Spatial Scene Level Shape Error Concealment for Segmented Video

Luis Ducla Soares[1], Fernando Pereira[2]

[1]Instituto Superior de Ciências do Trabalho e da Empresa – Instituto de Telecomunicações, Lisboa, Portugal
[2]Instituto Superior Técnico – Instituto de Telecomunicações, Lisboa, Portugal
[1]lds@lx.it.pt, [2]fp@lx.it.pt

Abstract. In this paper, an original spatial scene level shape error concealment technique for segmented object-based video scenes, to be used in error-prone environments such as mobile networks, is proposed. This technique is different from existing shape concealment techniques because it considers, not only the video objects that have to be concealed, but also the context in which these objects are inserted. The obtained results suggest that the use of this technique, instead of independently concealing video objects, significantly improves the subjective visual impact of scenes on the end-user, in addition to making the concealment operation itself easier since the spatially adjacent shape data from surrounding video objects can also be used.

Index Terms—Spatial error concealment, binary shape data, segmented video scenes

1. INTRODUCTION AND PROBLEM STATEMENT

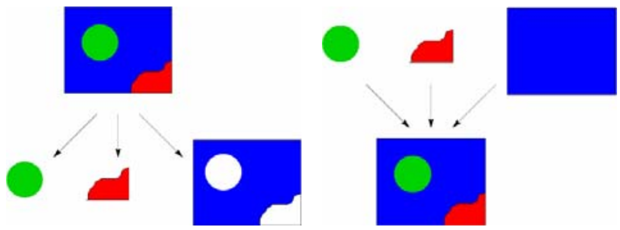
The emergence of the MPEG-4 object-based audiovisual coding standard [1] opened up the way for new video services, where scenes consist of a composition of objects. In order to make these object-based services available in error-prone environments, such as mobile networks or the Internet, with an acceptable quality, appropriate error concealment techniques dealing with shape and texture data are necessary.

Several such techniques have recently been proposed in the literature, e.g., [2][3][4][5][6][7] for shape and [8] for texture. These techniques, however, have a serious limitation in common, which is the fact that each video object is independently considered, without ever taking into account the scene context in which the objects are inserted. After all, just because a concealed video object has a pleasing subjective impact on the user, when it is considered on its own, does not necessarily mean that the subjective impact of the whole scene will be acceptable, when it is considered as a whole. An example of this situation is given in Figure 1, where a hole has appeared as a result of composing the scene from two independently concealed video objects.

When concealing a complete video scene, the way the scene was created has to be considered since this will greatly influence the way concealment has to be done. As shown in Figure 2, there are basically two different approaches that can be adopted to create object-based video scenes. The video objects in a scene can be defined either by segmentation of an existing video sequence (segmented scene) or by composition of pre-existing video objects whose shapes do not necessarily have to fit perfectly together (composed scene). These two cases will imply different problems and solutions when it comes to error concealment. Additionally, it is also possible to define a scene using a combination of both approaches above.

(a) (b)

Figure 1. Illustration of a typical scene concealment problem: (a) Original video scene; (b) Composition of two independently error concealed video objects.



(a) (b)

Figure 2. Two different scene types: (a) Segmented scene; (b) Composed scene.

In this paper, segmented video scenes (or the segmented parts of hybrid scenes) are considered since concealment of composed scenes can typically be limited to object concealment. Additionally, due to the subjective importance of shape data, as well as its influence in the decoding and display of texture data, concealment of shape errors will be investigated first; only binary shape data will be considered. Shape error concealment techniques can be either spatial or temporal, depending on the data that is used for the concealment, each one with its own advantages and drawbacks. In this paper, the spatial concealment approach will be considered and, therefore, only information from the same time instant will be used in the shape concealment process.

2. PROPOSAL FOR A SPATIAL SCENE LEVEL SHAPE ERROR CONCEALMENT TECHNIQUE FOR SEGMENTED VIDEO SCENES

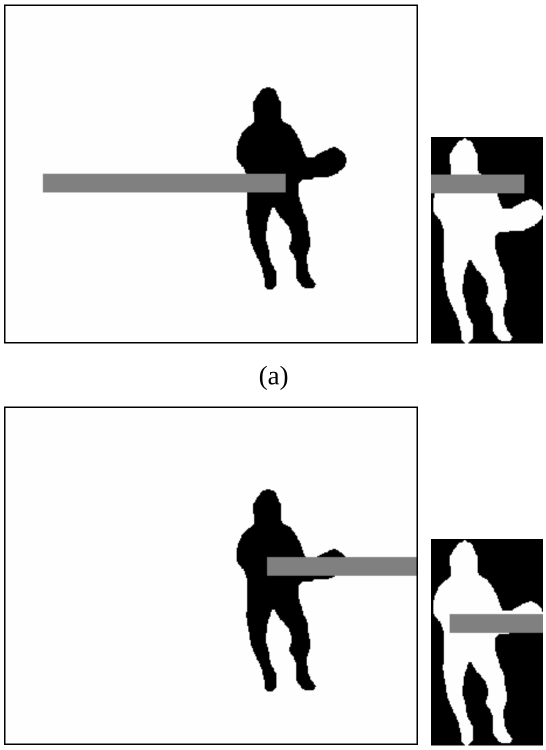
In segmented scenes, the various video objects have to fit together like the pieces in a jigsaw puzzle and, therefore, if there is any distortion in the shape data of these objects, holes or object overlapping will appear, leading to a subjective negative impact. However, the fact that the existing video objects have to perfectly fit in together can also be exploited when it comes to the concealment of binary shape data errors. In many cases, it will be possible to conceal at least some parts of the corrupted shape in a given corrupted video object by considering uncorrupted complementary shape data from surrounding objects. For those parts of the corrupted shape for which complementary data is not available because it is corrupted, concealment will be much harder.

Thus, for the concealment of corrupted shape data in segmented video scenes, the following two distinct cases, illustrated in Figure 3, have to be considered:

- **Correctly decoded complementary shape data** – In this case, the shape data from the surrounding video objects can be used to conceal the shape of the corrupted video object at hand since it is uncorrupted.
- **Corrupted complementary shape data** – In this case, the shape data from the surrounding video objects cannot be used to conceal the shape of the corrupted video object since it is also corrupted.

Although these two cases correspond to two different concealment situations, they may coexist in the same corrupted video object. Thus, it is possible that for a given corrupted video object some parts of the corrupted shape have correctly decoded complementary data in the surrounding objects, while for other parts the complementary data is not available because it is corrupted.

By considering these two cases, which have to be dealt with using different concealment approaches, a complete scene level shape error concealment solution can be proposed. The proposed solution corresponds to the following algorithm, which is divided in two consecutive steps:



(a)

(b)

Figure 3. Illustration of the two possible concealment situations for the Stefan video sequence objects (Background and Player): (a) Correctly decoded complementary shape data; (b) Corrupted complementary shape data.

1. **Concealment of corrupted shape with available complementary data** – In this first step, all the parts of the corrupted shape for which correctly decoded complementary shape data is available are concealed. To do this for a given corrupted video object, a complementary alpha plane has to be created which is basically the union of all the video objects in the scene except for the one currently being concealed. Afterwards, each corrupted shapel of the video object being concealed is set to the opposite of the corresponding shapel in the complementary alpha plane. Since the complementary alpha plane can also have corrupted parts, this is only done if the needed data is uncorrupted. This whole procedure is repeated for all video objects with corrupted shape. It should be noticed that, for those parts of the corrupted shape for which complementary data is available, this type of concealment recovers the corrupted shape without any distortion with respect to the original, which does not happen in the following step.

2. **Concealment of corrupted shape for which no complementary shape data is available** – In this step, the remaining corrupted shape data, which could not be concealed in the previous step, will be concealed. This step is divided in the following two consecutive phases:

   a) **Individual concealment of video objects** – Since the remaining corrupted shape of the various video objects in the scene has no complementary data available that can be used for concealment, it will be independently concealed. This can be done by using any of the available techniques in the literature. Here, however, since a spatial concealment technique is desired, individual concealment of video objects will be done by contour interpolation with Bézier curves, as proposed in [3].

   b) **Refinement of the individual concealment results** – As a result of the previous phase, holes or object overlapping may appear in the scene. The regions that correspond to holes or object overlapping are considered undefined, in the sense that they do not belong to any object yet. In this last phase, these undefined regions will be divided among the video objects around it. To do this, a morphological filter based on the dilation operation [9] is cyclically applied to the N objects in the scene, A_1,A_2, …, A_N, until all undefined regions disappear. The morphological operation to be applied to object A_j is the following:

$$A_j \oplus B - \left[\left[A_j \oplus B \right] \cap \bigcup_{i=1, i \neq j}^{N} A_i \right], \qquad (1)$$

   where the 3×3 structuring element B that is used for the dilation operation ⊕ is shown in Figure 4. This basically corresponds to dilating object A_j into the undefined region(s) adjacent to it. By cyclically applying this filter, the undefined regions will be progressively absorbed by the objects around them until they finally disappear. An example of this refinement is shown in Figure 5 for an undefined region which is surrounded by three objects.

0	1	0
1	1	1
0	1	0

Figure 4. Structuring element used for the dilation operation in the refinement of individual concealment results.

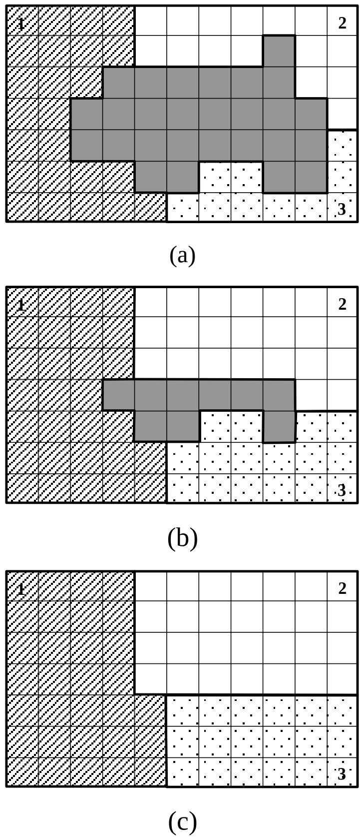


Figure 5. Elimination of undefined region by morphological filtering: (a) Initial undefined region; (b) Undefined region is shrinking; (c) Undefined region has been eliminated.

3. PERFORMANCE EVALUATION

In order to illustrate the performance of the proposed shape concealment process, Figure 6 should be considered. In this example, the four video objects of the News video scene (in Figure 6 (a)) have been corrupted, as shown in Figure 6 (b). In the remainder of Figure 6, the various steps of the concealment process are shown, leading to the final concealed video objects in Figure 6 (f). To compare these video objects with the original ones in Figure 6 (a), the *Dn* shape quality metric used by MPEG may be used. This metric is defined as

$$Dn = \frac{diff\_shapels}{original\_shapels}, \qquad (2)$$

where *diff_shapels* is the number of different shapels between concealed and original alpha planes, and *original_shapels* is the number of opaque shapels in the original alpha plane. This metric can also be expressed as a percentage,

$$Dn[\%] = 100 \times Dn. \qquad (3)$$

The obtained values are 0.0%, 1.0%, 0.7% and 0.9%, respectively for the Background, Dancers, Speakers and Logo video objects shown in Figure 6 (f). As can be seen, although the shapes of these video objects have been severely corrupted, the results are quite impressive, especially when compared to what is achieved by independent concealment alone. The main reason for such a great improvement over what is typically achieved with independent spatial concealment is the usage of the complementary data from surrounding objects during the concealment process, which does not happen when only independent concealment is performed.

To further illustrate the importance of considering the whole scene when performing concealment, instead of just considering the various objects independently, the following results should be considered. For these results, the News and Stefan video sequences, whose objects have been encoded according to the MPEG-4 Core Visual Object Type, have been considered. While the News sequence (CIF, 10 fps) has been encoded at a total rate of 192 kbps, the Stefan sequence (CIF, 15 fps) has been encoded at a total rate of 256 kbps. Additionally, since in MPEG-4 each video object plane (VOP) can be divided in several independently decodable units called video packets (VP), which can influence the concealment performance, each video object has also been encoded with 2, 4 and 8 VPs per VOP.

In Table 1 and Table 2, the average number of frames where corrupted shape had to be concealed is shown, as well as the average number of frames where the corrupted shape could not be concealed only with the complementary data from surrounding objects, respectively for the News and Stefan video sequences. The results, which correspond to averages over 50 different error patterns, are shown for several VP loss rates and for the three packetization schemes mentioned above. As can be seen, for lower VP loss rates, a large percentage of frames can be concealed by only considering the complementary shape data from surrounding objects; in these frames concealment is perfect. However, for larger VP loss rates, this is no longer the case and the concealment of corrupted shapes for which no complementary shape data is available becomes increasingly important.

In Table 3 and Table 4, for the same conditions as before, the average percentage of shapels that need to be concealed in each frame and the average percentage of shapels that need to be concealed after the parts of the corrupted shape for which complementary data is available have been concealed are presented. From these results, it becomes clear that the availability of the complementary shape data in surrounding objects can significantly ease the whole concealment process. This is especially true for lower VP loss rates, where it is more likely that the complementary shape data will be uncorrupted. For higher VP loss rates, however, the probability of having corrupted complementary shape data increases and, therefore, the use of the data from surrounding objects is not as advantageous, but it is still very useful.

Table 1. Average number of frames where corrupted shape could not be concealed only with the complementary data from surrounding objects for the News sequence (100 frames).

VP Loss Rate	Frames with corrupted shape (2VPs/4VPs/8VPs)			Frames where complementary data concealment was not enough (2VPs/4VPs/8VPs)		
1%	9.0	15.4	25.4	0.1	0.3	0.5
5%	37.3	54.2	74.3	3.5	7.0	11.3
10%	61.3	78.6	92.7	12.2	20.5	35.4
20%	85.6	96.0	99.6	35.8	53.3	74.5

Table 2. Average number of frames where corrupted shape could not be concealed only with the complementary data from surrounding objects for the Stefan sequence (150 frames).

VP Loss Rate	Frames with corrupted shape (2VPs/4VPs/8VPs)			Frames where complementary data concealment was not enough (2VPs/4VPs/8VPs)		
1%	7.0	12.8	21.8	0.1	0.1	0.4
5%	32.6	51.1	76.1	1.3	2.4	4.2
10%	57.9	84.3	112.7	5.1	7.8	13.7
20%	93.8	120.0	140.0	16.9	26.1	40.5

Table 3. Average percentage of undefined shapels before and after the complementary shape data from surrounding objects has been used for the News sequence (100 frames).

VP Loss Rate	Percentage of undefined shapels **before** any concealment has been applied (2VPs/4VPs/8VPs)			Percentage of undefined shapels **after** the complementary shape data has been used (2VPs/4VPs/8VPs)		
1%	1.06	1.17	1.08	0.02	0.01	0.01
5%	5.23	5.19	5.16	0.31	0.27	0.27
10%	10.35	10.16	10.26	1.10	0.96	1.03
20%	20.61	20.24	20.17	3.99	3.92	3.90

Table 4. Average percentage of undefined shapels before and after the complementary shape data from surrounding objects has been used for the Stefan sequence (150 frames).

VP Loss Rate	Percentage of undefined shapels **before** any concealment has been applied (2VPs/4VPs/8VPs)			Percentage of undefined shapels **after** the complementary shape data has been used (2VPs/4VPs/8VPs)		
1%	0.89	1.10	1.19	0.00	0.01	0.01
5%	5.15	5.09	5.27	0.08	0.08	0.08
10%	10.24	10.03	10.23	0.35	0.29	0.31
20%	19.93	20.10	20.12	1.25	1.18	1.15

4. FINAL REMARKS

In this paper, a spatial shape concealment technique for segmented object-based video scenes, such as those based on the MPEG-4 standard, is proposed. Results have been presented showing the ability of this technique to recover lost binary shape data in complete video scenes with rather small distortion, when compared to independent concealment of the various objects.

Finally, it is important to emphasize the relevance of shape concealment techniques, not only to achieve an acceptable shape quality, but also because the decoded texture quality obtained is highly dependent on the quality of the shape data (i.e., the texture data can only be correctly decoded if the shape data is correct). Therefore, for object-based video applications (with more than one object) to be actually deployed in error-prone environments, robust shape error concealment techniques for complete video scenes will have to be available.

5. REFERENCES

[1] ISO/IEC 14496-2, “Information Technology – Coding of Audio-Visual Objects, Part 2: Visual,” December 1999.

[2] S. Shirani, B. Erol, F. Kossentini, “A Concealment Method for Shape Information in MPEG-4 Coded Video Sequences,” *IEEE Transactions on Multimedia*, Vol. 2, No. 3, pp. 185-190, September 2000.

[3] L. D. Soares, F. Pereira, “Spatial Shape Error Concealment for Object-based Image and Video Coding,” *IEEE Transactions on Image Processing*, Vol. 13, No. 4, pp. 586-599, April 2004.

[4] G. M. Schuster, X. Li, A. K. Katsaggelos, “Shape Error Concealment Using Hermite Splines,” *IEEE Transactions on Image Processing*, Vol. 13, No. 6, pp. 808-820, June 2004.

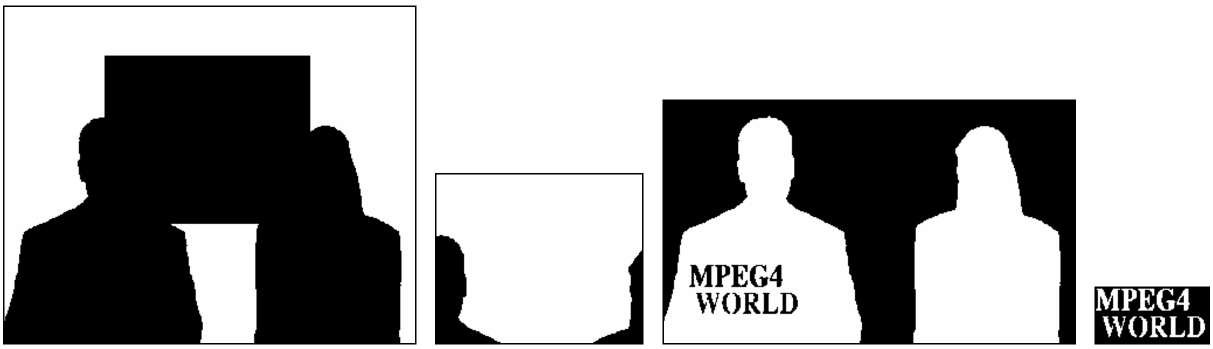
[5] P. Salama, C. Huang, “Error Concealment for Shape Coding,” *Proc. of the IEEE International Conference on Image Processing (ICIP 2002)*, Rochester, NY, USA, Vol. 2, pp. 701-704, September 2002.

[6] L. D. Soares, F. Pereira, “Motion-based Shape Error Concealment for Object-based Video,” *Proc. of the IEEE International Conference on Image Processing (ICIP 2004)*, Singapore, October 2004.

[7] L. D. Soares, F. Pereira, “Combining Space and Time Processing for Shape Error Concealment,” *Proc. of the Picture Coding Symposium (PCS 2004)*, San Francisco, CA, USA, December 2004.

[8] L. D. Soares, F. Pereira, “Spatial Texture Error Concealment for Object-based Image and Video Coding,” *Proc. of the EURASIP Conference on Signal and Image Processing, Multimedia Communications and Services (ECSIPM 2005)*, Smolenice, Slovak Republic, June 2005.

[9] R. C. Gonzalez, R. E. Woods, *Digital Image Processing*, 2nd Ed., Prentice-Hall, 2002.



(a)

(b)

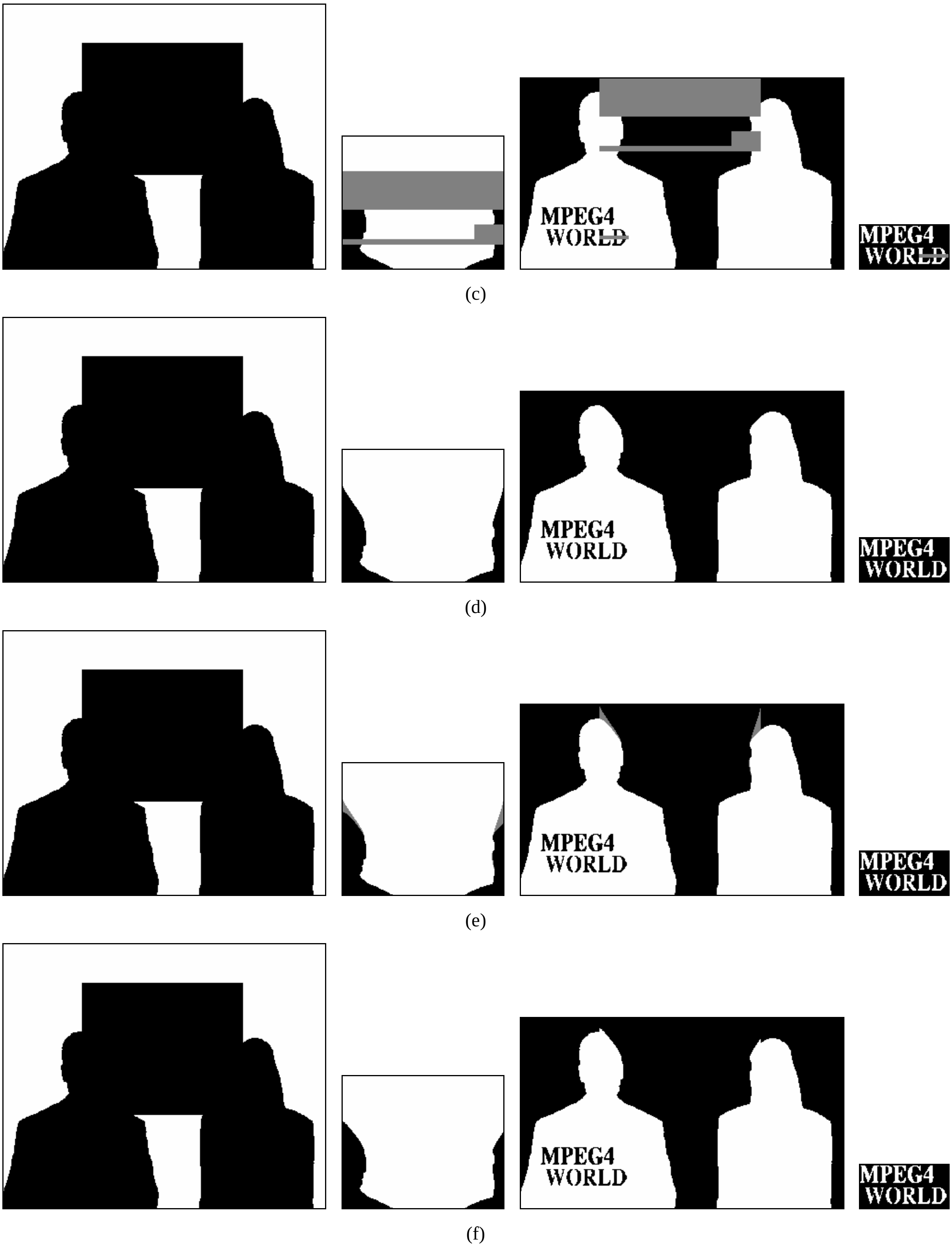


Figure 6. Illustration of the shape concealment process for the News video scene: (a) Original uncorrupted video objects (Background, Dancers, Speakers and Logo); (b) Corrupted video objects; (c) Video objects after the corrupted shape for which complementary data exists has been concealed; (d) Video objects after individual concealment; (e) Undefined shape regions that appear after individual concealment; (f) Final concealed video objects.