MOVING SHADOW REPRESENTATION BASED ON A LEVEL CURVES DECOMPOSITION

Henri Nicolas
IRISA/INRIA, Campus de Beaulieu
35042 Rennes, France
hnicolas@irisa.fr

ABSTRACT

This paper proposes a new model to represent the moving shadow areas in video sequences. Classical models assume that the ambient light received in a shadow area is constant, but in practice it is not always the case. In order to define a more precise model, the variations of the ambient light are here taken into account thanks to a level curves decomposition of the shadow area. A B-Spline approximation of these level curves is introduced in order to get a smooth and compact representation. Experimental results show that this model can efficiently represents shadow areas. Results have been applied to the problem of shadow removal for video editing applications.

1. INTRODUCTION

The illumination conditions in a natural video sequence are very important in most of image sequence analysis methods [1]. This is due to the fact that the direction of illumination, the illumination itself and its variations, create the global illumination of the scene and local phenomena such as the moving shadows, for which the Human Visual System is very sensitive [2][3]. The shadow influences on our visual perception is studied in [4]. It is shown that shadows are a robust and relevant information source to visually appreciate the spatial and temporal video content.

The estimation of the illumination conditions in natural video sequences is therefore an important issue for many video sequence analysis applications. This mainly consists in: 1) the definition of an illumination model [5][6], 2) the detection of the shadow areas [7], and 3) the estimation of the light source characteristics (direction and position) [8][9]. A critical point for many applications consists in the definition of an accurate shadow model able to correctly representing the variations of intensity generated by the shadow effect.

This is for example the case for applications such as video editing and virtual reality [10][11][12][13]. For example, if a video object has to be removed from an original sequence, a satisfactory visual result can be obtained only if the corresponding shadow is also correctly removed [14]. At the opposite, if a moving object has to be added in a natural video sequence, its shadow has also to be created. These applications require being able to artificially lightened or shaded a given area. An other application is video compression, where a shadow model can be used to improve the prediction phase within an hybrid video coding scheme.

In this context, this paper proposes a new shadow representation model. This model is based on the use of a reference image which is assumed to represent the scene background without any moving image. A level curves representation of the shadow area is therefore obtained in order to represent the variations of illumination due to the shadow effect itself but also to the ambient light variations.

2. SHADOW MODEL

Assuming that the light source is far away from the scene, and that the surface on which a shadow is projected is Lambertian, i.e., that there is no specular effect, the intensity of a pixel *p* when it is shaded or illuminated can be respectively expressed as [5]:

$$I(p) = \begin{Bmatrix} k_a(p)I^a(p) + k_d(p)I_s < \vec{N}(p), \vec{L}(p) > \\ k_a(p)I^a(p) \end{Bmatrix}$$

where $I^a(p)$ is the intensity of the ambient light at pixel *p* and in the image *I*, I_s the intensity of the unique light source, and k_a, k_d the reemission and diffusion coefficients. $\vec{L}(p)$ represents the direction of the light source, and $\vec{N}$ the normal of the surface at point *p*. The *shadow ratio* $R_t(p)$ between a shaded point in the image I_t and the same illuminated point in an illuminated reference image I_{ref} can therefore be expressed as:

$$R_t(p) = \frac{I_t(p)}{I_{ref}(p)} = \frac{k_a(p)I^a(p)}{k_a(p)I^a_{ref}(p) + k_d(p)I_s < \vec{N}(p), \vec{L}(p) >}$$

Assuming that the light source is far from the scene, and assuming that the surface on which the shadow is projected is plane, $\vec{N}$ and $\vec{L}(p)$ can be considered as constant. Furthermore, the ratio k_a / k_d can reasonably be considered as constant mainly if we consider that the shadow area is projected on the same surface. Under these assumptions we have:

$$R_t(p) = \frac{I^a(p)}{I^a_{ref}(p) + K}$$

0-7803-8554-3/04/$20.00 ©2004 IEEE.

where K is a constant. The ambient light received at a physical point p of a shadow area comes theoretically from any direction. If the received quantity of ambient light is identical in any direction, the ambient light is constant everywhere in the shadow area, and we have:

$$I^a(p) = I^{ref}(p) \;\Rightarrow\; R_t(p) = R \quad (\forall p \in S) \quad (1)$$

where S is the shadow area. A shadow detection process based on this model can provide with good moving shadow segmentation results as it is done for example in [15]. Nevertheless, during the segmentation process, it is necessary to tolerate relatively large variations of the shadow ratio. This is due to noise problems, but also to the fact that the ambient light is not always constant throughout the shadow area, with the risk to erroneously detect non-shadow areas. This is mainly the case near the object and shadow boundary where the ambient light is not fully received (if the object is physically in contact with the background surface). On real moving shadow areas, the constant ambient light assumption is therefore not always valid. In order to obtain a more precise representation of the shadow area, it is therefore necessary to define a more precise shadow model. In practice, a shadow area can be decomposed in two sub-shadow areas as:

$$S = S_1 \bigcap S_2$$

where S_1 is the area where the ambient light is constant and similar to the ambient light received in the reference image (i.e. without the moving object), and S_2 the area where the ambient light is lower than in the reference image. Then we have:

$$I^a(p) = \alpha(p) I^a_{ref}(p) \;\Rightarrow\; R_t(p) = \alpha(p)\frac{I^a_{ref}(p)}{I^a_{ref}(p) + \mathrm{K}}$$

with $$\begin{aligned} &\alpha(p) = 1 \;\; (\forall p \in S_1) \\ &\alpha(p) \in [0,1[\;\; (\forall p \in S_2) \end{aligned}$$

The theoretical value of $\alpha(p)$ can be obtained only if the 3D structure of the scene and the general illumination conditions in the scene are known which is obviously not the case here. Assuming that the reduction of the ambient light in S_2 is due to the occlusion created by the object, and if S_2 is convex, the Shadow Ratio function is generally also convex with a global minimum noted $S_2^{\min}$, where $S_2^{\min}$ may be a pixel or a set of pixels (if S_2 is not convex, it can be split into a set of convex sub-shadow areas). Furthermore, if the object is physically in contact with the background surface, $S_2^{\min}$ is connected to the shadow/object boundary. In order to take into account in the shadow representation model the variations of the ambient light, the ratio image, within the shadow area S_2, is segmented using a set of *Shadow Level Curves (SLC)*. A level *R(L)* is therefore associated to the *Shadow Level Curve* L. *L* is defined as the contour of the largest region *A* such as:

$$A(L) = \{p \,/\, R(p) < R(L)\}$$

with

$$A(\mathrm{L}_{i+1}) \subset A(\mathrm{L}_i) \text{ and } R(\mathrm{L}_i) > R(\mathrm{L}_{i+1}) \quad (1)$$

where $A(L_i, L_{i+1})$ represents the area included between L_i with L_{i+1}. The Shadow Ratio R in $A(L_i, L_{i+1})$ is interpolated using one of the two following methods:

Discontinuous approach: This model is appropriate when there is an abrupt modification of the ambient light.

$$R(p) = R(L_{i+1}) \quad \forall p \in A(L_i, L_{i+1})$$

Linear interpolation: this method is useful to represent smooth modifications of the ambient light. Each value $p \in A(L_i, L_{i+1})$ is linearly interpolated as follows:

$$R(p) = \frac{d(p,L^i)}{d(p,L^i) + d(p,L^{i+1})} R(L_i) + \frac{d(p,L^{i+1})}{d(p,L^i) + d(p,L^{i+1})} R(L_{i+1})$$

where $d(p, L^i)$ is the distance between L^i and p.

3. ESTIMATION METHOD

The method used to identify the proposed shadow model is decomposed into the four following phases:

1 - Definition of the shadow contours. The contours of the moving shadow areas and the reference image are assumed to be available. A robust version of the method proposed in [15] has been used to obtain the shadow segmentations used for the experiments presented in this paper. It also provides a global shadow ratio R_g for each detected shadow area.

2 - Choice of the level curves. The shadow ratios associated to the SLC are defined in the range $[R_g, R_{\min}]$, where $R_{\min}$ is the lowest value of the shadow ratio in *S*. If a new level curve L_{new} has to be introduced between two existing ones L_i and L_{i+1}, the shadow ratio associated to L_{new} is $R(L_{new}) = \frac{R(L_i) + R(L_{i+1})}{2}$ (or $\frac{R_g + R_{\min}}{2}$ for the first created SLC). A new level curve is introduced between L_i and L_{i+1} if the average prediction quality (in term of PSNR) within $A(L_i, L_{i+1})$ is lower than a predefined threshold.

3 - Definition of the level curves contours. A given level curve $L_i(R_i)$ can theoretically be defined as the contours of the largest region $A(L_i)$ such as:

$$A(L_i) = \{p \in A(L_{i-1}) \,/\, R(p) < R(L_i)\}$$

Nevertheless, in practice, the shadow ratio image *R* is computed using the original and reference images and it may be noisy. As a consequence, the detected region $A(L_i)$ may contains holes or may be split into several sub-regions. The estimated region $\tilde{A}(L_i)$ is therefore defined as the largest detected sub-regions (and includes holes).

4 - B-Splines representation. The shadow model proposed in this paper has been developed for video editing and compression applications. For such applications, it is important to obtain smooth synthesized shadows and a compact representation of the SLC since they may have to be coded. For this purpose, a cubic B-spline representation of the SLC is used. The nodes of the splines are defined on the original curve using a criterion based on the global variations of orientation of the curve. Starting from an initial node n_1 at point p_1, the variations of orientation of the curve L_i is evaluated in order to detect the most significant orientation changes. For that purpose, a pixel-based measure of the orientation is first calculated as follows:

$$\Theta(p_n) = Angle(\overrightarrow{p_1 p_n}, \overrightarrow{p_n p_{n+1}})$$

where p_n is the *n*th pixel of L_i since the previous node, and $Angle(\vec{A}, \vec{B})$ the angle between oriented vectors $\vec{A}$ and $\vec{B}$. A new node is introduced at pixel p_n if the global variation of orientation since the previous node is significant. This is evaluated using the following criterion:

$$\frac{\sum_{k=1,n} \Theta(p_k)}{n} > T,$$

with the constraint to have the distance between two successive nodes higher than $d_{\min}$, (i.e. $n > d_{\min}$). The first node is the node which corresponds to the maximal value of the previous criterion when k is in the range $\pm d_{\min}$. For the experiments presented here, we take $d_{\min} = 20$. Coefficients k_n of the cubic B-spline representation have been estimated by minimizing a distance criterion defined between the approximated and original curves. The criterion is the following:

$$\{k_n\} = \arg\min_k Z(L, \hat{L}_k)$$

where Z is the total area included between the original level curve L and its B-spline approximation $\hat{L}_k$.

4. EXPERIMENTAL RESULTS

Experimental results have been performed on the two real video sequences *Desk* and *Road* which contain moving shadow areas.

Figure 1 illustrates the approximation obtained on a real SLC compared to the original one. It can also be seen that the estimated nodes corresponds to significant modifications of the curve orientation. Figure 2 shows the reference and shadow ratio images for *Road*. Figure 3 shows an image of *Road* in which the shadow area (and the corresponding video object which has been segmented separately) has been removed. It can be seen that the obtained edited image contains no visible default in the removed shadow area using the level curve decomposition. At the opposite, strong degradations occur if the level curve shadow decomposition is not used.

Figure 4 and 5 show the obtained level curve decompositions. In each case, the ambient light is lower near the object/shadow boundary. The level curves have logically been introduced near the shadow/object boundary. Nevertheless, for *Road*, a small area has been introduced in the left part of the shadow. This is due to the PSNR-based criterion used to validate the creation of a new curve. For these experiments, the quality threshold is fixed to 39dB. In both cases, the linear model has been selected (using the PSNR-based criterion) to interpolate the shadow ratio value for the intermediate SLC while the discontinuous one is selected only for the areas corresponding to the highest and lowest ratio values. This is logical since the highest values correspond to the areas where the ambient light is not modified by the object. Similarly, the lowest values correspond to the area where the occlusion effect due to the object is maximal. In Figure 5 it can be seen that the decomposition is relatively stable temporally. This confirms that the ratio shadow variations come mainly from the ambient light variations, and that these variations are directly linked to the moving object structure and its position relatively to the light source (constant here since the object is a sphere).

5. CONCLUSION

This paper proposed a new model to represent moving shadow areas when the ambient light cannot be considered as constant. The model is based on a level curves representation of the shadow ratio image. Furthermore, a B-spline approximation has been defined in order to obtain a compact and smooth representation of the level curves. Experiments show that good shadow representations have been obtained on real video sequences. The method has been applied for shadow removal in the context of video editing application. A main perspective is related to the application of the proposed shadow model to improve the prediction phase included in most video compression algorithms.

REFERENCES

[1] D. Waltz. *Understanding lines drawings of scenes with shadows*. The Psychology of Computer Vision, P. H. Winston, Ed., Computer Science Series, pp. 19-91. McGraw-Hill, 1975.

[2] H. Chen, P. Belhumeur, and D. Jacobs. In search of illumination invariants. In Proc. of Computer Vision and Pattern Recognition, Los Alamitos, pp.254-261, June 2000.

[3] P. Mamassian and R. Goutcher. *Prior knowledge on the illumination position*. Cognition, Vol. 81, pp. B1-B9, 2001.

[4] P. Mamassian, D. C. Knill, and D. Kersten. *The perception of cast shadows*. Trends in Cognitive Sciences, Vol. 2, No. 8, pp. 288-295, 1998.

[5] B. Phong. *Illumination for computer generated pictures*. Comm. ACM Vol. 18, N°6, pp. 311-317, June 1975,

[6] J. Stauder, R. Mech, and J Osterman. *Detection of Moving Cast Shadows for Object Segmentation*. IEEE Trans. on Multimedia, Vol. 1, No. 1, pp. 65-76, March 1999.

[7] C. Jiang and M.O. Ward. *Shadow segmentation and classification in a constrained environment*. In Proc. of Computer Vision, Graphics, and Image Processing. Image Understanding, Vol. 59, No. 2, pp. 213-225, March 1994.

[8] RA. Pentland. *Finding an illumination direction*. Journal of the Optical Society of America, Vol. 72, pp. 448-455, 1982.

[9] W. Chojnacki, M.J. Brooks, and D. Gibbins. *Revisiting pentland's estimator of light source direction*. Journal of the Optical Society of America A, Vol. 11, No. 1, pp. 118-124, January 1994.

[10] P. Debevec. *Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography*. Siggraph, Annual Conf. Series, Orlando, pp.189-198, 1998.

[11] I. Sato, Y. Sato, and K. Ikeuchi. *Acquiring a radiance distribution to superimpose virtual objects onto a real scene*. IEEE Transactions on Visualization and Computer Graphics, Hans Hagen, Vol. 5, No. 1, pp. 1-12, IEEE Computer Society, 1999.

[12] H. Nicolas. *New Methods for Dynamic Mosaicking*. IEEE Transactions on Image Processing, Vol. 10, No. 8 pp. 1239-1251, August 2001.

[13] R. Szeliski. *Videomosaics for virtual environments*. IEEE Computer Graphics and Applications, Vol. 2, No. 16, pp. 22-30, march 1996.

[14] R. Cucchiara, C. Grana, M. Piccardi, A. Prati and S. Sirotti. *Improving Shadow Suppression in Moving Object Detection with HSV Color Information*. Proc. IEEE Cong. Intelligence Transportation System, pp.334-339, August 2001.

[15] J.M. Pinel and H. Nicolas. *Cast shadows Detection on Lambertian Surfaces in Video Sequences*. Proc. of VCIP, pp. 378-384, Vol. 5150, Lugano, Switzerland, July 2003.

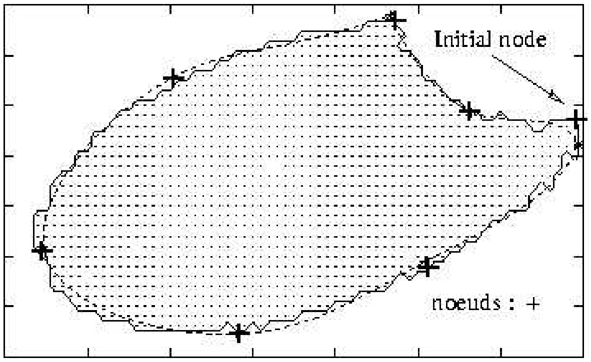

Figure 1: Illustration of the B-Spline approximation represented by the doted line on a real level curve detected on *Desk*.

Figure 2: Sequence *Road*. Left: reference image. Right: Shadow ratio image.

Figure 3: Sequence *Road*. Top: Original images of Desk. Middle : zoom on the shadow area. Bottom left: Edited image obtained after video object and shadow removal and using the SLC decomposition. Bottom right: zoom on the image obtained after shadow removal without using the SLC representation. Zoomed images have been lightened for visibility reasons.

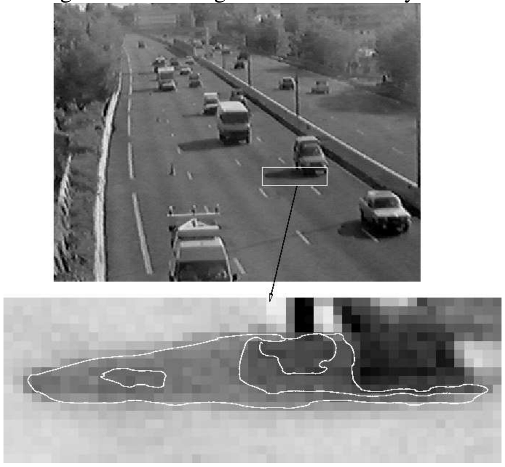

Figure 4: Sequence *Desk*. Zoom on the decomposed shadow area. The shadow ratio values associated to the SLC are: 0.58 (main area); 0.52 (for the small area on the left of the shadow); 0.47 ; 0.35 ; 0.28 (for the small area on the right).

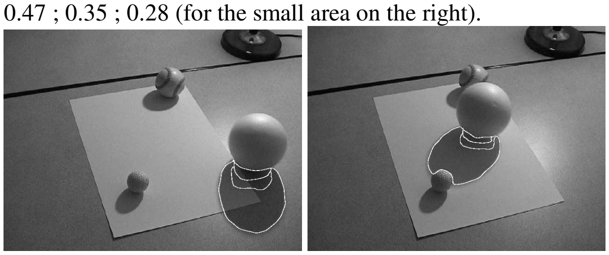

Figure 5: Sequence *Desk*. Shadow segmentation and two level curves (same shadow ratio for the two images). The shadow ratio values associated to the SLC are: 0.42 ; 0.35 ; 0.31.