MOTION ESTIMATION USING TANGENT DISTANCE

Jonathan Fabrizio and Séverine Dubuisson

Laboratoire d'Informatique de Paris 6 (UMR CNRS 7606) - Université Pierre et Marie Curie-Paris6
104 avenue du Président Kennedy, F-75016 PARIS
{Jonathan.Fabrizio, Severine.Dubuisson}@lip6.fr

ABSTRACT

In this paper, we present a method based on tangent distance to estimate motion in image sequences. Tangent distance combines an intuitive understanding and effective modeling of differences between patterns. This tool was first introduced and successfully applied in character recognition. It allows to compare patterns according to small transformations (translations, rotations, *etc*.). We show, how to take advantages of some properties of tangent distances to perform a robust motion estimation algorithm. Particularly, the presented algorithm can easily be adapted and optimized to various types of movements and can also be used to estimate optical flow in image sequences. Moreover, and despite a time of computation a bit long, this algorithm can be massively paralleled.

***Index Terms*—** Motion analysis, Optical flow, Tangent distance

1. INTRODUCTION

The estimation of motion in a sequence of images is a basic task in computer vision. A lot of new techniques have evolved. Optical flow [1] gives the distribution of the apparent movement of the brightness patterns in an image. A class of techniques that received a special interest for its simplicity and computational efficiency is based on the idea that for most points in the image, neighboring points have approximately the same brightness. This class of techniques is known as the gradient-based techniques [2] and is composed of two sub-classes, namely, global and local. Frequency domain approaches [3, 4], consider the motion estimation is based on the phase changes in the frequency domain. Block-based methods [5] find the best block from a previous frame, generally used to reconstruct an area of the current frame: a vector characterizes the displacement of a block of pixels. The computational complexity of a block-based motion estimation technique can be determined by three main factors: the search algorithm, the cost function evaluation and the search range. A lot of motion estimation algorithms have been developed to reduce the complexity of motion estimation with full search. Some try to reduce the cost function [6]. Many others are based on predefined search patterns to reduce the total number of search points [7, 8].

In this paper, we investigate the use of tangent distance in a motion estimation framework. Tangent distance was first introduced by P. Simard et al. in the 90's [9] and has been proved to be especially effective in the domain of character recognition. Because of its insensitivity to small transformations and thus its capacity to derive classifiers independent to small variations, this tool has mainly been used in pattern recognition systems [10, 11, 12, 13]. We can also cite specific applications on face location and recognition [14] and speech recognition [15].

This paper is organized as follows. In Section 2, we give a brief explanation of tangent distance. Then, in Section 3, the general principle of our algorithm is described followed by a brief discussion concerning important parameters. Finally, in Section 4, some results are presented, and concluding remarks are given in Section 5.

2. TANGENT DISTANCE

As lie operators [16] are known to be effective in detecting very small degrees of object motions, the tangent distance allows to compare two patterns according to small transformations. The set of all transformed patterns has highly nonlinear characteristics, and to obtain a tractable representation, we consider a linear approximation of the transformation using a Taylor expansion. Then, given an image I and I_r this image after a small rotation, we can linearly approximate I_r by:

$$I_r = I + \lambda_r \vec{V_r} \tag{1}$$

with $\vec{V_r}$ the tangent vector of the rotation and λ_r its contribution.

We can locally model the set of images of a point (seen as parametric curve) generated by each basic transformation by adding the tangent vector to the curve, weighted by a factor, at this point. This tangent vector derives, in this point, the set of pattern generated by this transformation. The tangent distance between two points P and P' is equal to the euclidean distance between the linearization of the parametric curves

Thanks to Pr. Milgram from UPMC-LISIF for the help he provides to us during this work.

through P and P' (see figure 2).
If we now consider a composition of transformations (rotations, translations, *etc.*), an approximation of I_m, transformation of I, is given by:

$$I_m = I + \lambda_1 \vec{V}_1 + ... + \lambda_i \vec{V}_i \quad (2)$$

where $\vec{V}_i$ is the vector tangent to the i^{th} transformation (figure 1) and λ_i represents its contribution.
For a given couple of images (I and J), we do not compare directly these images, but their approximations I_m and J_m, given by equations:

$$I_m = I + \lambda_1 \vec{V}_1 + ... + \lambda_i \vec{V}_i \quad (3)$$
$$J_m = J + \lambda'_1 \vec{W}_1 + ... + \lambda'_i \vec{W}_i \quad (4)$$

The comparison between the images is then performed by the minimization of the tangent distance, given by:

$$\min(d(I + \lambda_1 \vec{V}_1 + ... + \lambda_i \vec{V}_i, J + \lambda'_1 \vec{W}_1 + ... + \lambda'_i \vec{W}_i)) \quad (5)$$

where $d(i, j)$ is the euclidean distance between i and j.
In this minimization scheme, I and J are known, all $\vec{V}_i$ and $\vec{W}_i$ are tangent vectors and can be computed numerically or analytically [17]. On the other hand, all λ_i are unknown: the result of the minimization gives their optimal values. This means that the minimization not only gives a robust comparison between two patterns but also computes the proportion of every transformations applied to make them fit.

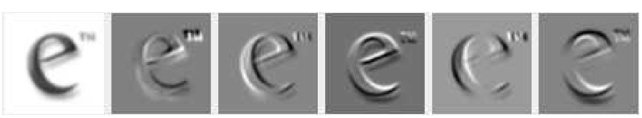

Fig. 1. Illustration of tangent vectors. From left to right, the original image, the tangent vector of a rotation, an horizontal translation, a vertical translation, an horizontal scale and a vertical scale.

3. GENERAL PRINCIPLE

3.1. Algorithm

The integration of tangent distances in a motion estimation framework is quite easy: we take a block B of pixels around the considered pixel in the current frame and compare this block to some blocks near the position of B in the next image. The comparison between blocks is performed using tangent distance and its minimization determines the block B' that best matches with B. The solution of the minimization also gives the values of λ_i (equation 5) indicating contributions of each transformation (translations, rotations, *etc.*). Then, the difference between positions of B and B', gives a first approximation of the global transformation (t_1 in figure 3). This approximation is improved by considering the values of

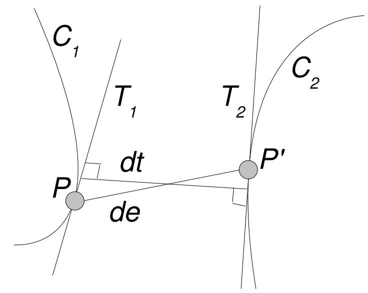

Fig. 2. Comparison between P and P': $[de]$ is the euclidean distance. Each 3D curve C_1 and C_2 represents possible transformations respectively for P and P'. T_1 and T_2 are the linearization of both curves C_1 and C_2 in P and P'. The tangent distance is the distance between T_1 and T_2 ($[dt]$). Nearest points of T_1 and T_2 correspond to the approximations of transformations of P and P'.

λ_i which give proportions of the other transformations (t_2 in figure 3). By the composition of the two transformations, we can determine the exact displacement (t in figure 3) of the considered pixel.
This process allows to test only few positions around the initial block without decreasing the search range. In our tests we never exceed 11 positions.
Transformations that have to be included in the computation depend on the possible movements in the sequence. If the movement is in the image plan (*i.e.* 2D), only translations and rotations are required. If the movement corresponds to the projection of a 3D motion, it is necessary to include scale factors. Moreover, in the case of deformable objects, this approach allows to add various transformations to ensure the robustness of the estimation. Then, because of the possibility given by tangent distance to integrate different kind of transformation models, the presented algorithm can be adapted to various situations.

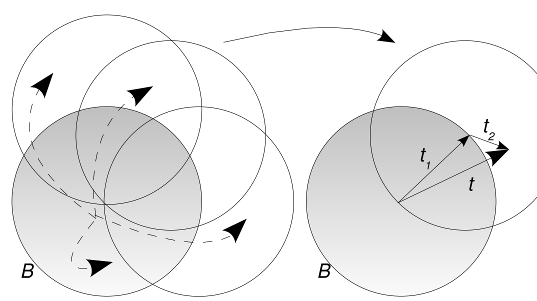

Fig. 3. The initial block B is compared to other blocks. Using the tangent distance minimization scheme, the block B' is selected and the transformation is computed. This transformation is given by the composition of a translation t_1 between B and B' and the transformation t_2 that makes B and B' exactly match.

3.2. Parameter Influence

Two main parameters must be considered to get optimal results. The first parameter concerns the preprocessing step and the second is the geometry of the block.
Computing tangent vectors necessitates a derivation of patterns. As the image is sampled, it must be smoothed to be derivable. We have smoothed the image using a Gaussian filter. The choice of the standard deviation σ of the Gaussian is very important : if σ is too small, the linear approximation (equation 2) becomes false and the movement is not detected (figure 4 on left). If σ is too big, the movement is extended over a too large area (figure 4 on right).
The geometry of the block is also determinant. If the block is too small, the movement is not detected. On the contrary, if the block is to big, the movement is spread out.
The Gaussian, the size of blocks and the maximum distance between the original block and all other blocks need to be adjusted to the magnitude of the movement.
These parameters have thus an influence on the quality of the results, but also an impact on the speed of the algorithm: a too large Gaussian and, moreover, a too big block, will too much increase executing time.

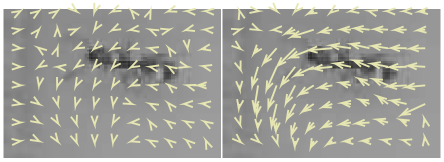

Fig. 4. Influence of the smoothing on the motion estimation results. On left the image was not smoothed, no motion is detected. On right the image was too much smoothed (Gaussian with $\sigma = 6$), the movement is spread out.

4. RESULTS

Our method has been tested on different image sequences, and we present the results obtained on three of them. The first one is a sequence of moving ants, the second one a video monitoring sequence. Results on both sequences can be seen as optical flows. The last sequence presents a moving and deformable stuffed toy.
The first sequence, called "Ant" sequence, shows ants moving on a flat surface. Each ant has a complex movement: translation while walking and deformations while moving its head, abdomen, or antennas. Despite the erratic and highly nonlinear displacements of ants, motions are well estimated by our approach (see figure 5). Notice that, due to the size of the blocks, the movement is a bit spread out around ants.
On the second sequence, one man is static, the other one is moving. Figure 6 shows that the moving man is well detected. The direction of his body is well estimated as well as the movement of his legs. Best is to notice that our algorithm not only detects the walking man but also its reflect on the window on right (not easy to see) and the reflection of another person in the center.
On the last sequence, the little stuffed toy is moving and deforming in space. To check if the movement is well estimated we have reconstructed a frame at t using the previous frame and the motion vectors estimated between frames t and $t-1$ using our approach. The real image is shown in figure 7.(a) and its reconstruction in figure 7.(b). Only the area in the rectangle delimited by the two red marks has been reconstructed. We can see that the reconstruction is quite good, despite an artifact in front of the head. Another interesting point is that the reconstruction on the boundaries between the reconstruction area and the real frame is quite good. The last and important point is that the leg, which has a non rigid motion, is also well reconstructed.
All these results show our motion estimation approach is very robust ("Ant" sequence), able to detect imperceptible motions ("Window" sequence) and can deal with non rigid displacements ("stuffed toy" sequence).

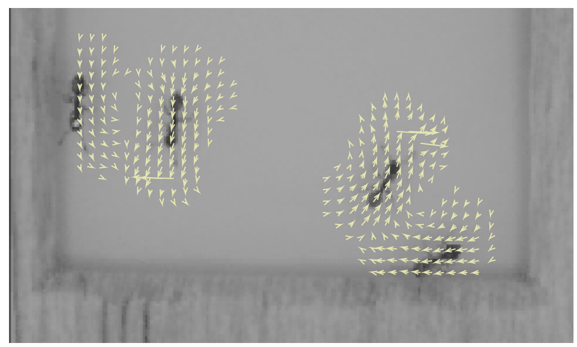

Fig. 5. Results on "Ant" sequence: all ant motions are well detected.

Fig. 6. Results on "Window" sequence. Walking men motions are well detected as well as their reflection (on right and center).

5. CONCLUSION

We have proposed an algorithm to estimate motion in images sequences using tangent distances. This method can also be

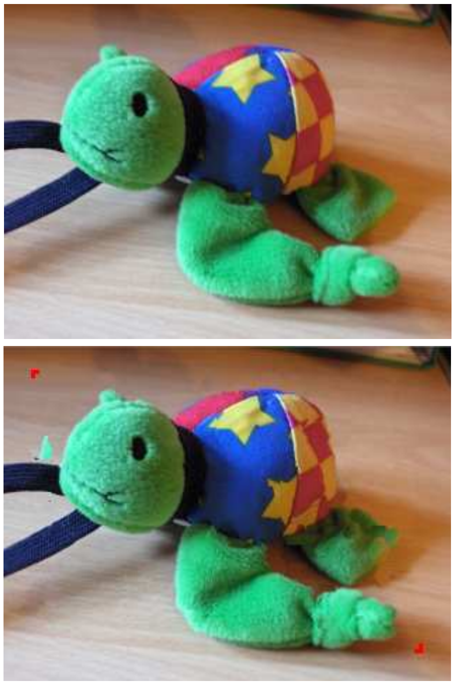
(a)
(b)

Fig. 7. Results on "Stuffed toy" sequence. (a) The original frame at t. (b) The reconstructed frame (into the rectangle marked by the two red marks) using (a) and the motions vectors estimated between frames $t-1$ and t. Results are good and the transition between the reconstruction and the real image is not visible.

used to estimate optical flow in images sequences. Our algorithm takes advantage of tangent distance: many distortions can be taken into account to make the motion estimation robust and adapted to the specific context of a sequence, as have shown our experimental tests. For example, we are currently integrating an illumination model to be more robust against high lightening variations. Another interesting point is that the algorithm can be easily paralleled: at low level, all computations are quite simple operators (linear algebra), and at high level, we can estimate the motion of pixels independently. We are implementing this algorithm on $FPGA$ and plan to improve it to make it work in real time.

6. REFERENCES

[1] B.K.P. Horn and B.G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.

[2] M. Elad, P. Teo, and Y. Hel-Or, "Optimal filters for gradient-based motion estimation," *IEEE international conference on computer vision*, pp. 559–565, 1999.

[3] R.W. Young and N.G. Kingsbury, "Frequency-domain motion estimation using a complex lapped transform," *IEEE Transactions on image processing*, vol. 2, no. 1, pp. 2–17, 1993.

[4] E. Bruno and D. Pellerin, "Robust motion estimation using spatial gabor-like filters," *Signal Processing*, vol. 82, pp. 297–309, 2002.

[5] Y. Huang, C. Chen, C. Tsai, C. Shen, and L. Chen, "Survey on block matching motion estimation algorithms and architectures with new results," *J. of VLSI Signal Processing*, vol. 42, no. 3, pp. 297–320, 2006.

[6] X. Lee and Y.Q. Zhang, "A fast hierarchical motion compensation scheme for video coding using block feature matching," *IEEE Transactions on circuit and systems for video technology*, vol. 6, no. 6, pp. 627–635, 1996.

[7] A.C.K. Ng and B. Zeng, "On macroblock partition for motion compensation," *IEEE International Conference on Image Processing*, pp. 1697–1700, 8-11 Oct 2006.

[8] S. Zhu and K.K. Ma, "A new diamond search algorithm for fast block-matching motion estimation," *IEEE Trans. on Image Processing*, vol. 9, 2000.

[9] P. Simard, Y. LeCun, J. Denker, and B. Victorri, "An efficient algorithm for learning invariances in adaptive classifiers," in *Proc. of International Conference on Pattern Recognition, The Hague*. 1992, pp. 651–655, IAPR.

[10] P. Simard, B. Victorri, Y. LeCun, and J. Denker, "Tangent prop: a formalism for specifying selected invariances in adaptive networks," in *Advances in Neural Information Processing Systems 4 (NIPS*91)*, Denver, CO, 1992.

[11] P. Simard, Y. LeCun, J. Denker, and B. Victorri, "Transformation invariance in pattern recognition, tangent distance and tangent propagation," in *Neural Networks: Tricks of the trade*, 1998.

[12] H. Schwenk and M. Milgram, "Constraint tangent distance for on-line character recognition," in *International Conference on Pattern Recognition*, 1996, pp. 515–519.

[13] Holger Schwenk, *Amlioration de classifieurs neuronaux par incorporation de connaissances explicites : Applications la reconnaissance de caractres manuscrits*, Ph.D. thesis, Université Pierre et Marie Curie - Paris VI, 1996.

[14] Robert Mariani, "A face location and recognition system based on tangent distance," in *Multimodal interface for human-machine communication*, River Edge, NJ, USA, 2002, pp. 3–31, World Scientific Publishing Co., Inc.

[15] Wolfgang Macherey, Daniel Keysers, Jörg Dahmen, and Hermann Ney, "Improving automatic speech recognition using tangent distance," in *Eurospeech 2001, 7th European Conference on Speech Communication and Technology*, Aalborg, Denmark, 2001, vol. III, pp. 1825–1828.

[16] W.D. Pan, S.-M. Yoo, and C.-H. Park, "Complexity accuracy tradeoffs of lie operators in motion estimation," *Pattern Recognition Letters*, vol. 28, no. 7, pp. 778–787, 2007.

[17] P. Simard, Y. LeCun, and Denker J., "Efficient pattern recognition using a new transformation distance," in *Advances in Neural Information Processing Systems*, 1993, vol. 5.