An Algorithm of Camera Sabotage Detection Using Contourlet

SHUANG LIANG
Peking University
School of Mathematical Science
No.5 Yiheyuan Road, Beijing
P.R.China
liangshuang12@pku.edu.cn

TING LI
Mechanical Engineering College
No.97 Heping west Road, Shijiazhuang
P.R.China
hbsdliting@sohu.com

WEI GUO
Hebei University of Technology
Science College
Xiping Road No.5340, Shijiazhuang
P.R.China
gwpku2008@aliyun.com

YU WANG
Peking University
School of Mathematical Science
No.5 Yiheyuan Road, Beijing
P.R.China
wangyu_amo@pku.edu.cn

Abstract: Contourlet is one of the new topics in image processing and video processing. Besides a lot of theoretical works about contourlet transform, its applications have roused enough interest as a critical means of multi-scale geometric analysis. This article, focusing on camera sabotage detection, extends the application of contourlet transform to video processing. A new algorithm to detect camera sabotage based on contourlet transform is proposed and experiment results show that the proposed algorithm is powerful and more efficient. Moreover, a comparison of results obtained by this algorithm and the one based on the 9-7 wavelet is made.

Key–Words: Camera sabotage detection, Contourlet transform, Filter design, McClellan transform

1 Introduction

Nowadays the intense video surveillance systems exit almost everywhere especially in public. Monitoring these systems requests a huge amount of time and effort from the security personnel. This prompts the development of intelligent video surveillance system which can trigger the alarm when specific events occur. One typical part of the specific events is camera sabotage made by some suspects intending to prevent the cameras from recording evidence. It is necessary to detect the camera sabotages automatically and to send the alarm in time.

Cases of camera sabotage include any deliberate behaviors on a camera, which can be divided into three categories: 1) obscured camera, the camera lens are occluded with opaque objects; 2) reduced visibility, the visibility of the camera frames is reduced because the focus setting of the camera is changed or the lens is sprayed paint; 3) camera lens displacement, the lens is turned round away from the designated field.

The definition of camera sabotage is earliest proposed by P.Gil-Jimenez [1]. The main idea in sabotage detection is to identify abnormal cases by comparisons between the video frames and the background frames or a group of previous frames. The algorithms of camera sabotage detection are usually classified into two categories: the algorithms based on non-background [2, 3, 4] and the algorithms based on background frames [1, 5, 6].

Though options of the measurement of frame dissimilarity are various, the measures based on two-dimensional transforms, which can capture more details and structure in visual, are more sensitive to the drastic changes. And these changes are more likely to be the sabotage we concern about. In this paper we choose a ”true” two-dimensional transform called contourlet [7], whose construction starts in discrete-domain directly unlike others. And contourlet allows subbands as many as need at each scale, which indicates the real directionality and anisotropy of the transform. Besides, the iterated filter banks in construction and multirate identities make contourlet computationally efficient and more useful in applications.

This paper, sufficiently taking advantage of contourlet transform, gives a new algorithm to detect camera sabotage. In the following sections, we first briefly review the basic principles and implementation steps of the contourlet transform, and then focus on the algorithm of camera sabotage detection. A design of contourlet filters is given through McClellan trans-

form and a process of concrete multirate identities is used to improve the computing efficiency. A comparison of results obtained by our algorithm and the one based on the 9-7 wavelet [5] is made. Experiment results show that the proposed algorithm is powerful and more efficient.

1.1 the Contourlet Transform Framework

Since M.N.Do and M.Vetterli proposed the contourlet, many scholars offer further developments and optimization to the theory and framework, such as critically sampled contourlet transform[8], non-subsampled contourlet transform[9], and so on.

The contourlet transform can provide multiresolution, localization and directionality in image processing as the best image representation requires. In addition, its anisotropy brings it distinctive superiority in capturing geometry and directionality of images. The contourlet decomposition consists of two steps: 1) multiscale step, decomposition by Laplacian pyramid (LP) to capture the singular points in images; 2) directional step, combination of quincunx filter banks (QFB) and shearing operators, called iterated directional filter banks (DFB), to get the singular points with same direction.

The Laplacian pyramid is one way to obtain a multi-scale decomposition of images. Actually, the construction of the Laplacian pyramid and Gaussian pyramid are indivisible. The LP decomposition at each level is the approximate difference of the two adjacent levels of Gaussian pyramid, which is also the bandpass image of the Gaussian pyramid. Do and Vetterli[7, 10] showed that the LP with orthogonal filters provided a tight frame with frame bounds one. Furthermore, they proposed the dual frame operator to achieve the optimal linear reconstruction.

The DFB used in the contourlet transform was proposed in the M.N.Do's doctoral theses[11]. In order to avoid modulating input images, conjugate mirror filter bank (QFB) of fan shape is employed. DFB consists of two parts: 1) a two-channel quincunx filter bank with fan filters to divide two-dimensional spectrum into horizontal and vertical directions; 2) the shearing operator to reorder the image samples without changing the number of them. In contourlet transform, DFB possesses binary tree structure comprised of shearing operators and quincunx filter banks and is used to obtain the desired spectrum decomposition of images[11, 12]. Therefore, filter banks are the main part of contourlet transform and the design of the fan filters is primary in image processing using contourlet transform.

The fan filter is one kind of two-dimensional filters which are usually constructed by McClellan transform[13, 14] or multiple phase representation. Here we adopt the former, which is well known and classical. Let $h(n) = \{h_N, \cdots, h_1, h_0, h_1, \cdots, h_N\}$ be a zero phase symmetric filter of length $2N+1$ and whose Fourier transform be

$$H(w) = \sum_{n=-N}^{N} h(n)e^{-inw} = \sum_{n=0}^{N} a(n)\cos nw,$$

where

$$a(n) = \begin{cases} h(0) & , n = 0 \\ 2h(n) & , elsewhere. \end{cases}$$

Replace $cos(nw)$ by a n-order Chebyshev polynomial of $cos(nw)$, $T_n(\cos nw)$, and use the iterative formula $cos(n+1)w = 2cos(w)cos(nw) - cos(n-1)w$, we rewrite $H(w)$ as

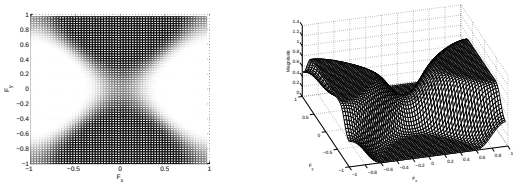
$$H(w) = \sum_{n=0}^{N} \tilde{a}_n \cos^n w.$$

The key point of McClellan transformation is to replace $\cos w$ by a zero phase two-dimensional filter $M(w_1, w_2)$, the typical choice of which is $M(w_1, w_2) = 0.5cos(w_1) + 0.5cos(w_2)$. Then through a series of complicated but elementary computation, the Fourier transform $H(w_1, w_2)$ of the two-dimensional filter can be written as

$$H(w_1, w_2) = H_0(w_1)[h(n_1, n_2)]H_0^T(w_2),$$
$$H_0(w) = (e^{-iNw}, \cdots, e^{-iw}, 1, e^{iw}, \cdots, e^{iNw})$$

where $[h(n_1, n_2)]$ is a $(2N+1)\times(2N+1)$ symmetric matrix. Since the matrix gotten in above is of diamond shape, a π phase shift in w_1 or w_2 direction is needed to make the two-dimensional filter be of fan shape. Taking $N = 4$ as an example, a filter $[h(n_1, n_2)]$ of diamond shape is

Taking a 31×31 filter of fan shape ($N = 15$) as example, which is shown in Figure 1. Its according one-dimensional filter is constructed based on the Hamming window with bandpass cutoff frequency $w_p = 0.4$, stopband cutoff frequency $w_s = 0.6$.



(a) (b)

Figure 1: The Two-dimensional Filter of Fan Shape

The multirate identities can turn a tree-structured DFB into a equivalent parallel channel filter bank. The DFB consists of filtering, downsampling and upsampling. In the analysis side it has the former two kind of building blocks. To obtain the equivalent parallel channel filter bank, the operations both of cascading and exchanging are needed.

The way to cascade two operations of the same type, filtering or downsampling, is to combine them into one operation. Taking downsampling for example, downsampling with Q_0 and Q_1 is equivalent to downsampling with D which is the product of Q_0 and Q_1 $(D = Q_0Q_1)$. It should be noted that when it comes to upsampling, the product order is switched $D = Q_1Q_0$.

Exchanging the order of downsampling and filtering is to replace a downsampling Q_0 operation followed by the filter $H(\xi)$ by a downsamped filter $H(Q_0^T\xi)$ and downsamplingQ_0. Each branch of DFB can be sorted into two building blocks of filtering and downsampling by exchanging the order. Then cascade the same type of operations to get the equivalent parallel channel filter bank. Figure 2 shows multirate identity of a branch of DFB with three level.

2 Camera Sabotage Detection Algorithm

When camera sabotage happens, the contexts of frames change dramatically. We can use the difference between the current frame and the background frame to identify the camera sabotages. According to the definition of sabotage, our detection aims at three categories: camera occlusion, visibility reduction and lens displacement.

Now that our detection is based on the difference between the current frame and the background frame, background estimation is had to be done first. There are several ways to establish the background frame. Here we choose the one proposed by Anil Aksay in their camera sabotage detection algorithm [5]. Let $I_n(x, y)$ be the intensity (brightness) value at position (x, y) in the n^{th} frame and $B_n(x, y)$ the intensity value of the n^{th} background frame at (x, y). Then the background frame of the n^{th} frame can be updated by the following formula:

$$B_n(x,y) = \begin{cases} aB_n(x,y) + (1-a)I_n(x,y) \\ ,if \quad D_I(x,y) \le T_n(x,y) \\ B_n(x,y), if \quad D_I(x,y) > T_n(x,y) \end{cases}$$

where $D_I(x, y)$ stands for $|I_n(x, y) - I_{n-1}(x, y)|$ and $B_0(x, y) = I_0(x, y)$. The parameter a controls the update rate with a value between 0 and 1. The threshold value $T_n(x, y)$ is updated as follows:

$$T_{n+1}(x,y) = \begin{cases} aT_n(x,y) + (1-a)cD_{I,B}(x,y) \\ ,if \quad D_I(x,y) \le T_n(x,y)) \\ T_n(x,y), if \quad D_I(x,y) > T_n(x,y) \end{cases}$$

where $D_{I,B}(x, y)$ stands for $|I_n(x, y) - B_n(x, y)|$, $c > 1$ is a parameter to be determined and $T_0(x, y)$ is chosen depending on experience. It is obvious that the bigger the parameter c is, the higher the threshold is and the lower the detection sensitivity is.

After having established the background frames, we are ready to set up the algorithm of sabotage detection. First, when a camera is obscured, most parts of the current frames are darkened and the difference between the current frame and the according background frame becomes bigger. Though there are many ways to measure such difference changing, here the histogram of $|I_n - B_n|$ is employed. Let $H(X) = \{H_1(X), \cdots, H_{32}(X)\}$ denote the $32-$bin histogram of an image (matrix) X. Since the coming in images I_n are darker, more gray values of the pixels in I_n become smaller and so $\max H(I_n)$ increases while $\max H(B_n)$ keeps unchanged. On the other hand, as most values of $|I_n - B_n|$ are amplified, $H_i(|I_n - B_n|)$ decreases in left bins (small $i's$) and increases in right bins (big $i's$). When the first several $H_i(|I_n - B_n|)$'s are reduced by a certain amount, camera occlusion is considered happened. Therefore, The camera is occluded if the following two inequalities hold.

$$\begin{aligned} &\max H(I_n) > Th_1 * \max H(B_n), \\ &\textstyle\sum_{i=1}^{32} H_i(|I_n - B_n|) > Th_2 * \sum_{i=1}^{3} H_i(|I_n - B_n|), \end{aligned} \tag{1}$$

where Th_1 and Th_2 are two thresholds.

Second, when the visibility of a camera lens is significantly reduced, a lot of details of the images will lose, which means many lines and curves of the images missing. The so called high frequency energy based on contourlet coefficients is a really suitable tool to measure such line missing. Let us consider an 8-subband contourlet transform, and denote the ith subband output of an image X as $C_i(X) = [C_i(x, y; X)]$, $i = 1, 2, \cdots, 8$, where (x, y) is the position of a pixel. The high frequency energy of image X is defined as

$$E(X) = \sum_{i=1}^{8} \sum_{x,y} |C_i(x, y; X)|.$$

At normal condition, the frequency energies of current frames $E(I_n)$ and background frames $E(B_n)$ have some kind of balance. When the visibility of the camera lens is significantly reduced, $E(I_n)$ will suddenly

decline. Hence, if the following inequality is satisfied, it can be determined that the reduced visibility happens.

$$E(I_n) < Th_3 * E(B_n), \tag{2}$$

where $Th_3 \in (0, 1)$ is another threshold.

Finally, when the camera lens is moved away, the numbers of singular points (edge points) of the video sequence probably do not change much, but the orientations of the edges must change a lot. Again, contourlet transform can play the role of detecting such changes. Let us define a matrix of orientational factors according to every subband of contourlet transform of an image X with 3 levels(8-subband) .

$$F_{C_i}(X) = [F_{C_i}(x, y; X)], \quad i = 1, 2, \cdots, 8.$$

$$F_{C_i}(x, y; X) = \begin{cases} 1 & , C_i(x, y; X) > Th_4 * M \\ 0 & , else \end{cases} \tag{3}$$

where M represents $\max C_i(x, y; X)$ and Th_4 is a threshold and usually picks value of 0.5. The orientational matching factor of the current frame I_n and the background frame B_n is defined as

$$F_{matching} = \sum_{i=1}^{8} \sum_{x,y} F_{C_i}(x, y; I_n) F_{C_i}(x, y; B_n). \tag{4}$$

Obviously, the bigger the orientational matching factor $F_{matching}$ is, the more similar I_n and B_n are. So the the ratio $F_{matching}/F_{total}$ is an estimation of camera lens displacement, where

$$F_{total} = \sum_{i=1}^{8} \sum_{x,y} F_{C_i}(x, y; I_n) + \sum_{i=1}^{8} \sum_{x,y} F_{C_i}(x, y; B_n).$$

That is to say, the camera is viewing toward to a different scene if

$$F_{matching}/F_{total} < Th_5, \tag{5}$$

where $Th_5 \in (0, 0.5)$ is also a threshold.

In summary, our algorithm includes three judgements, (1),(2) and (5) , to discriminate three kinds of camera sabotage,camera occlusion, visibility reduction and lens displacement. This algorithm sufficiently takes the advantage in edges and orientations of contourlet transform.

3 Experimental Results

Since there is no public video library of camera sabotage detection on Internet, our experiments are conducted on the videos recorded in the lab. Total 50 videos have been tested, including 15 normal ones and 35 ones with various sabotages. Some videos may contain more than one kind of sabotages.

According to judgements, (1),(2) and (5), the following four ratios are calculated as time evolves.

$$r_1(n) = \frac{\max H(I_n)}{\max H(B_n)},$$

$$r_2(n) = \frac{\sum_{i=1}^{32} H_i(|I_n - B_n|)}{\sum_{i=1}^{3} H_i(|I_n - B_n|)},$$

$$r_3(n) = \frac{E(I_n)}{E(B_n)},$$

$$r_4(n) = \frac{F_{matching}}{F_{total}},$$

where n represents ordinal number of frames.

Some detected sabotages of camera occlusion, visibility reduction and lens displacement are illustrated in Figure 3, Figure 4 and Figure 5. Figure 3 displays the test results of 3 videos with camera occlusion. The upper row shows the $r_1(n)$ curves and the lower row shows the $r_2(n)$ curves. It can be seen that both curves climb sharply when the camera is obscured. Figure 4 displays the test results of 3 videos with visibility reduction. The upper row shows the $E(I_n)$ curves and $E(B_n)$ curves respectively, the lower row shows the $r_3(n)$ curves. The subgraphs show that $r_3(n)$ curves have immediately responses when the camera's visibility is significantly lessened. Figure 5 displays the test results of 3 videos with lens displacement. The upper row shows the $F_{matching}$ curves and F_{total} curves respectively, the lower row shows the $r_4(n)$ curves. The subgraphs show that $r_4(n)$ curves drop to a much low level when the camera lens is moved away.

It should be pointed out that persistency check must be paid more attention to. Persistency check can reduce the false alarm since the camera sabotage caused by deliberate action usually lasts for some time. Hence, in application, the judgements mentioned above should be satisfied for some continual frames before an alarm is sent out.

To compare our algorithm with that in Ref 5, a statistic on the correctness and errors of detections on these 50 videos is shown in table 1, where PAT means percent alarm true and PED means percent event detected. We make the comparison when the PED is 100% for we focus more on the PED than PAT in practice. The data in table 1 demonstrate that our algorithm based on contourlet has better performance than the one based on the wavelet in different scenarios. Especially when the scenarios has abundant details(scenarios 2 and 3), the algorithm based on wavelet has a slight less performance than the one based on

contourlet. We attribute it to the superiority on image representation of contourlet.

4 Conclusions

In this paper, we propose an algorithm based on contourlet transform to detect camera sabotage. Besides, we design a fan filter needed in the transform and make the computation much more efficient by concrete multirate identities. The figures and data resulting from experiments show that the contourlet transform is a powerful tool in camera sabotage detection. We believe that the flexible multiscale and directional decomposition of the contourlet transform have the potential to be used more and more in image and video processing. And we expect a combination of camera sabotage detection and machine learning in the future to develop more intelligent surveillance system.

References:

[1] P. Gil-Jimenez, R. Lopez-Sastre, P. Siegmann, J. Acevedo-Rodriguez, S. Maldonado-Bascon, Automatic Control of Video Surveillance Camera Sabotage , *International Work-Conference on the Interplay Between Natural and Artificial Computation* Spain, June 18-21,2007, pp 222–231

[2] Evan Ribnick, Stefan Atev, Osama Masoud, Nikolaos Papanikolopoulos, Richard Voyles, *Advanced Video and Signal Based Surveillance 2006,* pp10, IEEE Computer Society Washington, DC, USA ,2006.

[3] Baojun Wang,Camera Tampering Detection Based On Corner,thesis , Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, 2009.

[4] Daw-Tung Lin, Chung-Han Wu,Real-Time Active Tampering Detection of Surveillance Camera and Implimentation on Digital Signal Prcessor, *International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Piraeus, 2012,pp 383-386.

[5] Anil Aksay, Alptekin Temizel, A. Enis Cetin, Camera Tamper Detection Using Wavelet Analysis for Video Surveillance, *Advanced Video and Signal Based Surveillance 2007*,IEEE Conference, London, 2007, pp 558–562

[6] Ali Saglam, Adaptive Camera Tamper Detection for Video Surveillance, thesis, the Graduate School of Informatics, the Middle East Technical University, 2009

[7] Minh N. Do, Martin Vetterli,The Contourlet Transform: An Efficient Directional Multiresolution Image Representation, *Image Processing*, IEEE Transaction on Volume 14 Issue 12,pp 2091-2106.

[8] Y. Lu, M. N. Do, CRISP-Contourlet: a critically sampled directional multiresolution image representation, *Wavelets: Application in Signal and Image Processing X*,SanDiego, California, USA, 2003, Proc.SPIE5207.

[9] A. L. Cunha, Jianping Zhou, N. Minh. Do, The Nonsubsampled Contourlet Transform: Theory Design and Application, *Image Processing, IEEE Transactions*,2006, vol15, issue 10, pp 3089-3101.

[10] M. N. Do, M. Vetterli, Framing pyramids, *IEEE Trans. Signal Proc.*,2003, pp 2329-2342.

[11] M. N. Do, Directional Multiresolution Image Representaiton, thesis, Swiss Federal Institute of Technology, Lausanne, Switzerland, 2001

[12] M. N. Do, Yue M. Lu, Multidimensional Filter Banks and Multiscale Geometric Representations, *Foundations and Trends in Signal Processing*,2011, Vol5, Issue3, pp 157-264

[13] Russell M. Mersereau, Wolfgang F. G. Mecklenbrauker, Thomas F. Quatieri, JR, McClellan Transformations for Two-Dimensional Digital Filtering:I-Design, *IEEE TRANSACTION ON CIRCUITS AND SYSTEMS*,VOL.CAS-23, No.7, 1976.

[14] Wolfgang F. G. Mecklenbrauker, Russell M. Mersereau, McClellan Transformations for Two-Dimensional Digital Filtering:II-Implementation, *IEEE TRANSACTION ON CIRCUITS AND SYSTEMS*, VOL.CAS-23, No.7, 1976.

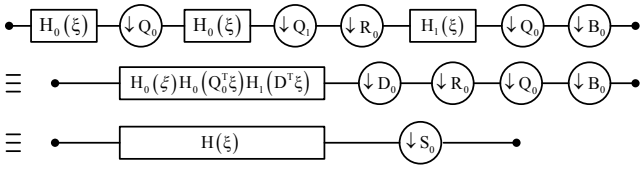
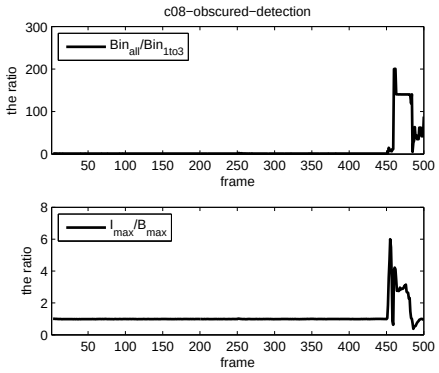


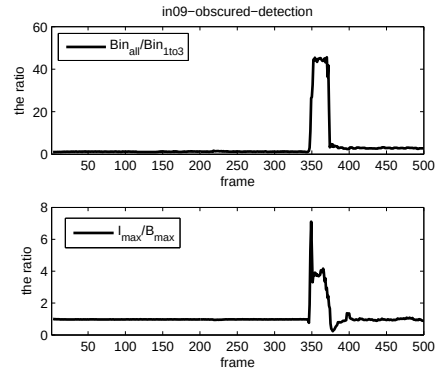
Figure 2: Multirate Identity

		PAT	PED
scenarios 1	9/7 wavelet	88.89%	100%
	contourlet	94.12%	100%
scenarios 2	9/7 wavelet	83.33%	100%
	contourlet	91%	100%
scenarios 3	9/7 wavelet	83.82%	100%
	contourlet	90.91%	100%

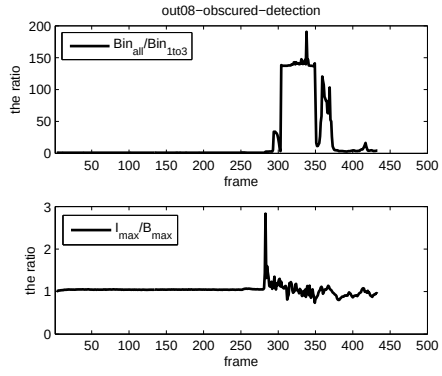
Table 1: Comparison of camera sabotage detection base on 9/7 wavelet and contourlet.



(a)

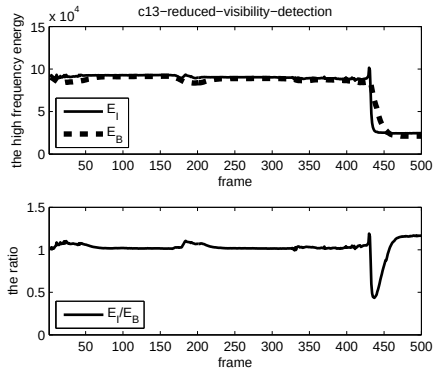


(b)

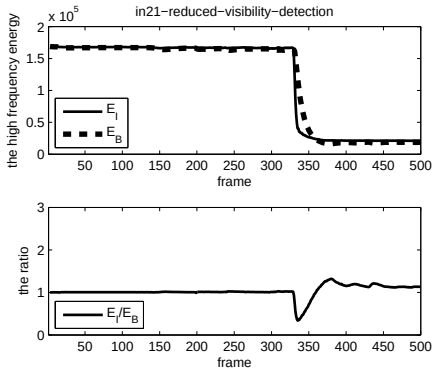


(c)

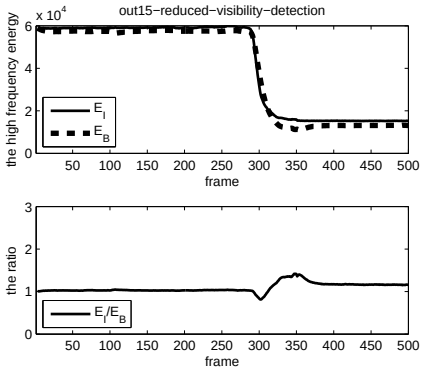
Figure 3: Obscured Detection



(a)

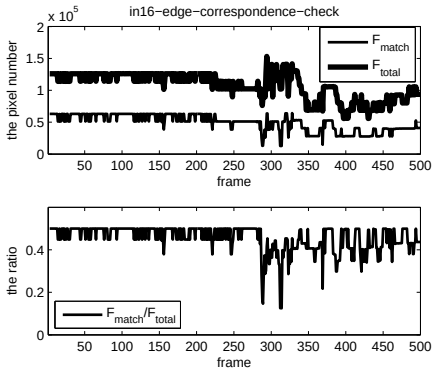


(b)

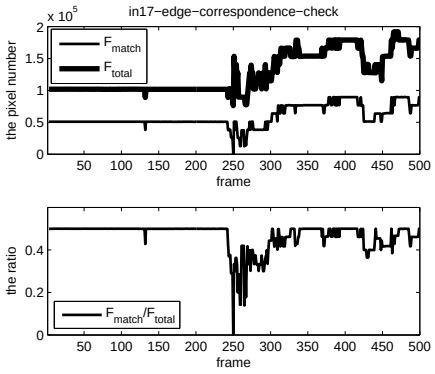


(c)

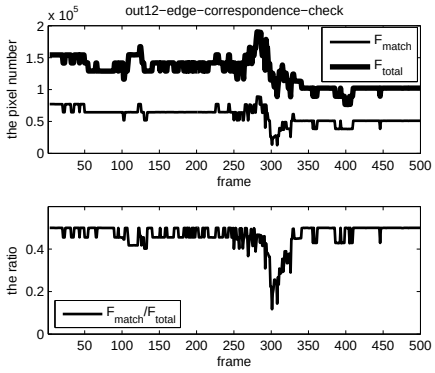
Figure 4: Reduced Visibility Detection



(a)



(b)



(c)

Figure 5: Edge Correspondence Detection