

Optimal Detection and Tracking of Feature Points using Mutual Information

Amaury Dame, Eric Marchand

Abstract—This paper proposes a new way to achieve feature point tracking using the entropy of the image. Sum of Squared Differences (SSD) is widely considered in differential trackers such as the KLT. Here, we consider another metric called Mutual Information (MI), which is far less sensitive to changes in the lighting condition and to a wide class of non-linear image transformation. Since mutual-information is used as an energy function to be maximized to track each points, a new feature selection, which is optimal for this metric, is proposed. Results under various complex conditions are presented. Comparison with the classical KLT tracker are proposed.

I. INTRODUCTION

Interest point tracking is a considerable domain of research. Such trackers are used in many application domains such as object tracking, navigation, augmented reality, motion estimation, etc. These works especially concern image sequences where displacements between one frame to another are small. Indeed, if the motion is small enough then windows including each interest point can be tracked by an energy minimization. The earliest studies on this subject are those performed by Lucas and Kanade [6] that use images difference as energy. Depending on the information of the window, each point cannot be tracked with the same efficiency. Different detectors have been then proposed to increase efficiency of the tracker. Many corner detectors, such as [5] [12], are examples of the latter.

The classical LK approach is based on a relation that links the luminance of a physical point at time t to its luminance at a later time $t + dt$ due to a relative motion of the observer with respect to that scene or to other events like illumination changes. Due to the complexity of this relation, aforementioned algorithms are most often based on the temporal luminance constancy hypothesis. However, it is well known that this constraint can be easily violated. Consequently, many authors have addressed this issue. The basic assumption of the luminance constancy, can be extended by the spatial gradient constancy assumption as proposed in [15]. Dealing with target tracking, under the assumption of a Lambertian scene, illumination variations can be taken into account [4]. This Lambertian hypothesis may be removed when modeling the illumination changes as a surface that evolves over time [13]. This latter work can be seen as related to [2] where the intensity variation between dt is expressed as a mixture of causes. A general framework has been proposed in [9] where local and constant modifications of the image intensity are estimated numerically.

Amaury Dame is with CNRS, IRISA, Rennes, France. Eric Marchand is with INRIA, Centre Rennes - Bretagne Atlantique, Rennes, France. This work is supported by DGA under contribution to student grant. `firstname.name@irisa.fr`.

Mutual information, especially studied in medical image registration, tolerates such changes. Therefore it is another metric well adapted and more robust for the point tracking problem. Since its derivative form have been recently studied in many works [14] [8] [3], its application on interest point tracking is considerable.

In this paper we present a tracker based on mutual information with its corresponding adapted windows. The remainder of this paper is organized as follows. In section II, an overview of the tracking method is tackled followed by a description of the proposed tracker. Section IV presents a description of the selection approach for points of interest. Finally, results are given in section V comparing our method and Lucas Kanade algorithm.

II. CLASSICAL SSD TRACKER

Using an energy function, the goal is to select and track points from images I_0 to I_t , i.e. track each corresponding template (small window around a point). Let us consider only one of these points and name the corresponding template T . Classically the displacement \mathbf{p} between T_0 and T_t is computed using SSD:

$$\hat{\mathbf{p}} = \arg \min_{\mathbf{p}} (SSD(T_0, w(T_t, \mathbf{p}))) \quad (1)$$

$$= \arg \min_{\mathbf{p}} \sum_{\mathbf{x} \in T_0} [T_t(w(\mathbf{x}, \mathbf{p})) - T_0(\mathbf{x})]^2 \quad (2)$$

where $w : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is the warp function corresponding to the displacement \mathbf{p} applied to \mathbf{x} . Here SSD will be replaced by mutual information.

III. MUTUAL INFORMATION-BASED TRACKER

A. Mutual information

Rather than comparing intensities, mutual information is the quantity of information shared between two random variables. Mutual information of two random variables X and Y is given by the following equation [11]:

$$MI(X, Y) = H(X) + H(Y) - H(X, Y). \quad (3)$$

$H(X)$ is the entropy of the signal X and $H(X, Y)$ is the joint entropy of the signal X and Y . If mutual information is maximized, then the two signals are aligned. The advantage of this function compared to SSD is that no linear relation is needed between the two signals [16].

B. Tracker

Considering mutual information on images, the problem to solve, previously described in equation (1) becomes:

$$\hat{\mathbf{p}} = \arg \max_{\mathbf{p}} (MI(T_0, w(T_t, \mathbf{p}))) \quad (4)$$

Let us note $w(T_t, \mathbf{p})$ as $T_t(\mathbf{p})$. Mutual information can also be defined by [11]:

$$MI(T_0, T_t(\mathbf{p})) = \sum_{r,t} p_{rt}(r, t, \mathbf{p}) \log \left(\frac{p_{rt}(r, t, \mathbf{p})}{p_r(r)p_t(t, \mathbf{p})} \right) \quad (5)$$

The joint probability $p_{rt}(r, t, \mathbf{p})$ is the probability that the couple $(T_0(\mathbf{x}), T_t(w(\mathbf{x}, \mathbf{p})))$ has a value of (r, t) on the images. It is classically obtained using the joint histogram $h(r, t)$ between T_0 and T_t :

$$\begin{aligned} p_{rt}(r, t, \mathbf{p}) &= \frac{1}{N_{\mathbf{x}}} \sum_{\mathbf{x}} h(r, t, \mathbf{p}) \\ &= \frac{1}{N_{\mathbf{x}}} \sum_{\mathbf{x}} \phi[r - T_0(\mathbf{x})] \phi[t - T_t(w(\mathbf{x}, \mathbf{p}))]. \end{aligned} \quad (6)$$

In the usual histogram computation, ϕ is the Kronecker function. This definition is inappropriate for our problem since T_t is warped, resulting in a no more discrete $T_t(\mathbf{x}, \mathbf{p})$. To overcome this difficulty, Collignon [7] introduced partial volume interpolation using ϕ as a simple first order B-spline: $\phi(x) = B_1(x)$. This means that, for each couple $(r, t) \in \mathbb{R}^2$, the four histogram entries $h(v_r, v_t)$, corresponding to the four neighbouring integer couples (v_r, v_t) , will be increased.

Typically, images T_0 and T_t have 256 possible values for each pixels. Under this form, mutual information is then computed on a 256×256 joint histogram. This definition supposes a very sharp maximum which is interesting for accuracy but reduces the domain of convergence. As an example, mutual information has been computed on a corner pixel using the previous definition with respect to (wrt) horizontal and vertical translation. As Fig. 1(a) shows, the domain of convergence is smaller than 1 pixel.

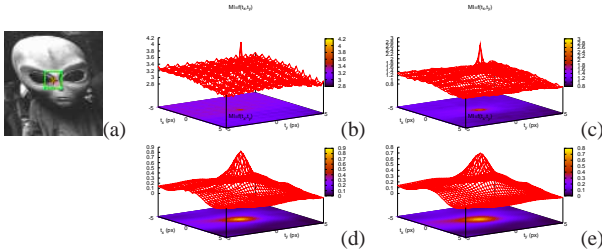


Fig. 1. Effects of the histogram's bin size on mutual information wrt translations. (a) Selected 15×15 window, and mutual information with (b) $N_c = 256$, (c) $N_c = 64$, (d) $N_c = 8$, (e) $N_c = 8$ and $\phi(x) = B_3(x)$.

The solution of this problem is to reduce the number of histogram's bins. Let N_c be the new number of bins. Equation (6) has to be changed to fit $(T_0, T_t) \in [0, 255]^2$ in the histogram of size $[0, N_c]^2$. A simple division is applied on both images to obtain $(\bar{T}_0, \bar{T}_t) \in [0, N_c]^2$.

$$\bar{T}_0(\mathbf{x}) = T_0(\mathbf{x}) \frac{N_c}{256} \quad \bar{T}_t(\mathbf{x}) = T_t(\mathbf{x}) \frac{N_c}{256}. \quad (7)$$

Figure 1 shows the effects of the bin size on the mutual information computation wrt translation. It is clear that the convergence of the tracker will increase using a smaller N_c as long as it is not too small to keep enough information. Even with a small N_c , mutual information, computed using partial volume interpolation, is still affected by artefacts. The

use of higher B-spline orders smoothes results and reduce interpolation effects [10] as depicts Fig. 1(e).

A non linear optimisation is used to find the maximum of the cost function using a Taylor expansion of the mutual information. Assuming that the shape of the maximum is parabolic, the optimisation is performed using a Levenberg-Marquardt like approach. The displacement parameter \mathbf{p}_t from the previous frame is iteratively updated by the following equation:

$$\mathbf{p}_{t+1} = \mathbf{p}_t - \lambda(\mathbf{H} + \mu \text{diag} \mathbf{H})^{-1} \mathbf{G}^\top \quad (8)$$

This algorithm is applied until \mathbf{p} reaches the convergence i.e., the displacement parameter corresponding to the current frame. Gradient and Hessian are given by [3]:

$$\begin{aligned} \mathbf{G} &= \frac{\partial MI(T_0, w(T_t, \mathbf{p}))}{\partial \mathbf{p}} \\ &= \sum_{r,t} \frac{\partial p_{rt}}{\partial \mathbf{p}} \left(1 + \log \left(\frac{p_{rt}}{p_r} \right) \right) \end{aligned} \quad (9)$$

$$\mathbf{H} = \frac{\partial^2 MI(T_0, w(T_t, \mathbf{p}))}{\partial \mathbf{p}^2} \quad (10)$$

$$= \sum_{r,t} \frac{\partial p_{rt}}{\partial \mathbf{p}}^\top \frac{\partial p_{rt}}{\partial \mathbf{p}} \left(\frac{1}{p_{rt}} - \frac{1}{p_r} \right). \quad (11)$$

The derivative of the joint probability is then required. Using the previous definition in (6) and passing derivative through the sum yields to:

$$\frac{\partial p_{rt}}{\partial \mathbf{p}}(r, t) = \frac{1}{N_{\mathbf{x}}} \sum_{\mathbf{x}} \frac{\partial \phi}{\partial \mathbf{p}}(r - T_0(\mathbf{x}) \phi(t - T_t(w(\mathbf{x}, \mathbf{p}))) \quad (12)$$

The partial derivate of ϕ is simply calculated by derivation decomposition: we know the warp function and its derivative function, as well as the image gradients, and finally the derivative B-spline is easily obtained by the following B-splines property:

$$\frac{\partial B_n(x)}{\partial x} = B_{n-1} \left(x + \frac{1}{2} \right) - B_{n-1} \left(x - \frac{1}{2} \right) \quad (13)$$

The warp function considered in this work is so that \mathbf{p} is a 4 dimension parameter taking into account a rotation of angle α , a scale λ and a translation \mathbf{T} :

$$w(\mathbf{x}, \mathbf{p}) = \lambda \mathbf{R}(\alpha) \mathbf{x} + \mathbf{T} \quad (14)$$

IV. FEATURES SELECTION

Since the following tracking method is not based on SSD, features optimal for other differential trackers (Harris points as reported in [12]) are not optimal for the presented tracker. Nevertheless, the method considered here to derive optimal features adapted to our feature is similar to the approach presented in [12]: the best features are the ones that, by construction, are "good" for our tracker. This goodness property is characterized by a well conditioned Hessian matrix \mathbf{H} . Let us consider that the observed transformation is a translation so that \mathbf{H} is 2×2 matrix. \mathbf{H} is well conditioned if both of its eigenvalues are large ie both are above a threshold λ .

\mathbf{H} is computed slightly the same way as defined in equation (9) using a translation as the warp function. The

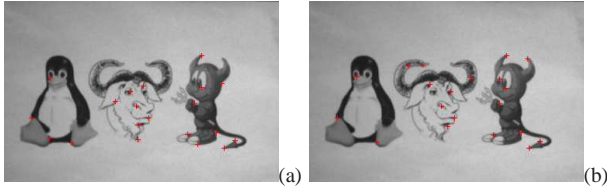


Fig. 2. Interest point selection. 30 best features using the (a) proposed selection and (b) Harris points selection with 17×17 windows.

difference, compared to the equations used in tracking, is that the computation of the joint probability and its derivative is preformed using a circular Gaussian centered on the interest point as a weighting function π . The joint probability is then computed as follows:

$$p_{rt}(r, t, \mathbf{p}) = \frac{1}{N_{\mathbf{x}}} \sum_{\mathbf{x}} \pi(\mathbf{x}) h(r, t, \mathbf{p}) \quad (15)$$

$$\pi(\mathbf{x}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\mathbf{x}-\mathbf{x}_c)^{\top}(\mathbf{x}-\mathbf{x}_c)}{2\sigma^2}}$$

where \mathbf{x}_c is the coordinates vector of the center of the window and σ^2 is the variance of the selected gaussian. Eigenvalues are simply obtained using a singular value decomposition (SVD). To determine the threshold λ , measures on noised uniform regions are done. As it is shown in Fig. 2, some of the extracted features are near the features extracted by a Harris detector. Indeed, regions of interest remains regions of the image that contain information allowing to distinguish horizontal and vertical motion. But this selection is optimal for a mutual Information tracker as experiments will show.

V. EXPERIMENTAL RESULTS

In this section we will firstly compare the tracker using Harris features and the optimal proposed features to finally compare results between the KLT algorithm and the Mutual Information tracker each one with its own optimal features. To estimate the accuracy of the trackers, a planar object is tracked, so that the transformation from one frame to the next one is an homography. When estimating the homography from teached points using robust techniques (M-estimation) the number of outliers is a good way to measure the efficiency of the tracking process.

A. Selection results

The goal of the first experiment is to show benefits of the optimal feature selection. To compare the effects of the selection of the initial features points, the proposed tracker is used on a sequence (see Fig. 3) using both selection processes. As shown in Fig. 4, the number of outliers using the proposed method is evolving slower than the one with Harris corner detector. The estimated homography computed from the initial to the current frame using the tracked points is then more accurate (See Fig. 3).

B. Window tracking

To validate the efficiency of our tracker, we compare its efficiency with the KLT tracker. A first experiment presents the behaviour of this two trackers toward an exemple of

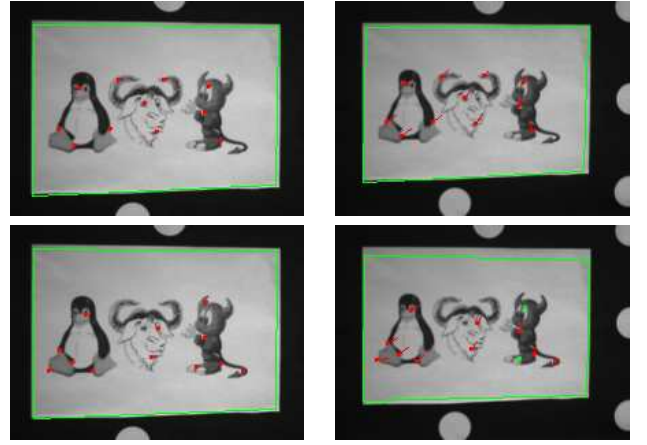


Fig. 3. Effects of the selection on the proposed tracker (frame 1 and 50). The first row is the proposed selection and second row is Harris selection. The green rectangle in the final image represents the rectangle from the first image transformed using the estimated homography.

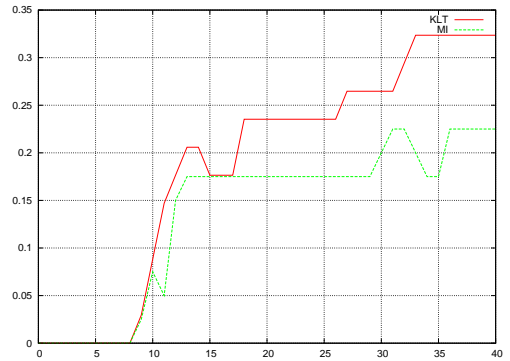


Fig. 4. Interest point selection. Evolution of the ratio of outliers on the same sequence using the proposed method and using Harris points.

point. The interest of mutual information in this kind of problem is that it is able to deal with illumination changes. To validate this robustness, we considere a corner point detected from Harris detector and add illumination changes to compare both energy functions. As Fig. 5 depicts, SSD is very sensitive to illumination changes, no convexity is visible on the cost function. However, mutual information is well adapted for this kind of changes. Using third order B-splines, the cost function and its gradients (computed using equation (9)) do not show any artefacts (See Fig. 5 and 6).

This exemple is very representative of the efficiency of the mutual information tracker compared to KLT toward illumination changes.

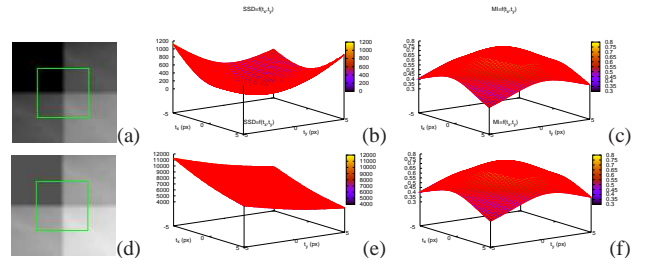


Fig. 5. Cost functions wrt translation without (first row) and with (second and third row) illumination changes. First column: templates, second column: sum of squared differences and third column: mutual information.

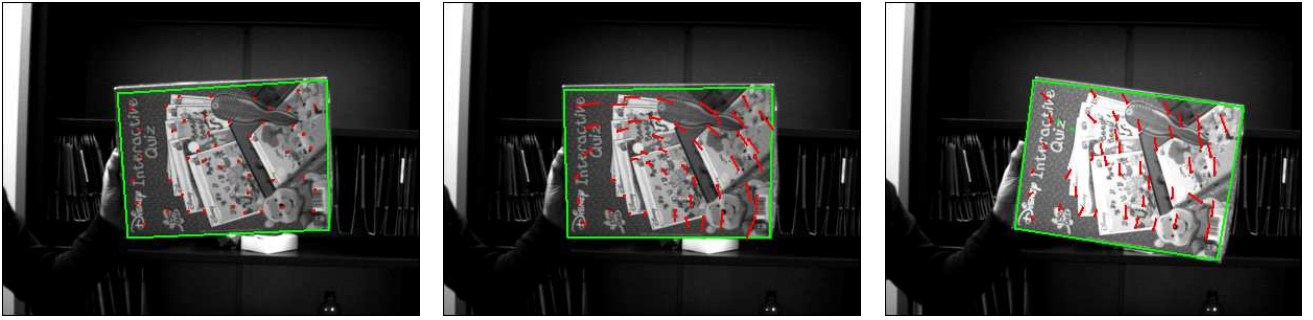


Fig. 7. Tracking using mutual information on the same image sequence, frame 1, 30 and 70. Red dots are the tracked points, red lines represent current velocity, green dots are outliers and the green rectangle represents the rectangle from the first image transformed using the estimated homography.

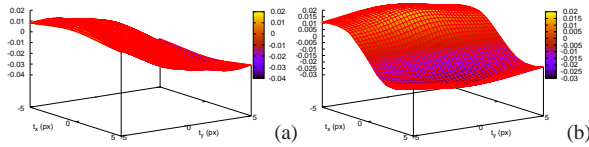


Fig. 6. Computed gradient wrt translations with illumination changes using the same templates as in Fig. 5. (a) Gradient wrt horizontal translation t_x , (b) gradient wrt vertical translation t_y

C. Tracker on image sequences

The last experiment applies both tracking methods on a sequence which contains illumination changes, specularity and saturation. The present KLT tracker uses the implementation of Stan Birchfield [1] (with and without illumination insensitivity). Figure 8 shows the number of outliers with respect to the sequence's frames. During all the sequence, indoor artificial illumination conditions cause oscillation in global illumination intensity. This variations are sufficient to cause the failure of the original KLT tracker while MI and illumination insensitive KLT are still efficient.

From the 50th to the last frame, the sequence is then affected by specularity and then by saturation. Figure 7 shows some frames of the sequence. Mutual information tracker is the first method to have outliers. Indeed, convergence domain of mutual information is smaller than the one of SSD in nominal conditions. Then, when a displacement is slightly large without illumination changes, the proposed tracker diverges whereas KLT keeps converging. However, when the sequences reaches illumination variations, the proposed tracker is more robust and the number of outliers of the KLT is quickly higher than the one of the proposed tracker.

VI. CONCLUSION

In this paper we presented a new feature point tracking method where features are selected optimally according to the specification of the tracker that uses mutual information, instead of generally used SSD. MI allows to cope with illumination changes and other non-linear image transformations. Since the problem statement is similar to the KLT and that different features are used, an hybrid tracker can be implemented to take advantages of both solutions.

REFERENCES

[1] Stan Birchfield. Derivation of kanade-lucas-tomasi tracking equation. <http://www.ces.clemson.edu/~stb/klt/>, January 1997.
 [2] M.J. Black, D.J. Fleet, and Y. Yacoob. Robustly estimating changes in image appearance. *Computer Vision and Image Understanding*, 78:8–31, 2000.

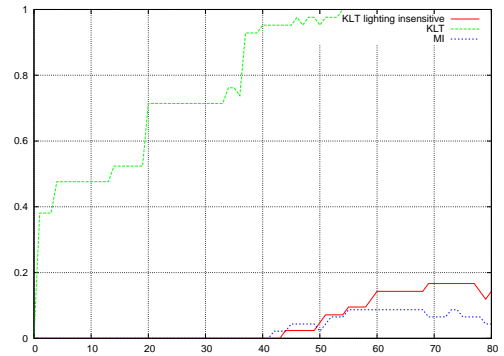


Fig. 8. Ratio of the number of outliers with the total number of points wrt frames.

[3] N.D.H. Dowson and R. Bowden. A unifying framework for mutual information methods for use in non-linear optimisation. In *European Conf. on Computer Vision*, pages 365–378, 2006.
 [4] G. Hager and P. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(10):1025–1039, October 1998.
 [5] L. Kitchen and A. Rosenfeld. Gray level corner detection. *Pattern Recognition Lett.*, 1, pages 95–102, 1982.
 [6] B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Int. Joint Conf. on Artificial Intelligence, IJCAI'81*, pages 674–679, 1981.
 [7] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *Medical Imaging, IEEE Transactions on*, 16(2):187–198, 1997.
 [8] F. Maes, D. Vandermeulen, and P. Suetens. Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information. *Medical Image Analysis*, 3(4):373–386, 1999.
 [9] S. Negahdaripour. Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(9):961–979, 1998.
 [10] J.P.W. Pluim, J.B.A. Maintz, and M.A. Viergever. Mutual information matching and interpolation artefacts. In K.M. Hanson, editor, *SPIE Medical Imaging*, volume 3661, pages 56–65. SPIE Press, 1999.
 [11] C. E. Shannon. A mathematical theory of communication. *Bell system technical journal*, 27, 1948.
 [12] J. Shi and C. Tomasi. Good features to track. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR'94*, pages 593–600, Seattle, Washington, June 1994.
 [13] G. Silveira and E. Malis. Real-time visual tracking under arbitrary illumination changes. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR'07*, Minneapolis, USA, June 2007.
 [14] P. Thévenaz and M. Unser. Optimization of Mutual Information for Multiresolution Image Registration. *IEEE Transactions on Image Processing*, 9(12):2083–2099, 2000.
 [15] M. Tistarelli. Multiple constraints for optical flow. In *European Conf. on Computer Vision, ECCV'94*, volume 1, pages 61–70, Stockholm, Sweden, May 1994.
 [16] Paul Viola and William M. Wells III. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997.