Improving the Correspondence Establishment Based on Interactive Homography Estimation*

Xavier Cortés, Carlos Moreno, and Francesc Serratosa

Universitat Rovira i Virgili, Departament d'Enginyeria Informàtica i Matemàtiques, Spain
{xavier.cortes,francesc.serratosa}@urv.cat,
carlosfrancisco.moreno@estudiants.urv.cat

Abstract. We present a method to find the correspondences between salient points of images in which an oracle can interact and influence on the final correspondences between points. The oracle looks up the current point correspondences and imposes new mappings or modifies some of them. This interaction influences on two aspects. On one hand, a new homography is computed such that the Least Square Error is minimized on the imposed correspondences. On the other hand, the Similarity Matrix between the set of points is modified and so, the interaction influences on the output of the Correspondence Algorithm. The method is independent of the algorithm to compute the homography and the correspondences. Practical evaluation shows that in few interactions of the oracle, the optimal correspondence is achieved.

Keywords: Homography Estimation, Iterative Closest Point, Interactive Correspondence, Hungarian Algorithm.

1 Introduction

In recent years, interaction between robots and humans has increased rapidly. Applications of this field are very diverse, ranging from developing automatic exploration sites [1] to using robot formations to transport and evacuate people in emergency situations [2]. Within the area of social and cooperative robots, the nature of interactions between a group of people and a set of accompanying robots has become a primary point of interest [3].

One of the low level tasks that these systems have to face is the automatic recognition of scenes and objects the robot visualises. Usually, the interpretation of scenes is performed trough two steps. Firstly, detecting salient points of the image and secondly, searching for the correspondences of these points and the salient points in the previously learned images [4]. Salient points (that play the role of parts of the image to be matched) are image locations that can be robustly detected among different instances of the same scene with varying imaging conditions. These points can be corners (intersection of two edges) [5], maximum curvature points [6] or isolated points of maximum or minimum local intensity [7]. There is an evaluation of

* This research is supported by the CICYT project DPI 2010-17112.

R. Wilson et al. (Eds.): CAIP 2013, Part II, LNCS 8048, pp. 457–465, 2013.

the most competent approaches in [8]. When salient points have been detected, several correspondence methods can be applied that obtain the homography that maps one image into the other [9], discards outlier points [10] or characterises the image into an attributed graph [11], [12].

Humans are very good at finding the correspondences between local parts of an image regardless of the intrinsic or extrinsic characteristics of the point of view. Current automatic methods to extract parts of images and their correspondences in non-controlled environments are far away of having the performance of a human. For this reason, in this paper, we propose a semi-automatic method in which humans can interact into the system when it is considered the quality of the automatically found correspondence is not good enough and then they impose a partial or total correspondence between some local parts of two scenes.

We advocate for a cooperative model in which robots have cameras and they have to identify objects and scenes in a cooperative manner with the aim of detecting specific objects or performing Simultaneous Localisation And Mapping (SLAM). We assume there is a master system that receives images from robots and finds the correspondences between images of these robots. These images may have been taken from different scenes or from the same scene but with completely different points of view, illuminations and so on. When scenes or objects from scenes have been localized, the master system sends new orders to the robots such as specific movements, grasping objects and so on. When a robot is not able to recognize the scene, then it stops and asks to the master system to localize it or send it the information of the scene. The master obtains the whole information through the correspondence between the current and past images of all the robots. Due to the environment (for instance, rescue inside buildings), we assume it is not possible to localise the robots through GPS, 3D triangulation or sensors on the wheels.

Figure 1 is a screen shot of the master system. We can visualize the cameras of robot 1 and 2 and the salient points automatically extracted. Blue lines represent the automatically extracted correspondences and the orange line represents the correspondence imposed by the human since he or she has realized one of the correspondences was wrong. This interaction influences over the obtained homography and also over the rest of the correspondences.

A similar interaction method was presented in [13]. In that case, there is only one robot and the human decides if a selected part of the image is a face and in the case that it is, imposes the name of the person.

Figure 2 shows a schematic view of our Interactive Correspondence Method (ICM). The user has access to the original images, both set of salient points and the

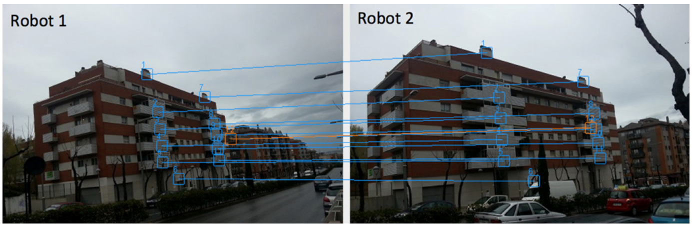

Fig. 1. Screen shot of the interactive system with robot 1 and robot 2 images

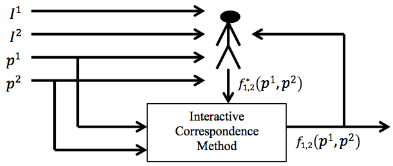

Fig. 2. Basic scheme of our Interactive Correspondence Method (ICM)

current correspondence $f_{1,2}$ between them. Then, the user imposes some point mappings $f^*_{1,2}$ and the output of the module is a correspondence in which the oracle has influenced. This process is repeated until the user considers the correspondence is good enough.

The rest of the paper is organized as follows. The next section summarizes the method to estimate a homography given some correspondence points. Section 3 schematically explains the main steps to find the correspondences of salient points of two images. Section 4 presents our interactive method to estimate the correspondences and section 5 shows the practical evaluation. Section 6 concludes the paper.

2 Homography Estimation Given a Correspondence

Given two images I_1, I_2 and a set of pairs corresponding points on these images $\{\{(x_1, y_1), (x'_1, y'_1)\},..., \{\{(x_n, y_n), (x'_n, y'_n)\}\}$ it is possible to obtain a transformation matrix F that projects each of the points in I_1 into the points in I_2 with a minimal error. Then $[x'_i, y'_i, 1]^T = F \cdot [x_i, y_i, 1]^T$, where $(x'_i, y'_i) \in I_2$ and $(x_i, y_i) \in I_1$.

If we suppose that the mean error has a normal distribution, then the least-square estimation is optimal [14]. Moreover, if we assume an affine transformation, then the transformation matrix is defined as

$$F = \begin{bmatrix} a & -b & c \\ b & a & d \\ 0 & 0 & 1 \end{bmatrix}$$

Where $a = S \cdot cos(\alpha)$, $b = S \cdot sin(\alpha)$, $S = scale$, c = translation in x and d = translation in y.

In this case the error function is usually estimated as follows:

$$E(a, b, c, d) = \sum_{i=1}^{n} [[ax_i - by_i + c - x'_i]^2 + [bx_i + ay_i + d - y'_i]^2]$$

We find the minimum through deriving the function; $\frac{\delta E}{\delta a} = 0$, $\frac{\delta E}{\delta b} = 0$, $\frac{\delta E}{\delta c} = 0$ and $\frac{\delta E}{\delta d} = 0$. We arrive to the following linear problem, A (a, b, c, d) ^{T} + B = 0. Where

$$A = \begin{bmatrix} \sum_i (x_i^2 + y_i^2) & 0 & \sum_i x_i & \sum_i y_i \\ 0 & \sum_i (x_i^2 + y_i^2) & \sum_i -y_i & \sum_i x_i \\ \sum_i x_i & \sum_i -y_i & n & 0 \\ \sum_i y_i & \sum_i x_i & 0 & n \end{bmatrix}, \text{ and } B = \begin{bmatrix} \sum_i (x_i x'_i - y_i y'_i) \\ \sum_i (x'_i y_i - x_i y'_i) \\ \sum_i -x'_i \\ \sum_i -y'_i \end{bmatrix}$$

We solve this linear system using LU factorization [15].

3 Automatic Correspondence Estimation

This section provides the explanation of an automatic correspondence selection based on homography estimation between two set of points.

Figure 3 shows the main steps of the automatic correspondence estimation. Any interest point extraction algorithm like, Harris [16], DoG [17] or LoG [18][19] (or manual detection) can be used to obtain the salient points P_1 and P_2, given images I_1 and I_2. While theses salient points are obtained, we proceed to obtain the homography $H_{1,2}$ between them, for instance using the Iterative Closest Point Algorithm [20]. Then, we project P_1, and obtain P_1' where $P_1' = H_{1,2} \cdot P_1$, to get the most similar possible correlative position of corresponding points from P_1' to P_2. Finally, we compute the similarity matrix $M_{1,2}$ between P_1' to P_2, given the Euclidean Distance between the points (the similarity is computed as the inverse of the distance). The last step is to solve the assignation problem (for instance, the Hungarian algorithm [21]), to get the correspondences $f_{1,2}$ from P_1 to P_2. Note that the correspondences function between P_1' and P_1 is the identity by construction.

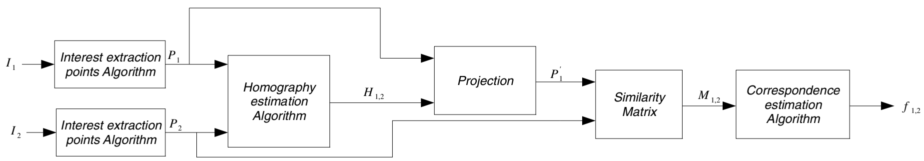

Fig. 3. Scheme of an Automatic Correspondence Estimation model

4 Interactive Correspondence Estimation

Figure 4 presents the main scheme of interactive correspondence estimation. It is based on the correspondence estimation model shown in section 3 but the oracle interaction is considered in two steps (interactive correspondence estimation and interactive similarity restrictions). Dashed lines in the scheme show the interactive part of the model.

Both images I_1 and I_2 and their sets of points P_1 and P_2 are presented to the oracle, together with the current correspondences $f_{1,2}$ and the oracle decides and imposes some correspondences between points $f_{1,2}^*$. Note that $f_{1,2}^*$ can represent a partial labelling between points. So, the oracle is asked to impose only some correspondences. In our experiment section, we suppose the oracle only imposes one correspondence each interaction. The iterative algorithm appends the new points mapping to the ones imposed in the previous iterations and stops when the oracle decides the current correspondences are good enough. The oracles feedback is used to estimate a homography between P_1 and P_2 and also to modify the similarity matrix.

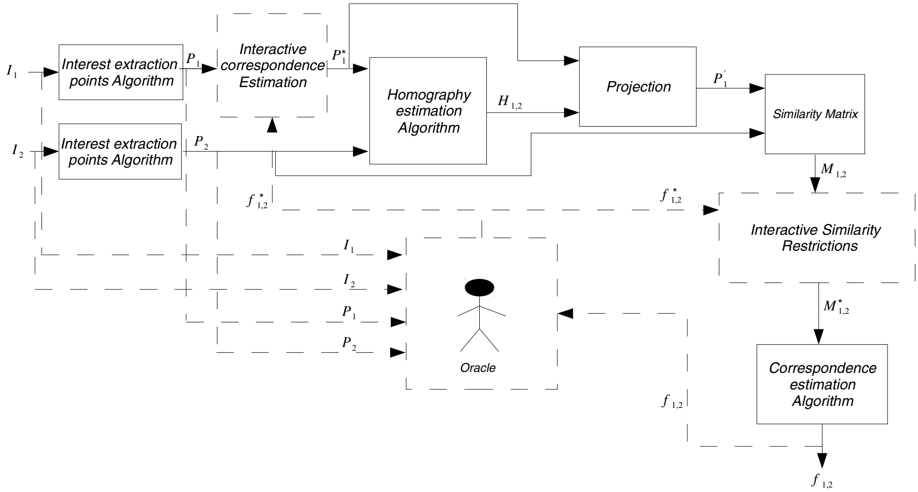

Fig. 4. Scheme of Interactive Correspondence Estimation model

In the first case (figure 5), the scheme of the interactive homography estimation is composed of two steps. In the first one, a homography $H^*_{1,2}$ is deducted (section 2) and in the second step, the set of points P_1 are projected obtaining the new set of points $P^*_1 = H^*_{1,2} \cdot P_1$.

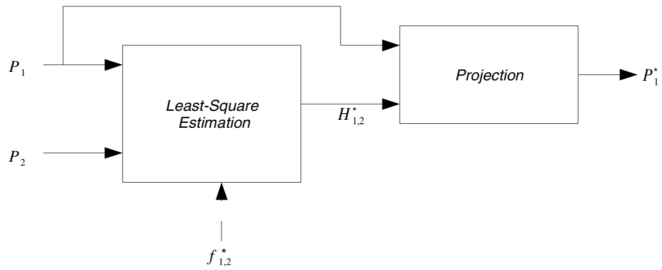

Fig. 5. Scheme of Interactive Homography Estimation

In the second case, the oracle intervenes over the similarity matrix before solving the assignation problem. The above algorithm we use is inspired in the algorithm presented in [22] although in that case, the interactive algorithm modified a cost matrix instead of a similarity matrix. The main idea is that if the oracle considers a pair of points has to be labelled, then, independently of their original similarity, the method imposes their similarity to be the maximum. Moreover, to force the correspondence to be a bijection, the model imposes a zero similarity to the other combinations of correspondences that the pair of points are involved.

Algorithm *Interactive Similarity Restrictions*

Input: *Similarity Matrix* $M_{1,2}$, *Oracle's correspondences* $f^*_{1,2}$

Output: *Similarity Matrix with Restrictions* $M^*_{1,2}$

```
Forall $c_j^i \in f^*_{1,2}$; $c_j^i$ is an oracle's correspondence from i to j
    $M_{1,2}(i, j) = \infty$;
    $M_{1,2}(i, b) = 0; \forall b \neq j$
    $M_{1,2}(a, j) = 0; \forall j \neq j$
End forall
```

We force a rigid transformation between images but non-rigid transformations appear in a real application. For this reason, if the interactive actions only influences on the homography estimation, the model may not achieve the correspondences imposed by the oracle. That is, although the oracle has imposed the matching between all points, it may happen that $f_{1,2} \neq f^*_{1,2}$. We show this effect in the practical evaluation.

5 Practical Evaluation

To experimentally validate our method, we have used four image databases.

APARTMENTS database [23]: It consists of ten images of an apartments building taken from different perspectives. Twelve salient points have been manually extracted from each image (figure 1). Each point represents the same part of the scene through the whole images, presenting a combination of rigid and non-rigid relations between them. The first 11 points represent parts of the building and the 12th point is the highest part of a lamppost.

HOTEL [24], **HOUSE** [24] and **FACES** [25]: They are composed of a set of reference points associated with different frames. Each test has T elements $\{P^t, P'^t, O^t\}, 1 \leq t \leq T$. Each element has a pair of sets of points P^t and P'^t and an oracle matching O^t. Each set of points P^t are the original points in the database. T is the number of sets of points of these datasets. Sets of points P'^t have been randomly generated through distorting P^t to make the automatic correspondence establishment more difficult for the automatic method (similar to [26]). First, we added a global artificial transformation for all points (rotation noise with a uniform probability distribution around 360º and translation noise with a normal distribution respect to initial position). And second, we added an individual noise for each point following a normal distribution. This noise simulates different viewpoints highly separated between them and makes the oracles interaction useful to find a good solution since the transformation between the original frame and the artificial one is non-rigid. Both sets of points P^t and P'^t have the same number of points. The oracle correspondences O^t has been computed together with the generation of the distorted sets of points P'^t. Nevertheless, by construction, we do not guarantee this correspondence to obtain the minimum cost.

The aim of this method is to minimize the distance between the automatically extracted correspondence and an ideal one. We suppose the ideal correspondence is the one that the oracle would impose when he had interacted on all the points. In these databases, these correspondences relate the same parts of the building in different scenes or the same parts of different faces. The metric used to evaluate the goodness of the method is the hamming distance between the ideal and the current correspondence at each iteration.

In the APARTMENTS dataset, each value of the results below is the average of the hamming distance obtained while executing the method for all combinations of pairs of the 10 images. In the HOTEL, HOUSE and FACES datasets, each result value is the average of the hamming distance while executing the method for each image and its corresponding set of points created artificially.

Figures 6 and 7 show the average hamming distance respect the number of iterations in the four different databases. In interaction 0, the correspondences are automatically computed (figure 3). In the interaction 1, the Interaction Homography Estimation module is not used due to two matchings are needed to deduct a unique affine homography. Nevertheless, the imposition on the similarity matrix is performed with the first oracles feedback. The interaction 2 is the first one that the whole model is applied. We show three different curves that represent: 1) the case in which the whole method is used ; 2) the case in which the interactive similarity restrictions are not imposed ; 3) the case in which the interactive correspondence estimation is not imposed (see figure 4).

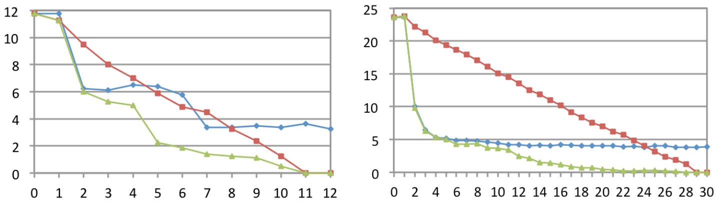

Fig. 6. Hamming distance respect iterations. **a**) APARTMENTS **b**) HOUSE

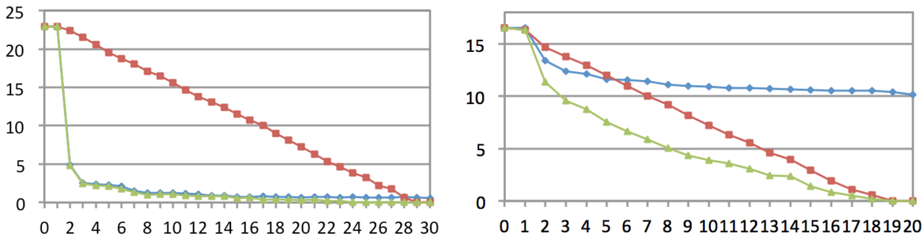

Fig. 7. Hamming distance respect iterations. **a**) HOTEL **b**) FACES

While using ( ), the model establishes an initial homography at second interaction, for this reason the hamming distance presents a steep descent. Throughout the other iterations, the model slightly updates the homography and imposes the restrictions and so the hamming distance decreases until arriving to zero. When restrictions are not imposed ( ), the hamming distance does not arrive to zero. This is because, in a real application, a rigid homography does not exist, which would allow to deduct the ideal correspondence applying only this transformation. Finally, in the case that only the restrictions are imposed but a homography is not estimated using the oracles information ( ), the hamming distances arrives to zero but more slowly than using the whole method. Through these experiments we show the need of using the whole method since one of the most important qualities that an interactive method needs to have is the ability to react in few oracles interaction.

6 Conclusions and Future Work

When a team of robots needs to have consistent shared information of its environment based on its computer vision systems, usually the completely automatic methods fail to find a good enough correspondence between the different parts of the scene. In these cases, an interactive method can improve the correspondence establishment and so increase the consistency of the shared information.

We have presented an interactive method that, given two images taken from different robots, the oracle imposes some correspondences between their local parts to estimate a homography and to restrict the point correspondences.

Results show that in the firsts iterations, there is an important increase on the quality of the correspondences. This result is important since it means it is not necessary a long-term dependence on the oracle but the initial knowledge of the oracle is crucial.

Nowadays, we are modelling the interaction on several sets of points simultaneously with the aim of obtaining a coherent set of homographies between different scenes. Moreover, we are also studding models for non-rigid homographies. Finally, we want to apply active query strategies at it was done in [26] and [27] to suggest the points to be imposed by the oracle.

Finally, this method is going to be expanded with a tracking method based on probabilities [28] and an automatic method to recognise written symbols, such as, “exit” or “elevator”, which is going to be based on a matching graph technique [29].

References

1. Trevai, C., Fukazawa, Y., Ota, J., Yuasa, H., Arai, T.: Cooperative exploration of mobile robots using reaction-diffusion equation on a graph. In: IEEE International Conference on Robotics and Automation, vol. 2 (2003)
2. Casper, J., Murphy, R.R.: Human–robot interactions during the robot-assisted urban search and rescue response at the world trade centre. IEEE Transactions on Systems, Man, and Cybernetics, Part B 33, 367–385 (2003)
3. Garrell, A., Sanfeliu, A.: Cooperative social robots to accompany groups of people. International Journal of Robotics Research 31(13), 1675–1701 (2012)
4. Xu, M., Petrou, M.: 3D Scene interpretation by combining probability theory and logic: The tower of knowledge. Computer Vision and Image Understanding 115(11), 1581–1596 (2011)
5. Tomasi, C., Kanade, T.: Detection and tracking of point features., Tech. Rep. CMUCS-91-132, Carnegie Mellon University (1991)
6. Han, W., Brady, M.: Real-time corner detection algorithm for motion estimation. In: Image and Vision Computing, pp. 695–703 (2005)
7. Rosten, E., Drummond, T.W.: Machine learning for high-speed corner detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 430–443. Springer, Heidelberg (2006)
8. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell. 27(10), 1615–1630 (2005)

9. Zhang, Z.: Iterative point matching for registration of free-form curves and surfaces. Int. J. Comput. Vision 13(2), 119–152 (1994)
10. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24(6), 381–395 (1981)
11. Sanromà, G., Alquézar, R., Serratosa, F., Herrera, B.: Smooth Point-set Registration using Neighbouring Constraints. Pattern Recognition Letters 33, 2029–2037 (2012)
12. Sanromà, G., Alquézar, R., Serratosa, F.: A New Graph Matching Method for Point-Set Correspondence using the EM Algorithm and Softassign. Computer Vision and Image Understanding 116(2), 292–304 (2012)
13. Villamizar, M., Andrade-Cetto, J., Sanfeliu, A., Moreno-Noguer, F.: Bootstrapping Boosted Random Ferns for discriminative and efficient object classification. Pattern Recognition 45(9), 3141–3153 (2012)
14. Pedhazur, E.J.: Multiple regression in behavioral research: Explanation and prediction, 2nd edn. Holt, Rinehart and Winston, New York (1982) ISBN 0-03-041760-0
15. Bunch, J.R., Hopcroft, J.: Triangular factorization and inversion by fast matrix multiplication. Mathematics of Computation 28, 231–236 (1974)
16. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proceedings of the 4th Alvey Vision Conference, pp. 147–151 (1988)
17. Lowe, D.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60(2), 91 (2004)
18. Lindeberg, T.: Feature detection with automatic scale selection. International Journal of Computer Vision 30(2) (1998)
19. Mikolajczyk, K., Schmid, C.: Scale and affine invariant interest point detectors (PDF). International Journal of Computer Vision 60(1), 63–86 (2004)
20. Zhang, Z.: Iterative point matching for registration of free-form curves and surfaces, pp. 119–152 (1992)
21. Kuhn, H.W.: The Hungarian method for the assignment problem Export. Naval Research Logistics Quarterly 2(1-2), 83–97 (1955)
22. Serratosa, F., Cortés, X., Solé-Ribalta, A.: Interactive graph matching by means of imposing the pairwise costs. In: ICPR 2012, pp. 1298–1301 (2012)
23. `http://deim.urv.cat/~francesc.serratosa/databases/`
24. Caetano, T.S., Caelli, T., Schuurmans, D., Barone, D.A.C.: Graphical Models and Point Pattern Matching. IEEE Trans. on PAMI 28(10), 1646–1663 (2006)
25. Jesorsky, O., Kirchberg, K.J., Frischholz, R.W.: Robust Face Detection Using the Hausdorff Distance. In: Bigun, J., Smeraldi, F. (eds.) AVBPA 2001. LNCS, vol. 2091, pp. 90–95. Springer, Heidelberg (2001)
26. Cortés, X., Serratosa, F.: Active-Learning Query Strategies applied to select a Graph Node given a Graph Labelling. In: GbR 2013, Austria, Vienna, pp. 61–70 (2013)
27. Cortés, X., Serratosa, F., Solé, A.: Active Graph Matching based on Pairwise Probabilities between nodes. In: SSPR 2012, Hiroshima, Japan, pp. 98–106 (2012)
28. Serratosa, F., Alquézar, R., Amézquita, N.: A Probabilistic Integrated Object Recognition and Tracking Framework. Expert Systems With Applications 39, 7302–7318 (2012)
29. Serratosa, F., Cortés, X., Solé, A.: Component Retrieval based on a Database of Graphs for Hand-Written Electronic-Scheme Digitalisation, ESWA 40, pp. 2493–2502 (2013)