COMBINING SPACE AND TIME PROCESSING FOR SHAPE ERROR CONCEALMENT

Luis Ducla Soares, Fernando Pereira
Instituto Superior Técnico / Instituto de Telecomunicações
E-mail: lds@lx.it.pt, fp@lx.it.pt

ABSTRACT

Several shape error concealment techniques for object-based video have already been proposed in the literature. These techniques, however, are either spatial or temporal, in the sense that only information from one of these domains is considered during the concealment process. While spatial error concealment techniques only use spatially adjacent shape information to perform the concealment, temporal error concealment techniques rely on shape information from other time instants. This means that these techniques inherently have the limitations of the specific domain being considered and, therefore, cannot be used with acceptable results for all possible situations.

In this paper, an adaptive spatio-temporal shape error concealment technique is proposed, which overcomes the drawbacks of working only in one domain, spatial or temporal, by adaptively choosing the most adequate type of concealment for each situation.

1. INTRODUCTION

The emergence of the MPEG-4 object-based audiovisual coding standard [1] opened up the way for new video services, where scenes are understood as a composition of objects. However, 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 both shape and texture data are needed.

In terms of shape concealment, several techniques have already been proposed in the literature, which can be divided in two major categories, depending on the information that is used for the concealment process. On one hand, spatial error concealment techniques, such as those proposed in [2-4], only use spatially adjacent shape information to perform the concealment and do not use any shape information from other time instants. This makes these techniques especially useful when the shape changes greatly in consecutive time instants, such as when new objects appear or are uncovered. On the other hand, temporal error concealment techniques, such as those proposed in [5-7], rely on shape information from other time instants to perform the concealment. Since, in most video object sequences, the shape data does not change that much in consecutive time instants, temporal techniques are typically able to achieve better concealment results than spatial techniques.

The main problem with these two types of shape error concealment techniques is that they cannot be used with acceptable results for all possible situations. In this paper, an adaptive spatio-temporal shape error concealment technique is proposed, which overcomes the drawbacks of working only in one domain, spatial or temporal.

2. PROPOSAL FOR A SPATIO-TEMPORAL SHAPE ERROR CONCEALMENT METHOD

The adaptive spatio-temporal shape concealment technique proposed here is based on two previously proposed shape concealment techniques by the same authors.

The first technique, spatial shape concealment [4], assumes that most contour segments are rather smooth and uses Bézier curves to interpolate those contours that were broken due to channel errors.

As for the second technique, temporal shape concealment [7], it is based on a combination of global and local motion compensation. First, it starts by assuming that the alpha plane changes occurring in consecutive time instants can be described by a global motion model and simply tries to conceal the corrupted alpha plane blocks by using the corresponding blocks in the global motion compensated previous alpha plane. However, since not all alpha plane changes can be perfectly described by global motion, an additional local motion refinement is then applied to deal with areas of the object that have significant local motion.

By designing a scheme that adaptively selects one of the two proposed shape concealment techniques depending on the shape characteristics, it should be possible to obtain the advantages of both solutions while compensating for their disadvantages and, thus, improve the final concealed shape quality. Without this adaptive method, one of the two proposed techniques would have to be chosen at the beginning of the communication, which can have its share of problems if the video characteristics change along the way (e.g., the alpha plane of a given video object can start by having a very smooth contour and, later on, become spatially very complex but with a well-behaved motion or different objects may have different behaviors). With the adaptive method, however, it is possible to perform concealment in a much more flexible and, hence, effective way. For instance, the following situations can be better dealt with:

- **Video objects with different characteristics –** Since the various video objects in the scene do not necessarily have similar characteristics, this means that different concealment techniques can be applied to different objects.
- **Time-varying characteristics of objects –** Additionally, since the characteristics of the video objects also change in time, the used shape concealment techniques can be adapted accordingly.
- **Space-varying characteristics of objects –** In fact, even within a given video object at a given time instant, different techniques can be used for the different parts of the video object, which may have different spatial and temporal variation characteristics.

This way, the block diagram for the proposed adaptive spatio-temporal shape concealment technique is presented in Figure 1. The first module is responsible for determining for each corrupted area to be concealed in the corrupted alpha plane which type of concealment technique, i.e., spatial or temporal, is the most adequate. Afterwards, the concealment itself is performed according to the selected approach. A detailed description of the spatial and temporal shape concealment techniques to be used (gray blocks in Figure 1) can be found in [4] and [7], respectively.

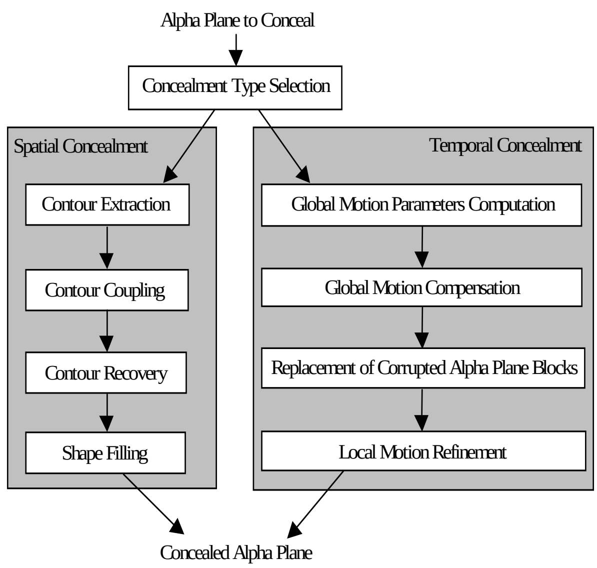


Figure 1 – Proposed adaptive spatio-temporal shape concealment process

2.1 Concealment Type Selection

In order to define an algorithm to select the most adequate shape concealment technique for each independent corrupted area in the alpha plane, the characteristics of the two proposed techniques should be considered. While the spatial concealment technique is only able to produce accurate concealment results when the corrupted segment of the contour to be concealed is relatively short, the temporal concealment has no such limitation and can produce accurate results independently of the length of corrupted contour segment, as long as the object is relatively well-behaved in terms of motion. In fact, for the cases where the corrupted contour segment is short, the spatial concealment technique is typically able to achieve comparable (or even better) results than the temporal technique. And, therefore, for such cases, the spatial concealment technique can be used independently of how well-behaved the object is in terms of motion. On the other hand, when the corrupted contour segment is rather long, the results obtained with the spatial concealment are expected to be worse than the ones obtained with the temporal concealment. This is especially true if the motion is well-behaved but, even if it is not, the results will still be acceptable due to the local motion refinement process included in the overall motion-based concealment [7]. Still for the case where the corrupted contour segment is rather long, when the motion is so complex that the temporal concealment technique is not able to produce good results, it is also unlikely that the spatial concealment technique will be able to do so.

The arguments above and exhaustive testing show that the decision of which concealment technique shall be applied to a given corrupted area should depend on the *length of the longest contour segment* that has been lost inside the corrupted area in question and has to be concealed. The problem with this is that this length cannot be determined since the contour has been lost in the corrupted area. However, it can be estimated by considering the factors on which it depends. For this, the following two factors were identified:

- **Distance between coupled contour endings –** The first factor on which the length of the corrupted contour segment depends is clearly the distance that separates the two corresponding contour endings. This is clearly illustrated in Figure 2, where the two corrupted alpha planes have corrupted areas with the same size but correspond to very different lengths of corrupted contour segments. For corrupted areas where more than two contour endings exist, the decoder has first to couple the available contour endings for each corrupted segment according to the procedure defined in [4] and then decide for which pair the contour endings are farther apart.

- **Size of the corrupted area –** However, the length of the corrupted contour segment does not depend only on the distance that separates the corresponding contour endings. In addition, it also depends on the size itself of the corrupted area. This happens because large corrupted areas may hide additional important information, which cannot be simply inferred from the contour endings alone. For instance, in Figure 3, three different corrupted alpha planes are shown with one corrupted area; in all three alpha planes, the distance that separates the coupled contour endings is similar. However, the length of the corrupted contour segment inside the corrupted area is very different. As can be easily understood, for the cases with larger corrupted areas, the spatial concealment technique will very unlikely yield a good result.

(a) (b)

Figure 2 – Examples of corrupted alpha planes where the corrupted areas have the same size, but the length of the corrupted contour segment is very different – (a) Short corrupted contour segment; (b) Long corrupted contour segment

(a) (b) (c)

Figure 3 – Examples of corrupted alpha planes where the distance separating the coupled contour endings is similar, but the size of the corrupted areas is very different – (a) Small corrupted area; (b) Large corrupted area; (c) Larger corrupted area

Therefore, based on these two factors that influence the usability of the spatial concealment technique for a given corrupted area, it is proposed that the decision of which type of shape error concealment should be used for each corrupted area be performed based on the metric S defined as:

$$S = d_{\max} \times A_{corrupted} \tag{1}$$

where d_{max} is the (Euclidean) distance, measured in shapels, separating the coupled contour endings farthest apart for the corrupted area in question and $A_{corrupted}$ is the size of the corrupted area in shape blocks. This way, to decide which shape concealment technique should be used for a given corrupted area, the decoder has to compare S with a threshold S_{th} according to the following rule:

$$\begin{cases} \text{if } S \le S_{th}, & \text{use the spatial shape concealment} \\ \text{if } S > S_{th}, & \text{use the temporal shape concealment} \end{cases} \tag{2}$$

As for the threshold S_{th}, it has to be determined according to the type of shape data in the considered video content. After all, the definition of what is a short corrupted contour segment and what is a small corrupted area are intrinsically related to the spatial variation characteristics of the shape data. As shall be seen later in Section 3, exhaustive tests have shown that, for most shape data corresponding to the applications where objects may be used (MPEG-4 object-based test sequences include a large variety of object-based content, notably from videotelephony, surveillance, and TV), a value of 138 should be used for the S_{th} threshold.

2.2 Adaptive Concealment of the Corrupted Alpha Plane

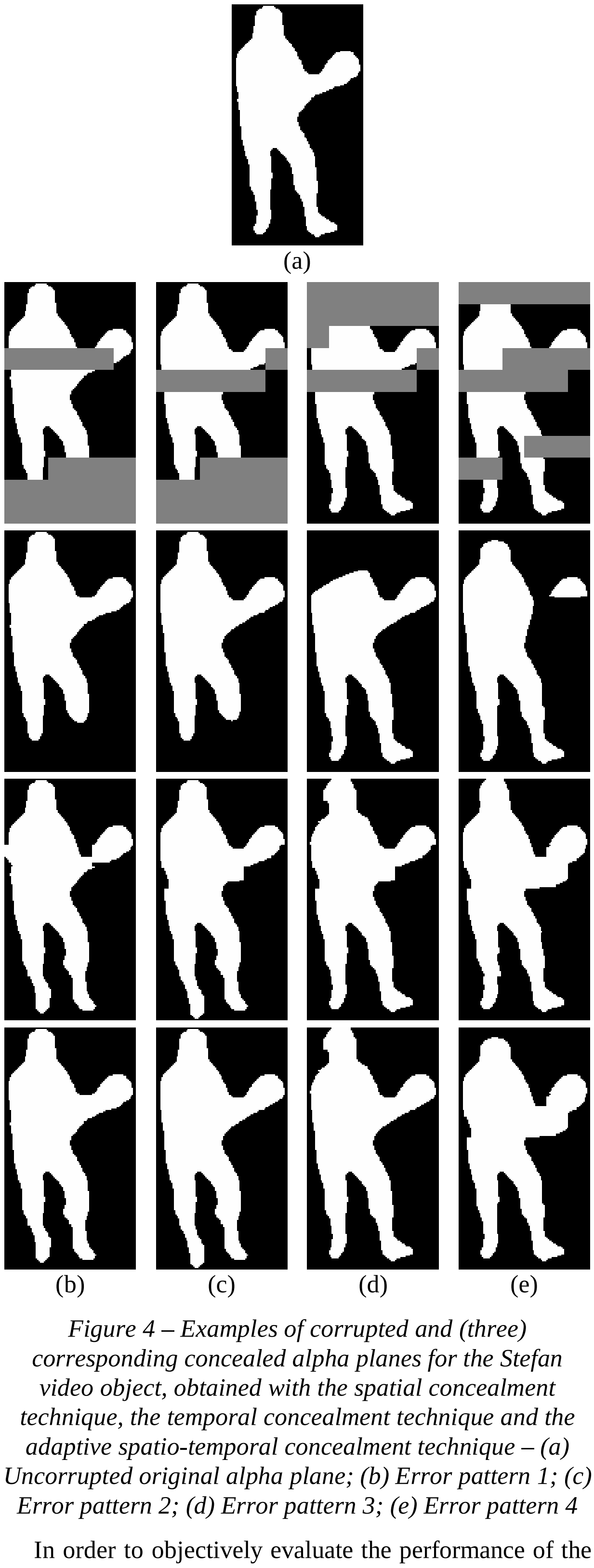
Finally, after the decision of which shape concealment technique (spatial or temporal) should be used has been taken, it is time for the concealment itself to be applied to each corrupted area. This is simply a matter of proceeding as explained in [4] and [7], where the considered spatial and temporal concealment techniques are described. This way, to conceal a given corrupted area in the alpha plane with the spatial technique, the concealment process described in [4] is followed. If on the other hand, the temporal technique is to be used, the concealment process described in [7] is followed.

Before giving some performance results, it should be noted that the encoder can positively influence the concealment decisions (and results) at the decoder, especially if it knows about the shape concealment selection method used at the decoder. To do this, the encoder has to carefully decide how to divide the video object data in several independently decodable units called video packets. For instance, if the encoder knows that a given part of the alpha plane, if corrupted, should be concealed with the spatial technique, it should use small video packets for that area in order for small corrupted areas to appear at the decoder. On the other hand, if it knows that a given area can, if corrupted, be easily concealed with the temporal concealment technique, it can use large video packets instead (and thus save the overhead associated with video packets).

3. PERFORMANCE EVALUATION

As explained in Section 2.1, before applying the adaptive spatio-temporal concealment technique to a given type of shape data, the S_{th} threshold has to be determined. For most shape data corresponding to the applications where objects may be used (represented by the MPEG-4 object-based test sequences), the S_{th} threshold was experimentally determined to be 138. This value was obtained by using Equation (1) and considering that the worst situation that can be well concealed with the spatial concealment technique corresponds to a corrupted area where the d_{max} value corresponds to the distance of a 16×16 shape block diagonally traversed (i.e., 23 shapels) and the critical $A_{corrupted}$ value corresponds to 6 shape blocks.

The first video object sequence on which the proposed adaptive spatio-temporal concealment technique was tested is the Stefan video object, of which a few illustrative examples are shown in Figure 4. These selected examples adequately reflect the obtained concealment results and, therefore, can be used to illustrate the performance of the method. In Figure 4 (a), the uncorrupted original alpha plane of a given decoded VOP from the Stefan video object is shown (VOP 2 in the original 300 VOP sequence). In the remainder of Figure 4, four different corrupted versions of the alpha plane in Figure 4 (a) are shown. Below each one of these corrupted alpha planes, three concealed alpha planes are included, obtained by using the spatial concealment technique [4], the temporal concealment technique [7] and the adaptive spatio-temporal concealment method proposed here. In Figure 4 (b) and (c), when the adaptive method is used, Stefan's feet are concealed with the temporal concealment technique and the rest of the alpha plane is concealed with the spatial concealment method. In Figure 4 (d), also for the adaptive method, Stefan's head is concealed with the temporal concealment technique, while the rest of the corrupted alpha plane is concealed with the spatial concealment technique. Finally, in Figure 4 (e), Stefan's head, as well as his legs, are concealed with the spatial technique; the rest is concealed with the temporal concealment technique. As can be seen from these examples, by using the adaptive spatio-temporal concealment technique, the subjective impact is generally improved. The only exception to this is Figure 4 (e), where the corrupted area corresponding to Stefan's head has been concealed with the spatial interpolation technique; better results would have been obtained with the temporal concealment technique, which has been used on the rest of the corrupted areas in that alpha plane. This less adequate decision could be changed by decreasing S_{th} but this lower value would create many more problems for other corrupted areas and, therefore, it was decided to not change the threshold.



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

Figure 4 – Examples of corrupted and (three) corresponding concealed alpha planes for the Stefan video object, obtained with the spatial concealment technique, the temporal concealment technique and the adaptive spatio-temporal concealment technique – (a) Uncorrupted original alpha plane; (b) Error pattern 1; (c) Error pattern 2; (d) Error pattern 3; (e) Error pattern 4

In order to objectively evaluate the performance of the adaptive spatio-temporal concealment technique, the *Dn* shape distortion metric used by MPEG can also be used here (the lower *Dn*, the better). This metric is defined as the number of different shapels between the decoded and original alpha planes divided by the total number of opaque shapels in the original alpha plane; it can also be expressed as a percentage,

$$Dn[\%] = 100 \times Dn\,, \tag{3}$$

as is the case in Table 2, where the values associated with the concealed alpha planes in Figure 5 are shown. As can be seen, in all but one case, the *Dn* value obtained when the adaptive spatio-temporal concealment technique is used is lower than for either one of the two individual techniques. The only exception happens for error pattern 4, where the *Dn* value achieved with the temporal concealment technique is lower than the one obtained with the adaptive spatio-temporal concealment technique.

Table 1 – Dn values for the various concealed alpha planes shown in Figure 4, obtained when the various shape concealment techniques are used

Concealment technique	Dn [%]			
	Error pattern 1	Error pattern 2	Error pattern 3	Error pattern 4
Spatial concealment only	9.52	9.77	12.63	9.87
Temporal concealment only	9.84	10.56	4.30	9.01
Adaptive spatio-temporal concealment method	8.67	9.27	2.99	9.56

The second video object sequence on which the proposed adaptive spatio-temporal concealment technique was tested is the First Man video object (of the Hall Monitor sequence), of which a few illustrative examples are shown in Figure 5. In Figure 5 (a), the uncorrupted original alpha plane of a given decoded VOP from the First Man video object is shown (corresponding to VOP 33[1] in the original 300 VOP sequence). The remainder of Figure 5 is organized as Figure 4. In Figure 5 (b), when the adaptive spatio-temporal concealment method is used, the First Man's head is concealed with the temporal concealment technique, while the foot is concealed with the spatial concealment method. In Figure 5 (c), also for the adaptive method, all the corrupted areas in the alpha plane are concealed with the temporal concealment technique. In Figure 5 (d), only the First Man's briefcase is concealed with the temporal concealment technique, while the rest of the corrupted alpha plane is concealed with the spatial concealment technique. Finally, in Figure 5 (e), only the First Man's upper torso is concealed with the temporal concealment technique; the rest is concealed with the spatial concealment technique. As can be seen from these examples, by using the adaptive spatio-temporal concealment technique, the subjective impact is generally improved. The only exception to this is Figure 5 (c), where all the corrupted areas in First Man have been concealed with the temporal concealment technique and, therefore, the subjective impact is exactly the same as when only the temporal concealment technique is used. This, however, does not represent a wrong decision by the decoder because, after all, the temporal concealment technique was the best choice in this case for all the corrupted areas. As for the *Dn* values associated with these concealed alpha planes, they are shown in Table 2 and correctly reflect what happens in terms of subjective impact. As can be seen, in all but one case, the *Dn* value obtained when the adaptive spatio-temporal concealment technique is used is lower than for either one of the two individual techniques. The only exception happens for error pattern 2, where the *Dn* value achieved with the temporal concealment technique is identical to the one obtained with the adaptive spatio-temporal concealment technique, which is understandable since all the corrupted areas have been concealed with temporal concealment technique.

Finally, from these two examples, it can be concluded that by using the adaptive spatio-temporal concealment technique, the subjective impact, as well as the *Dn* values, are generally improved. This just confirms that investing in the adaptive spatio-temporal concealment method was worthwhile as expected since the drawbacks of the elementary concealment techniques are compensated and the benefits accumulated.

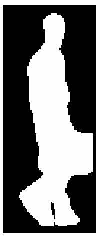
4. FINAL REMARKS

In this paper, two previously proposed shape error concealment techniques by the same authors are considered. Based on these two techniques, a method to adaptively combine them into an even more effective shape concealment scheme is proposed. By adaptively choosing one of the two concealment techniques (spatial or temporal) to be used to conceal a given corrupted area in an alpha plane, the achieved concealment results have been further improved to acceptable shape fidelity levels, thus increasing the quality of the object-based multimedia experience to be provided in error prone environments.

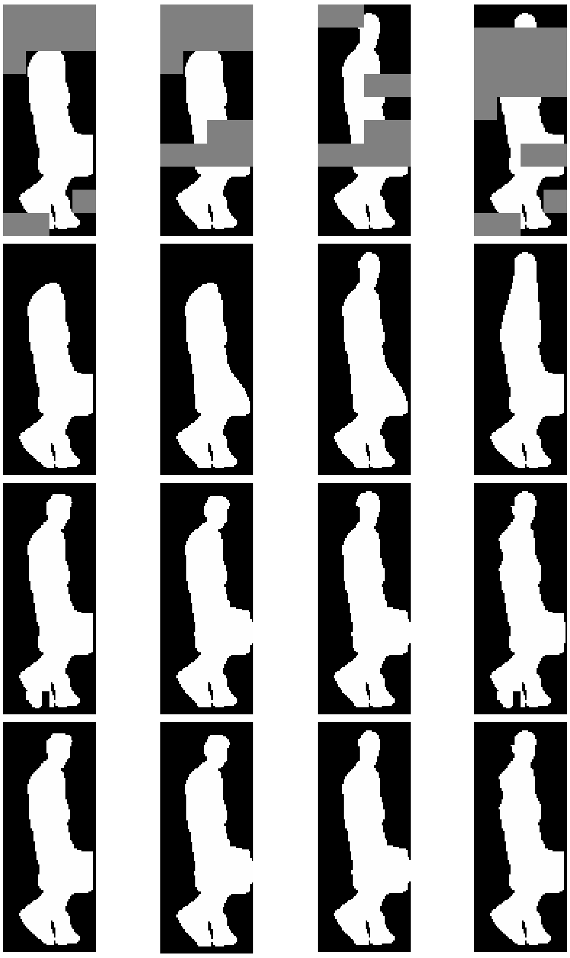
5. ACKNOWLEDGMENT

The work presented was developed within VISNET, a European Network of Excellence (http://www.visnet-noe.org).

[1] This VOP was used because the first few VOPs in the video object sequence are empty.



(a)



(b) (c) (d) (e)

Figure 5 – Examples of corrupted and (three) corresponding concealed alpha planes for the First Man video object, obtained with the spatial concealment technique, the temporal concealment technique and the adaptive spatio-temporal concealment technique – (a) Uncorrupted original alpha plane; (b) Error pattern 1; (c) Error pattern 2; (d) Error pattern 3; (e) Error pattern 4

Table 2 – Dn values for the various concealed alpha planes shown in Figure 5, obtained when the various shape concealment techniques are used

Concealment technique	Dn [%]			
	Error pattern 1	Error pattern 2	Error pattern 3	Error pattern 4
Spatial concealment only	7.78	10.60	3.12	5.35
Temporal concealment only	6.48	6.30	2.28	5.92
Adaptive spatio-temporal concealment method	4.17	6.30	2.13	3.23

6. 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] G. M. Schuster, X. Li, A. K. Katsaggelos, "Spline Based Boundary Loss Concealment," *Proc. of the International Conference on Image Processing*, Barcelona, Spain, September 2003.

[4] 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.

[5] M. R. Frater, W. S. Lee, M. Pickering, J. F. Arnold, "Error Concealment of Arbitrarily Shaped Video Objects," *Proc. of the International Conference on Image Processing*, Vol. 3, pp. 507-511, Chicago, IL, USA, October 1998.

[6] P. Salama, C. Huang, "Error Concealment for Shape Coding," *Proc. of the International Conference on Image Processing*, Vol. 2, pp. 701-704, Rochester, NY, USA, September 2002.

[7] L. D. Soares, F. Pereira, "Motion-based Shape Error Concealment for Object-based Video," *Proc. of the International Conference on Image Processing*, Singapore, October 2004.