This paper presents a new scalable block-based video compression scheme for video-surveillance applications. It is based on the detection and the modeling of cast shadows. Each block is classified according to its content, i.e. background, foreground, cast shadow and boundary. It is assumed that a reference image representing the fixed background of the scene is available which is generally realistic in the context of video-surveillance. The bitstream is decomposed into three layers containing respectively the foreground (basic layer), shadow and background update (optional enhancement layers) information. More bitrate can be dedicated to the coding of the foreground objects which is most relevant information. The shadows are coded using a precise model of the ambient light in the shadow area based on a level lines representation. The detected cast shadows are removed from the original images using the reference. Experimental results show that a compression gain can be obtained.

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]. The shadow influences on our visual perception is studied in [3]. 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 [4][5], 2) the detection of the shadow areas [6][7], and 3) the estimation of the light source characteristics (direction and position) [8]. 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 [9][10]. 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. In classical video compression schemes, moving shadows are generally not explicitly considered. The temporal activity generated by these shadow areas are therefore taken into account by the motion compensation phase, or by the prediction error coding. This is not satisfactory since moving shadows cannot be well representing by a motion model (except in specific situations, for example when a shadow is moving on a homogeneous surface). For applications related to video-surveillance, the moving shadows are generally not regions of interest. As a consequence, they can be removed or roughly coded without significant degradation of the quality of the decoded video. This paper proposes a new scalable video compression scheme which allows coding separately the moving cast shadows.

2. CODING SCHEME

The proposed video coding scheme is based on a scalable approach in which the images are decomposed into three main kinds of regions: the background, the moving cast shadows and the foreground objects. For video-surveillance applications, such decomposition is relevant since only the foreground objects are generally really useful. Higher visual quality can therefore be obtained on the foreground objects if less bitrate are dedicated to code the temporal variations in the background area generated by cast shadows. This is done here using a block-based representation, where each block is classified according to the following categories:

1. Foreground blocks
2. Background blocks
3. Shadow blocks
4. Shadow/foreground boundary blocks

This classification is obtained using a reference image (or a mosaic image if the camera is moving) which represents the background scene without any moving object. Such an image is
often available in the context of video-surveillance applications. The basic layer of our scalable bitstream contains only the information related to the foreground blocks. The detected background and shadow blocks are replaced by the corresponding blocks of the reference image. The first enhancement layer contains the shadow model parameters which allow to synthesis the shadow at the decoder side. The second enhancement layer contains the background update information. These enhancement layers are optional. To reach the best coding efficiency, we have chosen H264/AVC (MPEG4 Visual Part 10) to encode foreground and background blocks [10]. For the shadow/foreground boundary blocks, the shadow area is previously replaced by the corresponding texture of the reference image. The shadow ratio parameters are coding using a Huffman entropic coding. The level lines are coding using a Freeman coding technique. Figure 1 shows the block diagram of the proposed scheme.

3. BLOCK CLASSIFICATION

Each block of the original images is classified according to the four categories mentioned above. This is done using the following algorithm:

1- Detection of the background blocks. This detection is done using the difference between the original and reference ones. This difference image is threshold in order to eliminate isolated error points. Then the PSNR is calculated for each 8x8 block. Blocks for which the PSNR is higher than a threshold T are considered as background blocks.

2- Detection of the shadow blocks. For the remaining blocks, their compatibility with the shadow model is tested according to the following criterion:

\[
\frac{1}{\text{Card}(B)} \sum_{p \in B} (I_t(p) - R_t L_{\text{ref}}(p))^2 < T
\]

where B is the current block, \( L_{\text{ref}} \) is the reference image, \( R_t \in [0,1-\varepsilon] \) (\( \varepsilon \) is the shadow visibility threshold, set here to 0.9) is the global shadow ratio at time t (see the description of the shadow model in Section 4). For the first image, \( R_t \) is set as the value which provides the largest shadow area (for more details, see [9]). For a current image \( I_t \), the global shadow ratio is calculated as:

\[
R_t = \frac{1}{\text{Card}(S_{t-1})} \sum_{p \in S_{t-1}} I_{t-1}(p) L_{\text{ref}}(p)
\]

where \( S_{t-1} \) is the shadow area at time t-1. The remaining blocks are considered as foreground blocks.

3- Detection of the shadow/foreground boundary blocks. The blocks classified as foreground (shadow resp.) which have at least one neighborhood block classified as shadow (foreground resp.) may partially contains a shadow area. For these blocks, the shadow model is therefore tested on the four 4x4 sub-blocks using the two previous criteria to determine their type.

4. SHADOW MODEL

Assuming that the light source is far away from the scene, and that the surface on which a shadow is projected is plane and Lambertian, i.e., that there is no specular effect, the shadow ratio between a shaded point in the image \( I_t \) and the same illuminated point in \( I_{\text{ref}} \) can therefore be expressed as:

\[
R_t(p) = \frac{I_t^a(p)}{I_{\text{ref}}^a(p)} + K
\]

where K is a constant and \( I_t^a(p) \) is the intensity of the ambient light at pixel p and in the image I. The ambient light received at a physical point 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:

\[
I_t^a(p) = I_{\text{ref}}^a(p) \Rightarrow R_t(p) = R (\forall p \in S)
\]

Nevertheless, on real moving shadow areas, the constant ambient light assumption is not always valid. This is typically the case near the object/shadow boundary. In order to obtain a more precise representation of the shadow area, it is therefore necessary to define a more precise shadow model. 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:

\[
I_t^a(p) = \alpha(p) I_{\text{ref}}^a(p) \Rightarrow R_t(p) = \alpha(p) \frac{I_{\text{ref}}^a(p)}{I_{\text{ref}}^a(p) + K}
\]

with \( \alpha(p) = 1 (\forall p \in S_1) \)

\( \alpha(p) \in [0,1] (\forall p \in S_2) \)

The theoretical value of \( \alpha(p) \) can be obtained only if the 3-D structure of the scene and the general illumination conditions in the scene are known. If the object is physically in contact with the background surface, the minimal value of the ambient light is generally 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, is segmented using a set of Shadow Level Curves. A shadow ratio level \( R(L) \) is therefore associated to each Shadow Level Curve L, where L is defined as the contour of the largest region \( A \) such as:

\[
A(L) = \{ p \in S / R(p) < R(L) \}
\]
with $A(L_{i+1}) \subset A(L_i)$ and $R(L_i) > R(L_{i+1})$

$A(L_i, L_{i+1})$ represents the area included between $L_i$ and $L_{i+1}$. 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, then:

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

**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:

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

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

These level lines are computed in order to allow an optimal shadow synthesis. In practice, a new level line $L$ is included inside two existing level lines $L_i$ and $L_{i+1}$ using the following minimization criterion:

$$L = \arg \min_{L \in \mathcal{L}(L_i, L_{i+1})} \left\{ \sum_{p \in A(L_i, L_{i+1})} \left[ I_i(p) - R_i(p, L) f_{opt}(p) \right]^2 \right\}$$

The shadow areas are usually relatively small; it is therefore not computationally expensive to perform an exhaustive search to define the optimal decomposition. The number of created level lines depends on the available bitrate. It can also be determined by defining a minimum increase of the PSNR in the considered region to be valid. If it is not the case, the level line creation process stops.

Furthermore, the previous expression is tested for each interpolation technique (discontinuous or linear) of the shadow ratio for each shadow block. This information has therefore to be specified at the decoder for each shadow block.

### 5. EXPERIMENTAL RESULTS

Experimental results have been performed on video-surveillance sequences. Figure 2 shows an example of original images and the shadow ratio image. Figure 3 shows a level line decomposition obtained for a shadow area (sequence Road). Figures 4 shows the reference image used for Hall and the decoded images obtained with the basic layer only and with the first enhancement layer. A gain of around 15% can be obtained on the bitrate. Similarly, Figure 5 shows the decoded images before and after the shadow synthesis phase. The shadow information represents around 10-25% (depending on the test sequence) of the bitrate required to code the foreground area (and for a similar PSNR). For video-surveillance applications shadows are usually not useful. It can therefore be sufficient to code the basic layer containing only the foreground information. Enhancement layers containing the shadow information and the background update are therefore sent to the decoder only if the quality of the result is not satisfactory, for example if the foreground objects have been partially and erroneously classified as shadow areas by the block classification process. Figure 6 shows (for Road) the rate-distortion curves obtained by coding:

- The original sequence without the proposed shadow processing,
- The basic layer
- The basic and enhancement layers.

The performances are logically better for the sequence without shadows (basic layer) since the temporal variations have been significantly reduced. At low bitrate, the use of the proposed shadow modeling method allows a better compression. At higher bitrate, the variations of the ambient light within the shadow areas have to be more precisely taken into account. Thus, more level lines are introduced, but with a smaller gain in term of PSNR. As a consequence, the coding cost increases and the performance decreases.

### 6. CONCLUSION

This paper proposed a new method for video compression in the context of video-surveillance applications. The proposed method allows a scalable representation of the bitstream based on the separation of foreground, cast shadows and background update information. The proposed method allows reducing, at low bitrate, the coding cost by elimination of the moving shadows which are generally not useful for video-surveillance applications.

### REFERENCES


