RECOVERY OF TWO TRANSPARENT PRIMITIVE IMAGES FROM TWO FRAMES

Javier Toro, Rubén Medina† *and Djemel Ziou*

DMI, Université de Sherbrooke
Sherbrooke (Québec) J1K 2R1, Canada
†GIBULA, Universidad de Los Andes
Mérida 5101, Venezuela

ABSTRACT

The problem of recovering two primitive images from their transparent combination is explored. Given an image sequence, the case where each primitive image undergoes a dissimilar and invertible motion over time is considered. The authors show that discrete samples of the primitive images can be recovered from the observation of two consecutive image frames. In this context, consideration is given to the density of the recovered sample set and the requirements for proper reconstruction. On the whole, transparency separation and recovery is an ill-posed inverse problem in the sense that its solution is not unique. Recovery, however, can be successfully performed by finding the smoothest solution consistent with the data. Illustrative examples show that the proposed method provides good estimates of the primitive images.

1. INTRODUCTION

Transparency refers to the phenomenon where an image is perceived as the combination of two or more primitive images. The origin of such phenomenon is varied. Some common sources are specular reflections, dark filters, puffs of smoke, gauze curtains, and cast shadows [1].

Although there are many situations where transparency is observed, we are primarily interested in the case where primitive images move over time. A number of techniques have been designed to cope with this particular situation. Among them, there are techniques concerned with motion estimation [2, 3], detection of transparent layers [4] and the recovery of such layers [5]. Here, our attention will be focused on the latter, that is, on how to estimate the primitive images from their mixture.

A prototypical approach for transparency separation and recovery has been temporal integration after registration [3]. More recently, new operators have been proposed to replace the temporal integration procedure [5, 6]. Szeliski *et al.* [5] developed two techniques. In their first approach a constrained least square estimation procedure is used to recover the primitive images. In the second, minimum- and maximum-compositing operations are used to achieve the same task. In the same line, Weiss [6] proposed an approach to recover a single reflectance image given a number of observations where the illumination is the only changing primitive. The approach is formulated as a maximum-likelihood estimation problem where a temporal mean is applied to filtered versions of the observed images. These methods, however, require a large number of frames to achieve a clear recovery of the primitive images. Looking at the problem from a different perspective, Vernon [7] proposed an approach for recovering two primitive images using a considerably reduced number of frames. The method operates in the Fourier domain by relying on the fact that a translation of the space domain becomes a multiplication by an exponential of the frequency domain. A conspicuous limitation of this method is that it is restricted to scenes where motion processes are purely translational.

In this paper, we provide a better insight to previous results reported by the authors on the problem of transparency separation and recovery [8]. Particularly, we expand on issues related to the recovery of primitive images when large motions are observed. Altogether, the approach provides a general mechanism to recover two moving primitive images from two consecutive image frames.

2. PROBLEM FORMULATION

To begin our discussion, let $g_j : \mathbb{R}^2 \to \mathbb{R}$ be a primitive image whose spectra vanishes outside $\mathcal{B} = \{\bar{\omega} : |\bar{\omega}| \leqslant \omega_0\}$ and let $\varphi_j : \mathbb{R}^2 \to \mathbb{R}^2$ be an invertible map that describes its motion over time. Then a time-discrete image sequence generated as the transparent combination of two primitive images can be written as

$$f_i(\bar{x}) = g_1(\varphi_1^{i-1}(\bar{x})) + g_2(\varphi_2^{i-1}(\bar{x})), \quad i = 1, 2, 3, \cdots \tag{1}$$

where $\varphi_j^n(\bar{x})$ denotes the n compositions of the mapping on itself.

The aim is to determine the primitive images $g_1(\bar{x})$ and $g_2(\bar{x})$ given two consecutive image frames. We assume in this paper that our observation of the image frames is noise-free and that all motion processes are known. Although these assumptions are not normally met in usual conditions, existing filtering and motion estimation [2] techniques may alleviate this limitation.

3. TRANSPARENCY RECOVERY

3.1. Preliminary Definitions

We first review some concepts in difference calculus that are core to the method proposed here. A fundamental concept of this theory is the finite difference [9], which can be regarded as the discrete analogue to the derivative. More specifically, let $y : \mathbb{Z}^* \to \mathbb{R}$ be a discrete function. Then the finite backward difference operator $\Delta : \mathbb{R} \to \mathbb{R}$ is defined as

$$\Delta y(i) = y(i) - y(i-1). \tag{2}$$

This relationship connects the backward difference of a function with two successive values of that function. Conversely, the term integration or antidifference is used to denote the process of finding a function $y(i)$ whose difference is a given function $Y(i)$. Such operation is denoted by $y(i) = \Delta^{-1}Y(i)$. Function $y(i)$ is called the indefinite finite integral of $Y(i)$. When definite limits are applied we have the definite finite integral of $Y(i)$, which can be expressed as

$$\Delta^{-1}Y(i)\mid_1^n = y(n) - y(0) = \sum_{i=1}^{n} Y(i). \tag{3}$$

This relationship establishes a connection between the quantity $y(n) - y(0)$ and the sum of n terms of the given function $Y(i)$. If $y(0)$, the initial condition, is known, then the function $y(i)$ can be uniquely determined.

3.2. The Separation Scheme

We now turn to the problem of transparency separation and recovery. Consider a composite image sequence where one primitive image remains static and the other transforms according to a given motion map. We write the first two frames of this time-discrete sequence as

$$f_1(\bar{x}) = g_1(\bar{x}) + g_2(\bar{x}) \tag{4a}$$
$$f_2(\bar{x}) = g_1(\bar{x}) + g_2(\phi(\bar{x})). \tag{4b}$$

This particular case may seem restrictive when compared to that of Eq. (1). However, as motion is relative, whatever the given motion map $\varphi_1(\bar{x})$ may be, the image frame $f_2(\bar{x})$ of Eq. (1) can be transformed to make the primitive image $g_1(\bar{x})$ appear at rest when compared to that of the image frame $f_1(\bar{x})$. In which case, mapping $\phi(\bar{x})$ can be written in terms of $\varphi_1(\bar{x})$ and $\varphi_2(\bar{x})$ as $\phi(\bar{x}) = \varphi_2(\varphi_1^{-1}(\bar{x}))$. Consider now the discrete set of points $J_{\bar{x}_0}$, for some $\bar{x}_0 \in \mathbb{R}^2$, given by

$$J_{\bar{x}_0} = \{\bar{x}_0^i : \bar{x}_0^i = \phi^{-1}(\bar{x}_0^{i-1}),\ \bar{x}_0^0 = \bar{x}_0,\ i = 0, 1, 2, \cdots\}. \tag{5}$$

The set of points produced in this fashion is referred to as a discrete trajectory. Notice that the set $J_{\bar{x}_0}$ is isomorphic to the set $\mathbb{Z}^*$. As such, operations defined on functions that have $\mathbb{Z}^*$ as the definition set can also be formulated on functions defined on $J_{\bar{x}_0}$. Keeping this in mind, we define the following differences:

$$G_1(\bar{x}) = f_2(\bar{x}) - f_1(\phi(\bar{x})) \tag{6a}$$
$$G_2(\bar{x}) = f_1(\bar{x}) - f_2(\bar{x}). \tag{6b}$$

If we then restrict $G_1(\bar{x})$ and $G_2(\bar{x})$ to the definition set $J_{\bar{x}_0}$, we obtain:

$$G_1(\bar{x}_0^i) = g_1(\bar{x}_0^i) - g_1(\bar{x}_0^{i-1}) \tag{7a}$$
$$G_2(\bar{x}_0^i) = g_2(\bar{x}_0^i) - g_2(\bar{x}_0^{i-1}). \tag{7b}$$

We can then obtain the values of $g_1(\bar{x})$ and $g_2(\bar{x})$, for any $\bar{x} \in J_{\bar{x}_0}$, by applying the antidifference operator to these two expressions. The recovery, however, is unique up to an additive constant value – the unknown initial conditions. Note that we obtain samples of the primitive images by integrating along a predetermined trajectory in the image given by the set $J_{\bar{x}_0}$. We can explore the whole image by choosing sufficient starting points $\bar{x}_0$, for example, along a line crossing the trajectory described by $J_{\bar{x}_0}$. We denote the totality of sampling points, generated through the set of starting points $\{\bar{x}_{0_i}\}_{i=0}^m$, as $J = \bigcup_{i=0}^m J_{\bar{x}_{0_i}}$. For each of these trajectories, the recovered samples of each primitive image have their own unknown initial condition. The totality of these unknowns is denoted here by the set Θ. There is therefore a great deal of ambiguity in the recovery of the primitive images as there is nothing to relate the integrals along adjacent trajectories. This ambiguity can be removed if additional constraints are introduced.

Before proceeding, notice that in practice our observation of a given image is confined within the bounds of a viewing frame, which can be denoted as a closed region $\Omega \subset \mathbb{R}^2$. The proposed separation scheme operates by relating two values of an image pattern, each given at a particular position in the image domain. Consequently, for the separation scheme to be able to function, given an $\bar{x} \in \Omega$ there must exist a $\bar{y}$ also in Ω such that one is the image of the other under transformation ϕ^{-1}. The aggregate of these points is given by the set $\Gamma = \{\,\bar{x} : (\bar{x} \in \Omega \wedge \phi^{-1}(\bar{x}) \in \Omega) \vee (\bar{x} \in \Omega\ \wedge\ \phi(\bar{x}) \in \Omega)\}$. Evidently, the region of the primitive images that can be recovered is limited to the set Γ.

3.3. Finding a Unique Solution

A reasonable assumption that can be used to constrain the solution is smoothness. The rational behind this assumption is that, except for a few discontinuities, the intensity of a viewed scene does not change abruptly. The method we use to compute a solution that is both smooth and sufficiently close to the data consists in finding the functions $\hat{g}_1(\bar{x})$ and $\hat{g}_2(\bar{x})$ that minimise

$$\int_\Gamma |\nabla \hat{g}_1(\bar{x})|^2 + |\nabla \hat{g}_2(\bar{x})|^2\, d\bar{x} \tag{8}$$

among $\hat{g}_1(\bar{x})$ and $\hat{g}_2(\bar{x})$ that satisfy

$$\sum_{\bar{x} \in J \cap \Gamma} (f_1(\bar{x}) - \hat{g}_1(\bar{x}) - \hat{g}_2(\bar{x}))^2 \leqslant C. \tag{9}$$

Eq. (8) is a regularisation term that favours smooth solutions. In this functional, ∇ denotes the gradient operator and $|\cdot|$ the Euclidean norm. The constraint of Eq. (9) is aimed at restricting the search to those functions that are close to the data. As the data is assumed noiseless, the constant C is set equal to zero. Functions $\hat{g}_1(\bar{x})$ and $\hat{g}_2(\bar{x})$ are approximation mappings that have the same values as the recovered $g_1(\bar{x})$ and $g_2(\bar{x})$ for $\bar{x} \in J \cap \Gamma$. For simplicity, these mappings are taken here as bilinear interpolants. As they directly depend on the values that the unknown parameters of the set Θ may take, the minimisation simplifies to finding these parameters. The downhill simplex method, implemented in MATLAB, is used to conduct the search. Contrary to the regularisation scheme proposed in [8], where the formulation is first derived for translation motions and then used to face more general cases, the method proposed here is better suited to deal with general motions for it explicitly takes into account the spatial distribution of recovered samples.

Observe that this approach does not provide a unique solution. It can be seen that if $\hat{g}_1(\bar{x})$ and $\hat{g}_2(\bar{x})$ minimise (8) constrained to (9), it is possible to construct two further functions $\hat{g}_1(\bar{x}) + k$ and $\hat{g}_2(\bar{x}) - k$, for k constant, that also are a solution. To remove this ambiguity, one of the unknown parameters of the set Θ is arbitrarily set to a constant value.

3.4. Remarks on the Quality of the Solution

The solution thus obtained does not necessarily reproduce the actual primitive image patterns but an approximation of them. One anticipated source of error is the non-conformity of the primitive images with our prescribed criterion of smoothness [8]. Another source is the interpolation procedure needed to approximate the primitive images at all other points not included in J. In what follows and further on in this section, we assume that images are defined in $\mathbb{R}^2$ and that a non-bounded set J can be constructed. If the set J is not sufficiently dense, in other words, sampling points are sparsely distributed, then our approximation of the images will not be satisfactory whatever the interpolation method used. This problem appears when motions are large. We illustrate what we mean by "large" by considering the simplest motion – translation. Motion map $\phi(\bar{x})$ can then be written as $\phi(\bar{x}) = \bar{x} - \bar{v}$, where $\bar{v}$ is a constant velocity vector. Clearly, for any $\bar{x}_0$, the trajectory described by the set of points in $J_{\bar{x}_0}$ is a straight line. For simplicity, the set J can be constructed by choosing different starting points uniformly spaced along a line at an angle of the trajectory prescribed by vector $\bar{v}$, that is, $\bar{x}_{0_i} = \bar{x}_{0_0} + i\bar{w}$, for some $\bar{x}_{0_0}$ and some $\bar{w} \neq k\bar{v}$. Notice that if $\bar{w}$ is chosen perpendicular to $\bar{v}$, then, by the theory of two-dimensional regular sampling, the primitive images can be exactly reconstructed from their samples if $|\bar{w}| \leqslant \pi/\omega_0$ and $|\bar{v}| \leqslant \pi/\omega_0$. The former can easily be satisfied; the latter, however, depends on the observed motion. We say that $\bar{v}$ is "large" if the latter is not fulfilled. Accordingly, "large" is a notion directly connected to the bandwidth of the signal we are dealing with.

In the general case of arbitrary motions, where points in J may not be uniformly distributed, the results of Clark *et al.* [10] on the reconstruction of functions from nonuniformly spaced samples can be used to assess the density of J. In brief, their result for the two-dimensional case ([10], Theorem 4) can be stated as follows. Suppose that a function of two variables $f(\bar{x})$ is sampled at points in the set J. If there exists a one-to-one continuous mapping $\gamma : \mathbb{R}^2 \to \mathbb{R}^2$ such that the set J is transformed into a uniformly spaced set of sampling positions U, then the function $h(\bar{x}) = f(\gamma^{-1}(\bar{x}))$ can be exactly reconstructed from its samples if its frequency content is enclosed by the baseband described by the sampling set U ([10], condition of Theorem 3: The Uniform Two-Dimensional Sampling Theorem).

A mechanism that can be used to assess the density of J is to compare the frequency content of $f(\bar{x})$ with that of $h(\bar{x}) = f(\gamma^{-1}(\bar{x}))$. Consider a function $f(\bar{x})$ whose Fourier transform satisfies $F(\bar{\omega}) = 0$ if $\bar{\omega} \notin \mathcal{B}$. Assume that the baseband described by the sampling sequence U tightly encloses the frequency content of this function. Now, consider the Jacobian of $\gamma^{-1}(\bar{x})$,

$$A = \frac{\partial \gamma^{-1}(\bar{x})}{\partial \bar{x}}. \tag{10}$$

This quantity can be thought of as an instantaneous linear transformation of $f(\bar{x})$. In the frequency domain, this linear transformation becomes: $f(A\bar{x}) \xrightarrow{\mathfrak{F}} \frac{1}{|A|}F(A^{-1}\bar{\omega})$, where $|A|$ is the determinant of matrix A. In our case, the instantaneous bandwidth of $f(\bar{x})$ for any $\bar{x}$ is constant and given by $\mathcal{B}$. Then, for frequencies $\bar{\nu}$ $(= A^{-1}\bar{\omega})$ such that $F(\bar{\nu}) \neq 0$, i.e., $|\bar{\nu}| \leqslant \omega_0$, we have

$$|\bar{\omega}| = |A\bar{\nu}| \leqslant \|A\||\bar{\nu}| \leqslant \|A\|\omega_0, \tag{11}$$

where $\|A\|$ is the spectral norm of A. Then, for $|\bar{\omega}| \leqslant \omega_0$ to hold, in other words, for the instantaneous frequency content of $f(A\bar{x})$ to lie in $\mathcal{B}$, or equivalently, for the instantaneous frequency content of $f(A\bar{x})$ to be enclosed by the baseband described by the uniform sampling sequence U, the condition

$$\|A\| \leqslant 1 \tag{12}$$

must be satisfied. That is to say, A must be nonexpansive. To illustrate this criterion we look once again at the case of pure translation. Take the uniform sampling sequence U to be of rectangular geometry, with basis vectors satisfying $|\bar{u}_1| = |\bar{u}_2| = \pi/\omega_0$. For simplicity, assume that $\bar{u}_1 = k\bar{v}$, for some $k > 0$, and that the set J is constructed using $\bar{w} = \bar{u}_2$. Then, it is not difficult to verify that for any $\bar{v}$, the mapping $\gamma(\bar{x})$ can be written as a single affine transformation for the whole image domain and that a diagonal form of matrix A can be written as $\mathrm{diag}(1/k, 1)$. Using now the criterion of Eq. (12) we obtain that perfect reconstruction is possible only when $k \geqslant 1$, i.e., when $|\bar{v}| = |\bar{u}_1|/k \leqslant \pi/\omega_0$, as already known. In general, if $\|A\| > 1$, the situation can be remedied by augmenting the density of J. This can be accomplished by including additional sampling points in such a way that the distance between the new added points and their neighbours in the existing set is minimised. Intuitively, we can see that at some point of this process the criterion of Eq. (12) will be met for the whole image domain.

Notice that in practice, reconstruction formulas employ a finite number of samples in a region about $\bar{x}$. An implication is that only an approximate reconstruction can be achieved. This limitation, however, is of no major consequence because the support of an image is finite, anyhow. In addition, the local nature of practical reconstruction formulas turns to our convenience for two reasons. Firstly, the determination of the coordinate transformation $\gamma(\bar{x})$ is greatly simplified (a procedure to obtain $\gamma(\bar{x})$ is given in [10]); and secondly, and more important, Eq. (12) is a local constraint: although it provides a means to draw a conclusion about the instantaneous frequency content of the signal, it does not provide, in general, an indication of the bandlimitedness of the signal globally.

4. EXPERIMENTAL RESULTS

To determine the performance of the proposed method several tests were carried out on real images using synthetic motion maps. For simplicity, each test image sequence was generated according to Eq. (4). The first frame of the mixture is shown in Fig. 1(c). The two primitive images $g_1(\bar{x})$ and $g_2(\bar{x})$ are shown in Figs. 1(a) and 1(b) respectively. An affine transformation was used to describe the motion of $g_2(\bar{x})$. A first test was carried out to explore the performance of the technique when the motion of $g_2(\bar{x})$ is purely translation. In this test, the pixels of this image were moving leftward at 6 pixels/frame. Figs. 1(d) and 1(e) show the recovered primitive images obtained from an undersampled reconstruction of them. As anticipated, the definition of the recovered images is very poor. The rms errors of these images are 9.72 and 25.31 respectively. Figs. 1(f) and 1(g) show the recovered primitive images when the criterion of Eq. (12) is satisfied. At a first glance they are indistinguishable from the original images. Differences are due to the departure of the primitive images from the idealised assumption of smoothness. Both images have an rms error of 8.68. Figs. 1(h) and 1(i) show the recovered images when the motion is more irregular. Here, the pixels of the primitive image $g_2(\bar{x})$ also move leftward but have their velocity linearly varied from 4.5 pixels/frame to 2.5 pixels/frame. We included as many additional

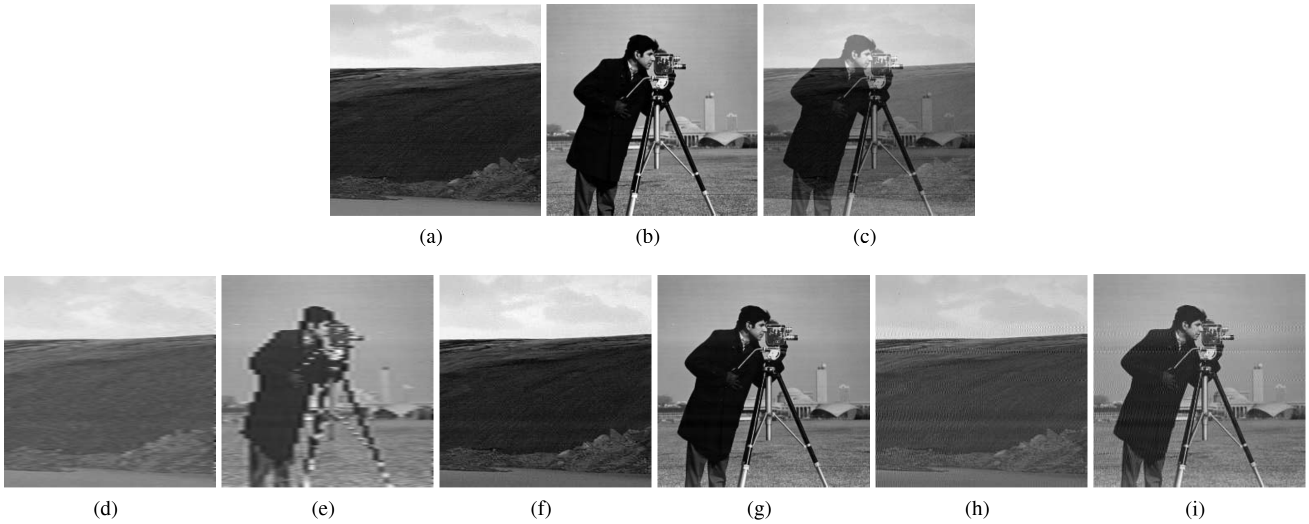

Fig. 1. Experimental results. Figs. (a) and (b) are the original primitive images. The first frame of their mixture is shown in Fig. (c). Results obtained for the case of pure translation are shown in Figs. (d), (e), (f) and (g). Figs. (d) and (e) show the recovered primitive images obtained from undersampled reconstructions. Figs. (f) and (g) show the recovered primitive images obtained from a dense set of samples. Results when a more irregular motion was considered are shown in Figs. (h) and (i).

samples as required to comply with Eq. (12). Although the recovered images look very similar to the original ones, their rms errors are 21.5 and 21.7 respectively. We attribute the differences to the interpolation and integration procedure needed to recover the images at the positions specified by J. In these experiments, bicubic interpolation was used. Clearly, small departures of the image samples from their actual values get accumulated in the integration process. These errors, which are different for each trajectory, may also explain the visible ripples in the recovered images.

5. CONCLUSION

We have presented a mechanism for transparency separation and recovery. The preceding sections of this paper have shown that it is possible to reconstruct two primitive images from two consecutive frames when observed motions are invertible. It has been shown that the primitive images can be perfectly reconstructed along discrete trajectories specified by the observed motions. To be able to recover the whole image, reconstruction along several of these trajectories is required. A criterion has been established for determining if the totality of the chosen trajectories suffices to achieve proper reconstruction. Part of our continuing work in this area is aimed at addressing the problems encountered in the experimental results and the eventual extension to cope with noisy observations.

6. REFERENCES

[1] E. H. Adelson and P. Anandan, "Ordinal characteristics of transparency," in *AAAI Workshop on Qualitative Vision*, 1990, pp. 77–81.

[2] J. R. Bergen, P. J. Burt, R. Hingorani, and S. Peleg, "A three-frame algorithm for estimating two-component image motion," *IEEE Trans. on PAMI*, vol. 14, no. 9, pp. 886–896, Sep 1992.

[3] M. Irani, B. Rousso, and S. Peleg, "Computing occluding and transparent motions," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 5–16, Jan 1994.

[4] T. Darrell and E. Simoncelli, "Separation of transparent motion into layers using velocity-tuned mechanisms," Tech. Rep. 244, Massachusetts Institute of Technology, Oct 1993.

[5] R. Szeliski, S. Avidan, and P. Anandan, "Layer extraction from multiple images containing reflections and transparency," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2000, vol. 1, pp. 246–253.

[6] Y. Weiss, "Deriving intrinsic images from image sequences," in *International Conference on Computer Vision*, Jul 2001, vol. 2, pp. 68–75.

[7] D. Vernon, "Decoupling Fourier components of dynamic image sequences: a theory of signal separation, image segmentation and optical flow estimation," in *European Conference on Computer Vision*, Jan 1998, pp. 69–85.

[8] J. Toro, F. Owens, and R. Medina, "Using known motion fields for image separation in transparency," *Pattern Recognition Letters*, vol. 24, no. 1-3, pp. 597–605, Jan 2003.

[9] V. Lakshmikantham and D. Trigiante, *Theory of difference equations : numerical methods and applications*, Academic Press, Boston, 1988.

[10] J. J. Clark, M. R. Palmer, and P. D. Lawrence, "A transformation method for the reconstruction of functions from nonuniformly spaced samples," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-33, no. 4, pp. 1151–1165, Oct 1985.