

EXPLOITING LOCAL AUTO-CORRELATION FUNCTION FOR FAST VIDEO TO REFERENCE IMAGE ALIGNMENT

Arif Mahmood, Sohaib Khan

{arifm, sohaib}@lums.edu.pk

CV Lab, Department of Computer Science,
Lahore University of Management Sciences, Pakistan

ABSTRACT

Digital images of natural scenes are usually characterized by strong spatial correlation between adjacent pixels which has been successfully exploited in the coding of still and moving pictures. In this work we show that the strong spatial correlation of natural images can also be used to speedup the video to reference image alignment algorithms. To this end, we divide the search locations in the reference image into groups. The target frames are matched with only one location in each group, while on the remaining locations we evaluate exact theoretic upper bounds on the correlation coefficient. These bounds are used to eliminate majority of the search locations and thus result in significant speedup without effecting the value or location of the global maxima. In our experiments, up to 83.3% search locations are found to be eliminated and the speedup is up to 5.3 times the FFT based implementation and up to 7.9 times the spatial domain techniques.

Index Terms— Fast Algorithm, Video to Reference Alignment, Bounds on Correlation Coefficient

1. INTRODUCTION

Digital images of natural scenes are usually characterized by strong spatial correlation between adjacent pixels. This spatial correlation has been successfully exploited by the still picture coding algorithms like JPEG and by the moving pictures codecs like MPEG and H.264/AVC. In this paper we show that the strong spatial correlation can also be exploited to speedup the video to reference image alignment algorithms. This results in an efficient, close to real time implementation of a number of applications in image processing that use video to reference image alignment as a key component. Such applications include object detection, model based landmark extraction, automatic camera tracking and video geo-registration over large search space [1].

Although video to reference image alignment involves pattern matching, it is inherently different from the block matching for motion compensation [2]. This is because the block matching algorithms are applied to video frames that appear temporally close to each other and are acquired by the

same sensor, hence the level of dissimilarity is low. On the other hand, video to reference image alignment may require pattern matching between two images acquired at different times, under different illumination conditions, and by using sensors with different spectral responses, resulting in high dissimilarity. The inherent differences between block matching in video compression and the video to reference image alignment applications, render the standard matching techniques [3], [4] used in the former to be inaccurate for the latter. In contrast, the correlation coefficient, which is usually criticized for its high complexity for block matching applications [5], turns out to be a more accurate and robust similarity measure for most of the video to reference image alignment applications [6], [7]. While a number of schemes have been investigated to reduce the time complexity of the correlation, elimination of potential matching locations based upon correlation bounds has not been well explored before us. We have proposed an L_1 norm based upper bound on correlation coefficient in [8] for fast block matching algorithms. This bound remains useful for blocks of very small sizes, for example 4×4 and 8×8 , while for the larger sized target frames it approaches Cauchy Schwartz inequality and cannot be used for elimination of potential search locations. In [9] we have proposed another bound on correlation coefficient which is based upon normalized angular distance and an elimination algorithm using that bound for fast video to reference image alignment as well. The Transitive Elimination Algorithm (TEA) [9] exploited inter-frame correlation to obtain speedup and therefore had a limited scope because strong inter-correlation over multiple frames is not always guaranteed. In the current paper, we propose a third bound on correlation coefficient which is based upon normalized Euclidean distance and we theoretically compare this bound with the bound based upon angular distance. We find that the bound based upon angular distance is tighter than the bound based upon Euclidean distance. Moreover, we also propose a new algorithm for video to reference image alignment which has significantly wider scope than the Transitive Elimination Algorithm, because the currently proposed algorithm does not require the video frames to be inter-correlated.

The currently proposed algorithm for fast video to reference image alignment exploits the local auto-correlation function of the reference image to establish bounds on the correlation coefficient, and then uses these bounds to eliminate unsuitable search locations. In the reference image, the spatially consecutive candidate search locations are grouped into Group Of Locations (GOL). The target frame is matched only with the central location in each GOL, while for the remaining locations upper bound on correlation is evaluated. If at a specific location, the upper bound is found to be smaller than the yet known maxima, that location is discarded because it can never become the best match location. If at a location, the elimination condition is not satisfied then actual correlation will be evaluated. If a large number of search locations are eliminated from the search space by bound comparison, the over all video to reference image alignment process becomes very fast. For the execution time comparison, the proposed algorithm is compared with FFT based frequency domain implementation, Bounded Partial Correlation (BPC) algorithm [10], fast spatial domain implementation [11] and the TEA [9], on the same dataset as used by [9]. In our experiments, the currently proposed algorithm is found to be 5.32 times faster than the FFT based implementation, 2.88 times faster than BPC and 7.90 times faster than fast spatial domain implementation and 1.13 times faster than TEA. Although there is some improvement in elimination performance and execution time over TEA, the main advantage of the currently proposed algorithm is its broader scope.

2. PROBLEM DEFINITION

We consider a digital image as a two dimensional sequence of pixels indexed at discrete space (i, j) . A target frame f of size $m \times n$ pixels is to be matched at all valid search locations in the reference image R of size $s \times t$ pixels. For the purpose of matching, the reference image R is considered to be divided into overlapping locations l_{i_o, j_o} , each of size $m \times n$ pixels, where (i_o, j_o) are the coordinates of the first pixel of the reference location. The similarity measure used to match the frame f with search location l_{i_o, j_o} is the correlation coefficient defined as:

$$\rho_{i_o, j_o}^f = \frac{\sum_{i=1}^m \sum_{j=1}^n (f(i, j) - \bar{f})(l_{i_o, j_o}(i, j) - \bar{l}_{i_o, j_o})}{\sqrt{\sum_{i=1}^m \sum_{j=1}^n (f(i, j) - \bar{f})^2} \sqrt{\sum_{i=1}^m \sum_{j=1}^n (l_{i_o, j_o}(i, j) - \bar{l}_{i_o, j_o})^2}}, \quad (1)$$

where \bar{f} and \bar{l}_{i_o, j_o} represent the mean of f and l_{i_o, j_o} . For each frame f , the maxima of ρ_{i_o, j_o}^f will yield the best matching candidate search location defined as (i_{max}, j_{max}) . The primary goal is to minimize overall computations over a sequence of frames, without changing the value and location of the global maxima for any individual frame. For this purpose, following information may be used:

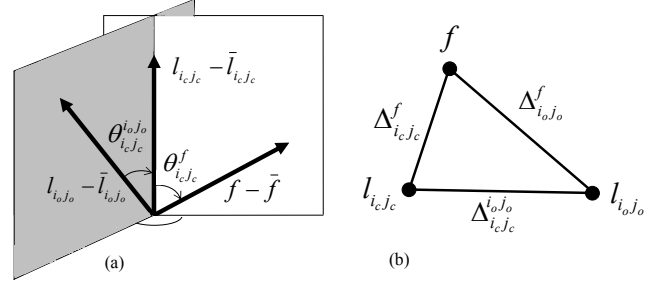


Fig. 1. (a) Normalized angular distance between vectors. (b) Normalized Euclidean distance between points.

1. All candidate locations in the reference image are grouped into Group Of Locations (GOL), such that spatially consecutive locations belong to the same GOL.
2. In a specific GOL, the correlations of the spatially central location, l_{i_c, j_c} , with other locations in that GOL are available and may be used for computational benefit.
3. For a frame f , an initial guess of the best match location is available from a previously matched frame.

We intend to match the frame f only at the central location in each GOL, while for the remaining locations we want to find exact bounds on the correlation, so that unsuitable locations in each GOL may be identified and discarded to reduce the redundant calculations.

3. DERIVATION OF BOUNDS

We want to calculate upper bound on ρ_{i_o, j_o}^f , for all non-central locations in each GOL, so that unsuitable reference locations may be identified and eliminated from the search space. If at a specific location, upper bound on ρ_{i_o, j_o}^f turns out to be smaller than the current known maxima, ρ_{max} , actual calculations of ρ_{i_o, j_o}^f at that specific location becomes redundant and may be skipped without any loss of accuracy. For this purpose, we intend to use the bounds on correlation coefficient, derived by using its relationship with normalized angular distance (Figure 1a):

$$\rho_{i_o, j_o}^f = \cos(\theta_{i_o, j_o}^f), \quad (2)$$

where θ_{i_o, j_o}^f is the angular distance between $[f - \bar{f}]$ and $[l_{i_o, j_o} - \bar{l}_{i_o, j_o}]$, considering them vectors in $m \times n$ dimensional space. Let θ_{i_c, j_c}^f be the angular distance between vectors $[f - \bar{f}]$ and $[l_{i_c, j_c} - \bar{l}_{i_c, j_c}]$ and $\theta_{i_c, j_c}^{i_o, j_o}$ be the angular distance between vectors $[l_{i_o, j_o} - \bar{l}_{i_o, j_o}]$ and $[l_{i_c, j_c} - \bar{l}_{i_c, j_c}]$. By the triangular inequality of distance measures [9]:

$$|\theta_{i_c, j_c}^f - \theta_{i_c, j_c}^{i_o, j_o}| \leq \theta_{i_o, j_o}^f. \quad (3)$$

Taking cosine of both sides and simplifying, we get the following upper bound on $\rho_{i_o j_o}^f$ [9]:

$$\rho_{i_o j_o}^f \leq \rho_{i_c j_c}^{i_o j_o} \rho_{i_c j_c}^f + \sqrt{(1 - (\rho_{i_c j_c}^{i_o j_o})^2)(1 - (\rho_{i_c j_c}^f)^2)}. \quad (4)$$

where $\rho_{i_c j_c}^{i_o j_o}$ is the correlation coefficient between central location, $l_{i_c j_c}$, and a non-central location, $l_{i_o j_o}$, with in a GOL. A similar bound can also be derived by using the relationship between correlation coefficient and the normalized Euclidean distance [8]:

$$\rho_{i_o j_o}^f = 1 - \frac{1}{2} [\Delta_{i_o j_o}^f]^2, \quad (5)$$

where normalized Euclidean distance $\Delta_{i_o j_o}^f$, is defined as:

$$\Delta_{i_o j_o}^f = \sum_{i=1}^n \sum_{j=1}^m \left[\frac{f(i, j) - \bar{f}}{\sigma_f} - \frac{l_{i_o j_o}(i, j) - \bar{l}_{i_o j_o}}{\sigma_{i_o j_o}} \right]^2. \quad (6)$$

Let $\Delta_{i_c j_c}^f$ be the normalized Euclidean distance between f and $l_{i_o j_o}$ and $\Delta_{i_c j_c}^{i_o j_o}$ be the distance between $l_{i_o j_o}$ and $l_{i_c j_c}$, then by the triangular inequality of distance measures (Figure 1b):

$$|\Delta_{i_c j_c}^f - \Delta_{i_c j_c}^{i_o j_o}| \leq \Delta_{i_o j_o}^f, \quad (7)$$

Squaring both sides and using Equation (5), another upper bound on the correlation coefficient is obtained:

$$\rho_{i_o j_o}^f \leq (\rho_{i_c j_c}^{i_o j_o} + \rho_{i_c j_c}^f - 1) + 2\sqrt{(1 - \rho_{i_c j_c}^{i_o j_o})(1 - \rho_{i_c j_c}^f)}. \quad (8)$$

Both of the bounds given by Equations (4) and (8) can be used to eliminate unsuitable search locations in a GOL, however a tighter bound can produce more elimination. Therefore we theoretically compare these two bounds and show that the bound derived by the normalized angular distance in Equation (4) is tighter than the bound based upon the normalized Euclidean distance given by Equation (8).

Theorem 1 *Upper bound on correlation coefficient based upon normalized angular distance is tighter than the bound based upon normalized Euclidean distance*

Proof: From Equation (5):

$$\Delta_{i_c j_c}^{i_o j_o} \Delta_{i_c j_c}^f = 2\sqrt{(1 - \rho_{i_c j_c}^{i_o j_o})(1 - \rho_{i_c j_c}^f)}. \quad (9)$$

Since $0 \leq \Delta \leq 2$ and $-1 \leq \rho \leq +1$, therefore following inequality will always be true:

$$\frac{\Delta_{i_c j_c}^{i_o j_o} \Delta_{i_c j_c}^f}{4} + \frac{1}{2} \sqrt{(1 + \rho_{i_c j_c}^{i_o j_o})(1 + \rho_{i_c j_c}^f)} \leq 1. \quad (10)$$

From the non-negativity property of the distance measures:

$$\Delta_{i_c j_c}^{i_o j_o} \Delta_{i_c j_c}^f \geq 0. \quad (11)$$

Combining Equations (10) and (11):

$$\Delta_{i_c j_c}^{i_o j_o} \Delta_{i_c j_c}^f \left[1 - \frac{\Delta_{i_c j_c}^{i_o j_o} \Delta_{i_c j_c}^f}{4} - \frac{1}{2} \sqrt{(1 + \rho_{i_c j_c}^{i_o j_o})(1 + \rho_{i_c j_c}^f)} \right] \geq 0. \quad (12)$$

Adding factor $1 - \frac{1}{2} [(\Delta_{i_c j_c}^{i_o j_o})^2 + (\Delta_{i_c j_c}^f)^2]$ on both sides of the equation and simplifying by combining the related terms we get:

$$\begin{aligned} & \rho_{i_c j_c}^{i_o j_o} \rho_{i_c j_c}^f + \sqrt{(1 - (\rho_{i_c j_c}^{i_o j_o})^2)(1 - (\rho_{i_c j_c}^f)^2)} \\ & \leq (\rho_{i_c j_c}^{i_o j_o} + \rho_{i_c j_c}^f - 1) + 2\sqrt{(1 - \rho_{i_c j_c}^{i_o j_o})(1 - \rho_{i_c j_c}^f)}, \end{aligned} \quad (13)$$

which shows that the upper bound on correlation coefficient based upon angular distance is tighter than the bound based upon Euclidean distance. Therefore in the following sections, we will use the bound based upon normalized angular distance for eliminating the unsuitable search locations.

4. CORRELATION ELIMINATION ALGORITHM

4.1. Basic Idea

In Correlation Elimination Algorithm (CEA), all valid search locations in the reference image are divided into GOLs, such that the spatially consecutive locations are assigned to the same group. In each GOL, the correlation coefficient between the spatially central location and the other locations, $\rho_{i_c j_c}^{i_o j_o}$, is pre-computed and remains available along with the reference image. During the matching process, each target frame is initially matched with the central location of each GOL to get $\rho_{i_c j_c}^f$. At all non-central locations in the GOL, upper bound on correlation coefficient is evaluated using the Equation (4). If at a search location, the upper bound on correlation is found to be smaller than the current known maxima, the actual correlation at that location can never exceed the current maxima. Therefore, the correlation calculations at that specific location becomes redundant and may be skipped without causing any change in value or location of the global maxima. However, if the upper-bound at a specific location evaluates larger than the current known maxima, at that location correlation will be computed. Thus, if large number of search locations are eliminated by the bound comparison, correlation computations are done on very few locations and the overall matching process becomes very fast.

4.2. Size and Shape of GOL

In order to get good elimination performance, local auto-correlation within a GOL must be strong. Therefore, the size and shape of GOL at a specific reference image location depends upon the height, width and shape of the local 2D auto-correlation function at that location. If local auto-correlation function has significant height over a larger width, a larger sized GOL may be selected having same shape as the

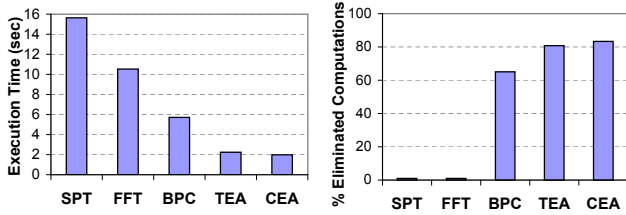


Fig. 2. Comparison of Correlation Elimination Algorithm (CEA) with Spatial (SPT), FFT, BPC and TEA: (a) Average execution time per frame. (b) Average % computation elimination .

local auto-correlation contour. Moreover, the width and shape of the local-autocorrelation function varies from location to location in the same image, therefore in general the size and shape of GOL will also vary. However, in order to define arbitrary size and shape of GOL, additional memory structures and computations are required. In order to avoid these overheads, we assume all GOLs to be of same shape and same size. This assumption renders the implementation of CEA significantly simple at the cost of some loss of performance.

4.3. Computation of Local Auto-Correlation Function

In the proposed algorithm, local auto-correlation within all GOLs must be available before starting the actual matching process. The computation of local auto-correlation function appears to be an additional overhead cost which is equivalent to the cost of matching one target frame on all non-central locations in the reference image. Since this is only one time cost, it is distributed over all target frames which are to be matched against one specific reference image. Thus, if the number of target frames is significantly large, the cost of local auto-correlation function becomes insignificant. Moreover, in most of the video to reference image alignment applications, the size of target frame is a fixed design parameter and reference image is available before starting the actual matching process. In such cases, the local auto-correlation function on the reference image can be pre-computed off line. Therefore the cost of local auto-correlation function computation is either nil or negligibly small.

5. EXPERIMENTS AND RESULTS

In order to demonstrate the concept of CEA, the multi-satellite multi-temporal image dataset of [9] is used. The reference image is an 800K pixels satellite image and the video sequence consists of 250 frames, each of size 101×101 pixels. All search locations in the reference image are divided into groups, each consisting of 3×3 locations. Each target frame is matched with the central location of each group and for the remaining eight locations, bounds are evaluated. At a specific location, if bound is found to be smaller than

the yet known maxima, correlation calculations at that location are skipped. In our experiments, the CEA algorithm resulted in 83.34% search locations elimination and the execution time is 1.98 Sec/Frame, on a 1.6GHz, 1GB RAM, IBM ThinkPad computer. The spatial domain time is observed to be 15.64 Sec/Frame and the frequency domain time as 10.53 Sec/Frame. The BPC algorithm is implemented assuming 20% correlation area and 80% bound area [10]. The execution time of BPC is observed to be 5.70 Sec/Frame and the average elimination as 65%. The execution time of TEA [9] is 2.24Sec/Frame and the elimination is 81.00%. According to these results, the speedup of the CEA algorithm over spatial domain is 7.9 times, over FFT based implementation is 5.32 times, over BPC implementation is 2.88 times and over TEA is 1.13 times. Thus CEA has shown some improvement in elimination performance and execution time over TEA, however, the main advantage of the CEA is its broader scope.

6. REFERENCES

- [1] M. Shah and R. Kumar, *Video Registration*, Kluwer Academic Publishers, Boston, 2003.
- [2] M. Ghanbari, "Standard Codecs: Image compression to advanced video coding," *IEE Telecom. Series 49, Institute of Electrical Engineers*, vol. 49, 2003.
- [3] W. Li and E. Salari, "Successive elimination algorithm for motion estimation," *IEEE Trans. Image Processing*, vol. 4, no. 1, pp. 105–107, January 1995.
- [4] Montrucchio and B. Quaglia, "New sorting-based lossless motion estimation algorithms and a partial distortion elimination performance analysis," *IEEE Trans. CSVT*, vol. 15, pp. 210–212, 2005.
- [5] W. K. Pratt, *Digital Image Processing, 3rd Ed.*, 2001.
- [6] M. Irani and P. Anandan, "Robust multi-sensor image alignment," *IEEE ICCV*, 1998.
- [7] Y. Sheikh and M. Shah, "Aligning dissimilar images directly," *ACCV*, 2004.
- [8] A. Mahmood and S. Khan, "Early termination algorithms for correlation coefficient based block matching," in *IEEE ICIP*, September 2007.
- [9] A. Mahmood and S. Khan, "Exploiting inter-frame correlation for fast video to reference image alignment," in *ACCV*, November 2007.
- [10] di Stefano L., S. Mattoccia, and F. Tombari, "ZNCC-based template matching using bounded partial correlation," *Patrr. Rec. Ltr.*, vol. 26, no. 14, 2005.
- [11] J.P. Lewis, "Fast normalized cross-correlation," *Vis. Inter.*, pp. 120–123, 1995.