IEEE International Conference on Multimedia and Expo 2003 (ICME'03)

SHADOW-RESISTANT TRACKING IN VIDEO

Hao Jiang and Mark S. Drew
Simon Fraser University, Vancouver, Canada
{hjiangb,mark}@cs.sfu.ca

ABSTRACT

In this paper, we present a new method for tracking objects with shadows. Traditional motion-based tracking schemes cannot usually distinguish the shadow from the object itself, and this results in a falsely captured object shape, posing a severe difficulty for a pattern recognition task. In this paper we present a color processing scheme to project the image into an illumination invariant space such that the shadow's effect is greatly attenuated. The optical flow in this projected image together with the original image is used as a reference for object tracking so that we can extract the real object shape in the tracking process. We present a modified snake model for general video object tracking. Two new external forces are introduced into the snake equation based on the predictive contour such that (1) the active contour is attracted to a shape similar to the one in the previous video frame, and (2) chordal string constraints across the shape are applied so that the snake is correctly maintained when only partial features are obtained in some frames. The proposed method can deal with the problem of an object's ceasing movement temporarily, and can also avoid the problem of the snake tracking into the object interior. Global affine motion compensation makes the method can be applicable in a general video environment. Experimental results show that the proposed method can track the real object even if there is strong shadow influence.

1. INTRODUCTION

Many computer vision tasks are made more difficult by the presence of shadows, confounding the recognition of objects and presenting a major challenge to almost every algorithm that depends on visual information. For example, the disambiguation of edges due to shadows and those due to material changes has a long history in computer vision research [1]. In fact, consideration of shadows as a cue for image understanding goes back to the beginnings of machine vision research [2]. Graphics and digital photography also must deal with shadow information in such tasks as color correction [3] and dynamic range compression [4].

Particularly in object tracking tasks we can expect shadows to present a confounding factor, since we need to distinguish the moving object from the shadows moving with it. Usual approaches involve *classification* of shadow pixels and material pixels, with perhaps some further classification of self-shadowing from cast shadows. Here we take an opposite point of view: instead of attempting to deal with shadows directly, we circumvent the issue as much as possible by trying to simply eliminate, or at least greatly attenuate shadows, thus obviating the problem.

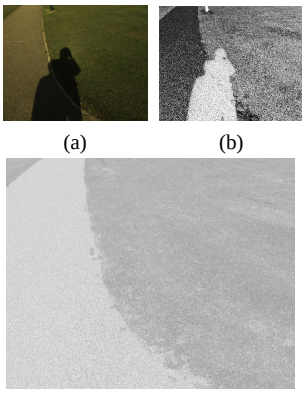
Some approaches do indeed try to eliminate shadows by detecting and correcting them. In the context of surveilling roads for the presence of pedestrians [5], under sunlight, the 'core lines' of a walking human are extracted, along with those of the shadows, based on the motion detection map. The internal and external parameters of the video camera are fixed in that method. Again in the context of surveillance, in [6] a scheme based on image subtraction is presented. The image captured by the first camera is first projected onto the road plane and then further projected to the image plane of the second camera. The road maps of two images will map perfectly while parts such as walking humans will not map well. Using the second image to subtract the projection of the first image will eliminate everything on the road plane including the moving shadows. In [7], pixels are classified on the basis of a statistical method. The features used include the illuminance and normalized chrominance vector. Color change under changing illumination is described by a von Kries rule — each color channel is approximately multiplied by a single overall multiplicative factor [8]. Pixels are classified into background, foreground, and shadow, based on maximum a posteriori classification, and spatial information is also applied to improve the dense region classification result. In [9], the shadow detection problem is studied based on a model similar to the Phong model. Heuristic methods are presented to classify the shadow and foreground object.

In [10], an attempt is indeed made to apply a color *invariant* approach to shadow classification, in the context of still image segmentation. The idea is to develop simple illumination invariant features so as to obtain an image which reflects surface materials only. Since cast shadows only change the illumination of backgrounds, the illumination invariant features will attenuate shadow effects. One of the approximately illumination invariant spaces devised by Gevers et al. [11] is first used to transform the color space. The transform favored is the space $c_1c_2c_3$, defined via

$$\{c_1, c_2, c_3\} = \arctan\left[\frac{\{R, G, B\}}{\{\max(G, B), \max(R, B), \max(R, G)\}}\right]$$

This color space has the virtue of being approximately invariant to shading and intensity changes, albeit only for matte surfaces under equi-energy white illumination.

Here we take another tack entirely: rather than redefining color in general, we concentrate on considering how the current camera handles differences in illuminant color. After all, what is a shadow? We can approximate it by the idea that areas not in shadow are directly illuminated, say by sunlight, but also receive indirect lighting, say from sky light. And areas in the umbra are likewise illuminated indirectly but lack a direct component — in essence, then, the color of the light combination is different. As a simple approximation we say that combinations of light are characterized by their Planckian temperature T; this is the idea underlying the concept of correlated color temperature [12]. If we characterize lighting by Planck's law, then the effect of lighting change amounts to temperature change. Since T occurs in an exponential in Planck's law, we can remove the effect of lighting entirely by taking logarithms of color ratios and projecting perpendicular to the direction of lighting change [13]. This direction is dependent on what camera we're using — e.g., a webcam [14] may produce a different direction than does a camcorder.



(a) (b)

(c)

Fig. 1. (a): Original image. (b): Ostensibly invariant c_1, c_2, c_3 image. (c): Grayscale shadow-invariant image. [Color images may be viewed at www.cs.sfu.ca/∼mark/ftp/Icme03/icme03.pdf]

What we give up, in this way, is the 3-dimensional nature of color: the resulting illumination-, and hence shadow-invariant (or at least resistant) image is a 1-D grayscale image. As an example, Fig. 1(a) shows a color image with an obvious shadow: the sun is behind the two people and their shadows fall across the path. Fig. 1(b) shows the same image, again in color, but in the ostensibly invariant color space $c_1c_2c_3$. The grayscale image Fig. 1(c), projected orthogonal to the T direction, is clearly superior.

In this paper, we specifically make use of the new type of illumination invariant image. If lighting is approximately Planckian, then as the illumination color changes, a log-log plot of 2-dimensional {log(R/G), log(B/G)} values for any single surface forms a straight line, provided camera sensors are fairly narrowband [13, 15, 14]. Thus lighting change reduces to a linear transformation along an almost straight line. Finding that line is a calibration task. We can use any target image, such as a Macbeth ColorChecker [16] to find the best direction of all such lines, and then apply that same direction to any new image. The projection orthogonal to the lighting-change direction greatly attenuates shadowing.

We further present an inertia enhanced snake model for tracking objects with shadows. We introduce two inertia terms. The first one adds an energy term based on the predictive contour; the second is a chordal constraint which tries to maintain the shape of the contour from frame to frame. This term attracts the active contour to converge to a shape similar to the one in the previous video frame. As well, instead of simply using the predictive contour to re-initialize the snake, we construct a new initial contour by a uniform expansion along the normal of the previous contour. This scheme prevents the contour from erroneously tracking features inside the true boundary of the object. At the same time, the new inertia energy terms make the snake ignore distracting elements. As well, if the object stops moving temporarily, the snake will evolve according to the inertia terms in the predictive contour and converge to a shape that corresponds to the motion prediction result and similar to the shape in the last frame. We adopt an affine motion model for global motion estimation and camera motion compensation with the result that our scheme can work in a general video environment.

2. SHADOW RESISTANT GRAYSCALE IMAGE

Suppose a surface is illuminated by Planckian lighting. In Wien's approximation of Planck's law, illumination $E(\lambda)$ is given by

$$E(\lambda) = Ic_1\lambda^{-5}e^{-\frac{c_2}{\lambda T}}$$

where I is intensity, T is temperature, and c_1 and c_2 are constants. Suppose narrowband camera sensors are approximately spike sensitivities $Q_k(\lambda) = q_k\delta(\lambda-\lambda_k)$, $k = 1..3$. In a Lambertian model, the *color* ρ_k for a point with normal vector $\mathbf{n}$ and reflectance $S(\lambda)$, illuminated from direction $\mathbf{a}$, is given by

$$\rho_k = \mathbf{a}\cdot\mathbf{n}\int E(\lambda)S(\lambda)Q_k(\lambda)d\lambda = c_1\mathbf{a}\cdot\mathbf{n}IS(\lambda_k)\lambda_k^{-5}e^{-\frac{c_2}{\lambda_k T}}q_k$$

Plotting the log-ratios $r = \log[\rho_1/\rho_2]$ and $b = \log[\rho_3/\rho_2]$, for a given reflectance surface we have a linear relationship

$$r - \log[\frac{q_1S(\lambda_1)\lambda_2^5}{q_2S(\lambda_2)\lambda_1^5}] = (b - \log[\frac{q_3S(\lambda_3)\lambda_2^5}{q_2S(\lambda_2)\lambda_3^5}])\frac{\lambda_1-\lambda_2}{\lambda_3-\lambda_2}\frac{\lambda_3}{\lambda_1}$$

But no matter what the surface, every surface's line has the same slope. Therefore, in the direction orthogonal we arrive at an illumination invariant grayscale image.

2.1. Camera Calibration

To calibrate our particular camera, we use a set of color patches, as in Fig. 2. Let P be the collection of log-log ratio pair sets $\{(r_i^k, b_i^k)|i \in I, k \in K\}$ where $r_i^k = \log(R_i^k/G_i^k)$ and $b_i^k = \log(B_i^k/G_i^k)$; (R_i^k, G_i^k, B_i^k) is the color of patch i under illumination k in the RGB color space. We first shift the log-log ratio vector such that the center of the cluster corresponding to one patch under different illuminations is located at the origin of the coordinate system. The best line in Fig. 3 is found by Least Squares or a robust method. The invariant image is calculated as the grayscale image formed by projection of any log-log pixel onto the orthogonal direction.

3. SNAKE MODEL WITH INERTIA TERMS

3.1. Snake Equation with Predictive Contour Inertia and Chordal String Constraints

In this section, we enhance the robustness of active contours in the tracking problem by a new snake equation (cf. [17]):

$\min_{X(s),Y(s)} \int_s \mathcal{I}(X(s)) + \mathcal{I}(Y(s))$
$+\frac{\gamma}{2}(E(X(s), C(s)) + E(Y(s), D(s))) + \rho/2G(X(s), Y(s))ds,$
with $\mathcal{I}(X) = \frac{\alpha}{2}|\nabla X(s)|^2 + \frac{\beta}{2}|\nabla^2 X(s)|^2 + P(X(s))$

where $X(s)$ is the active contour in the current frame, and $Y(s)$ is an auxiliary contour coupled with $X(s)$. It is $X(s)$ *shifted by half the arc length indices*: if $s = 0..1$ for $X(s)$, then $Y = Y(s')$, $s' = 0..1$, corresponding to n going from $N/2$ to N, and then from 0 to $N/2$. Then we try to keep chords (or more properly, diameters) $X(s)$-to-$Y(s')$, for $s, s' = 0..1$, the same as in the previous frame — this is an inertial shape constraint.

Curves $C(s)$ and $D(s)$ are the prediction contours from the previous frame. Energy terms $E(\cdot,\cdot)$ and $G(\cdot,\cdot)$ are two inertial terms. As in a traditional snake, the internal energy of the active contour is coded via a term $(\alpha/2)|\nabla(\cdot)|^2 + (\beta/2)|\nabla^2(\cdot)|^2$. Finally, $P(\cdot)$ is the external force based on the feature of interest, such as object motion. We use the natural choice

$$E(A(s), B(s)) = \|A(s) - B(s)\|^2$$

and for $G(\cdot,\cdot)$,

$$G(A(s), B(s)) = [\|A(s) - B(s)\| - d(s)]^2$$

where $d(s)$ is a distance function from $X(s)$ to $Y(s)$, for chords parameterized by $s = 0..1$. In a Euclidean norm, the resulting Euler Equation leads to an evolution equation in artificial parameter t as follows:

$$\begin{aligned}\frac{\partial X}{\partial t} &= \alpha X_{ss} - \beta X_{ssss} - \nabla P(X) + \gamma(C - X) \\ &\quad + \rho\frac{(Y(s)-X(s))}{\|X(s)-Y(s)\|}(\|X(s) - Y(s)\| - d(s)) = 0, \\ \frac{\partial Y}{\partial t} &= \alpha Y_{ss} - \beta Y_{ssss} - \nabla P(Y) + \gamma(D - Y) \\ &\quad + \rho\frac{(X(s)-Y(s))}{\|X(s)-Y(s)\|}(\|X(s) - Y(s)\| - d(s)) = 0\end{aligned}$$

As is usual, we actually replace the potential term $-\nabla P(X)$ in the above equations by a generalized force term $F_{ext}(X)$.

The initial state $X(s)$ is $X_0(s)$; the initial state of $Y(s)$ is coupled via $Y_0(s) \equiv Y_0[1/n] = X_0[\frac{1}{N}((n + \frac{N}{2}) \bmod N)]$ with integer n; $D[1/n] = C[\frac{1}{N}((n + \frac{N}{2}) \bmod N)]$; and $d(s) = \|X_{last\_frame}(s) - Y_{last\_frame}(s)\|$. The coupled equations try to maintain the shape of the contour from frame to frame by a *chordal string constraint*. Weight ρ controls the degree of shape cohesiveness. The other new inertia force term in this modified active contour for the tracking problem is based on the prediction contour. Weight γ controls the degree the prediction has on the contour tracking. Note that the solution for the tracking contour is $X(s)$: $Y(s)$ is an accessory contour used only in the solution process, and in fact has solution $X[\frac{1}{N}((n + \frac{N}{2}) \bmod N)]$.

3.2. Contour Prediction and Smoothing

We predict the future contour position and shape by the method of block-wise motion estimation. For every point (x, y) on the previous contour X_{last_frame}, a square block of width d is constructed centered on the pixel. The best matching block center in a search window of size w is selected as the predicted point:

$$\begin{aligned}&(\widehat{\Delta x}, \widehat{\Delta y}) = \arg\min_{\Delta x\in(-w/2,w/2),\Delta y\in(-w/2,w/2)} \\ &\int_{\xi=x-d/2}^{x+d/2}\int_{\eta=y-d/2}^{y+d/2}|u(\xi + \Delta x, \eta + \Delta y, t) - u(\xi, \eta, t - \Delta t)| \\ &d\xi d\eta\,, \qquad C = X_{last\_frame} + (\widehat{\Delta x}, \widehat{\Delta y})\end{aligned}$$

where u is the image sequence.

Motion estimation sometimes fails to estimate the correct future contour position. This will occur if some part of the previous contour does not fall at the boundary of an object, a situation very common for snake tracking of objects with concave boundaries. We propose a new approach for smoothing the prediction contour: we apprehend the smoothing process as a self-evolving curve without external force:

$$\frac{\partial C}{\partial t} = \alpha_0 C_{ss} - \beta_0 C_{ssss}$$

A stopping time has to be specified so that the curve will not distort too much while smoothing the singular points. One problem of the smoothing process is that the contour shrinks during the process of smoothing. Therefore, we cannot use the prediction contour as the initial contour for the next frame tracking. So we calculate the initial contour by a uniform expansion of the previous frame's tracking result:

$$X_{ini} = X_{last\_frame} - c\,\mathbf{n}$$

where $\mathbf{n}$ is the inwards normal of X_{last_frame} and c is a constant.

3.3. Dominant Motion Compensation; External Force

To remove the dominant camera motion, we adopt an affine model; the affine flow field can be represented as

$$\theta_1(x, y) = p_1 + p_2x + p_3y,\ \theta_2(x, y) = p_4 + p_5x + p_6y$$

(a) Illumination 1. (b) Illumination 2. (c) Illumination 3.

Fig. 2. The color chipboards used for camera calibration.

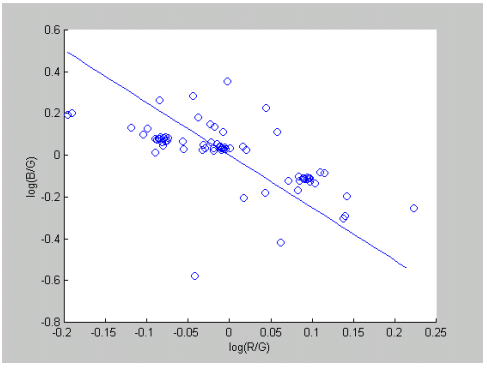


Fig. 3. Regression in log-log plot gives lighting change direction.

where $\vec{p} = \{p_1, p_2, p_3, p_4, p_5, p_6\}^T$ are the parameters to be estimated. In matrix notation,

$$\vec{\theta}(x, y) = A(x, y)\vec{p},\ \text{with } A(x, y) = \begin{pmatrix} 1 & x & y & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x & y \end{pmatrix}$$

The optical flow equation can be written as

$$\nabla u^T(A\vec{p}) + u_t = 0$$

with Least Squares solution

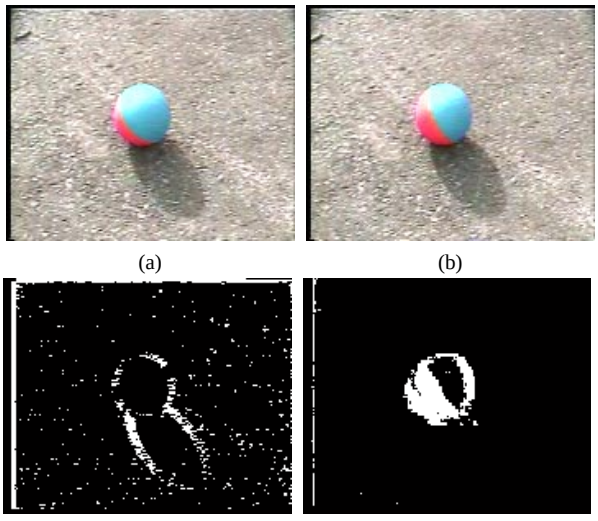
$$\vec{p} = (\sum A^T\nabla u\nabla u^T A)^{-1}(\sum -A^T\nabla u u_t)$$

In our tracking scheme, two motion detection maps are generated. The first one is based on the original video with global motion compensated. The other one is based on the shadow invariant global motion compensated images. We then use a simple thresholding scheme to detect the motion feature in both sequences. The intersection of both these motion detecting results produces an image segmentation map which is used to calculate the external force field of the snake based on the Gradient Vector Flow scheme [18].

4. EXPERIMENTAL RESULTS

Using a commercial camcorder (a Canon ES60), we first calibrated the camcorder to obtain the shadow invariant orientation. The Macbeth color chipboard with 24 patches is shown in Fig. 2. Images were formed under under three standard illuminants, comprising a daylight and two different indoor lightings. (Note, however, that standard lights are not at all necessary to calibrate a camera.) Fig. 3 shows that a scatter plot of the center-shifted log-log ratio data gives the illumination invariant orientation of the camcorder.

Fig. 4 shows the motion detection result for a two-color ball rolling on the ground based on the original image sequence and the illumination invariant sequence. The traditional motion detection scheme produces large errors on both the object's and shadow's boundary. Motion detection based on the shadow invariant image obtains much better results: the shadow's influence is nearly totally removed, and we also note that the background of the shadow



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

Fig. 4. Motion detection map. (a) Frame 1; (b) Frame 2; (c) Motion map by traditional scheme; (d) Motion map by shadow invariant scheme.

(a) Frame 1 (b) Frame 7 (c) Frame 13
(d) Frame 19 (e) Frame 25 (f) Frame 31
(g) Frame 37 (h) Frame 43 (i) Frame 50

Fig. 5. Tracking result with proposed method for the `baby` sequence.

invariant image motion detection result is much clearer. Recall that to increase robustness we use the intersection of the motion detection map of the original image sequence and that for the shadow invariant image sequence as the final motion detection map. Fig. 5 shows the result for tracking a baby's head, with an adjacent prominent shadow evident, under indoor lighting. The result shows that the contour is well tracked without being distracted by the shadow.

5. CONCLUSION

We utilize a new method for effective shadow *removal*, for tracking objects with shadows. Without the confounding factor of shadow edges, the real object can be tracked. This could be very useful for higher level vision processing such as gesture or behavior recognition. Future research directions include making the algorithm work in real time applications and using extra information from static scenes in webcam imagery.

References

[1] R. Gershon, A.D. Jepson, , and J.K. Tsotsos, "Ambient illumination and the determination of material changes," *J. Opt. Soc. Am. A*, vol. 3, pp. 1700–1707, 1986.

[2] D.L. Waltz, "Understanding line drawings of scenes with shadows," in *The Psychology of Computer Vision*, P.H. Winston, Ed., pp. 19–91. McGraw-Hill, 1975.

[3] P.M. Hubel, "The perception of color at dawn and dusk," in *7th Color Imaging Conference*, 1999, pp. 48–51.

[4] J.J. McCann, "Lessons learned from Mondrians applied to real images and color gamuts," in *7th Color Imaging Conference*, 1999, pp. 1–8.

[5] K. Onoguchi, "Shadow elimination method for moving object detection," in *14th Int. Conf. on Patt. Rec.*, 1998, pp. 583–587.

[6] Y. Sonoda and T. Ogata, "Separation of moving objects and their shadows, and application to tracking of loci in the monitoring images," in *4th Int. Conf. on Signal Proc.*, 1998, pp. Vol. 2:1261–1264.

[7] I. Mikic, P. Cosman, G. Kogut, and M. Trivedi, "Moving shadow and object detection in traffic scenes," in *ICPR2000*, 2000, pp. vol. 1:321–324.

[8] G.D. Finlayson, M.S. Drew, and B.V. Funt, "Color constancy: diagonal transforms suffice," in *Proc. Fourth Int. Conf. on Comp. Vision, Berlin*. IEEE, 1993, pp. 164–171.

[9] J. Stauder, R. Mech, and J. Ostermann, "Detection of moving cast shadows for object segmentation," *IEEE Trans. on Multimedia*, vol. 1, pp. 65–76, 1999.

[10] E. Salvador, A. Cavallaro, and T. Ebrahimi, "Shadow identification and classification using invariant color models," in *ICASSP2001*, 2001, pp. 7–1.

[11] T. Gevers and A.W.M. Smeulders, "Color-based object recognition," *Patt. Rec.*, vol. 32, pp. 453–464, 1999.

[12] G. Wyszecki and W.S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulas*, Wiley, New York, 2nd edition, 1982.

[13] G.D. Finlayson and S.D. Hordley, "Color constancy at a pixel," *J. Opt. Soc. Am. A*, vol. 18, no. 2, pp. 253–264, Feb. 2001, Also, UK Patent application no. 0000682.5. Under review, British Patent Office.

[14] G.D. Finlayson, S.D. Hordley, and M.S. Drew, "Removing shadows from images," in *ECCV 2002: European Conference on Computer Vision*, 2002, pp. 4:823–836, Lecture Notes in Computer Science Vol. 2353, http://www.cs.sfu.ca/∼mark/ftp/Eccv02/shadowless.pdf.

[15] G.D. Finlayson and M.S. Drew, "4-sensor camera calibration for image representation invariant to shading, shadows, lighting, and specularities," in *ICCV'01: International Conference on Computer Vision*. IEEE, 2001, pp. II: 473–480.

[16] C.S. McCamy, H. Marcus, and J.G. Davidson, "A color-rendition chart," *J. App. Photog. Eng.*, vol. 2, pp. 95–99, 1976.

[17] H. Jiang and M.S. Drew, "A predictive contour inertia snake model for general video tracking," in *ICIP'02*, 2002, http://www.cs.sfu.ca/ mark/ftp/Icip02a/icip02a.pdf.

[18] C. Xu and J.L. Prince, "Snake, shapes, and gradient vector flow," *IEEE Trans. on Im. Proc*, vol. 7, pp. 359–369, 1998.