contours, in Real-Time Computer Vision, C.M. Brown and D. Terzopoulos (eds), pp. 3-33, Cambridge University Press, 1994.

Theo Gevers is an associate professor of Computer Science at the University of Amsterdam, The Netherlands. His main research interests are in the fundamentals of image database system design, image retrieval by content, theoretical foundation of geometric and photometric invariants and color image processing.